Numerical Contributions to the Asymptotic Theory of Robustness

Von der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

von

Matthias Kohl
geboren am 04.08.1973 in Vilseck i.d. Opf.

1. Gutachter: Prof. Dr. H. Rieder (Universität Bayreuth)
2. Gutachter: Prof. Dr. S. Morgenthaler (EPFL Lausanne)

Tag der Einreichung: 09.09.2005
Tag des Kolloquiums: 15.12.2005

Einführung und Zusammenfassung

Dieser Dissertation ist eine CD beigefügt, welche die .pdf und .ps Version dieses Dokuments sowie das Windows® Installationsprogramm und den Quellcode für `R 2.1.1 patched` (vgl. R Development Core Team (2005)) enthält. Zusätzlich befindet sich auf dieser CD: Unser `R` bundle `RobASt` (vgl. Anhang D), welches aus den `R` Paketen `distrEx`, `RandVar`, `ROptEst`, `RobLox`, `ROptRegTS` und `RobRex` besteht, sowie die erforderlichen `R` Pakete `distr` (vgl. Ruckdeschel et al. (2005)), `setRNG` (vgl. Gilbert (2004)) und `evd` (vgl. Stephenson (2004)).
Um auf unsere Thematik hinzuführen, beginnen wir mit einem einleitenden Beispiel, das deutlich macht, warum robuste Statistik nötig ist. Daran anschließend folgt eine ausführliche deutsche Zusammenfassung dieser in Englisch verfassten Dissertation.

Robuste Statistik ist notwendig!

Diese Aussage findet sich ausführlich begründet zum Beispiel in Abschnitt 1.1 von Huber (1981) und Kapitel 1 von Huber (1997) bzw. in Abschnitt 1.2 von Hampel et al. (1986).

Basierend auf linearer Regression und Kovarianzmatrizen gibt Marazzi (1993), in seiner Einleitung, eine schöne Motivation für robuste Methoden.

Wir verwenden stattdessen das noch einfachere eindimensionale normale Lokationsmodell; d.h., $P_\theta = \mathcal{N}(\theta, \sigma)$, wobei $\sigma = 1$ bekannt ist. Obwohl dies wohl das bekannteste Modell in der robusten Statistik ist, werden einige neue Ideen und Aspekte (Resultate für endliche Stichproben, höhere Ordnungsasymptotik) präsentiert.

In unserem Ansatz, der Setup von infinitesimalen Umgebungen, sehen wir — wie auch Huber (1997) S. 61 — das Ziel von Robustheit darin, gegen Abweichungen von den Voraussetzungen abzusichern, die jenseits oder in der Nähe der Grenzen der Feststellbarkeit liegen. Der Zweck dieser Einleitung ist es, quantitativ zu zeigen, dass solche kleinen Abweichungen nicht triviale Effekte auf statistische Verfahren haben können und zugleich nicht mit Sicherheit von Anpassungstests entdeckt werden können; vergleiche Bemerkung 4.2.7 von Rieder (1994). Im Gegensatz dazu sind robuste Verfahren sehr stabil und verlieren nur wenig Effizienz im idealen Modell.

Gross Error Modell

Wie in Unterabschnitt 1.2c von Hampel et al. (1986) festgestellt, sind $1-10\%$ "falsche Werte" (gross errors (grobe Fehler), Ausreißer) typisch für regelmäßig anfallende Daten. Solche realen Datensätze können durch das bekannte "gross error Modell" (Konvexkontamination)

$$Q = (1-\varepsilon)P_\theta + \varepsilon H$$

modelliert werden, wobei H ein beliebiges Wahrscheinlichkeitsmaß und $\varepsilon \in [0,1]$ die Menge an gross errors (Kontamination) ist; vergleiche Tukey (1960).

Infinitesimale Umgebungen

In unserem asymptotischen Setup, welches auf Umgebungen basiert, die mit der Rate $\sqrt{n}$ schrumpfen, haben wir ε mit $r/\sqrt{n}$ für ein $r \in [0,\infty]$ zu identifizieren. Eine Motivation für dieses Schrumpfen in Termen der Ausreißerwahrscheinlichkeit ist in Ruckdeschel (2005a) gegeben. Des Weiteren verwenden wir im Fall von endlichen Stichproben eine Modifikation dieses Modells. Das heißt, für die Stichprobengröße $n \in \mathbb{N}$ und Zufallsvariablen $U_1, \ldots, U_n \overset{\text{u.i.v.}}{\sim} \text{Binom}\,(1, r/\sqrt{n}\,)$ arbeiten wir stattdessen mit den folgenden bedingten Wahrscheinlichkeiten

$$Q_n(r) = \Big\{ \mathcal{L}\left([(1-U_i)X_i + U_i Y_i]_{i=1,\ldots,n} \,\Big|\, \sum U_i < n/2 \right) \Big\}$$

wobei $X_1, \ldots, X_n \overset{\text{u.i.v.}}{\sim} P_\theta$, $(Y_1, \ldots, Y_n) \sim H_n \in \mathcal{M}_1(\mathbb{B}^n)$ und alle Zufallsvariablen stochastisch unabhängig seien. Diese Modifikation ist durch die Beobachtung motiviert, dass kein sinnvoller Schätzer verwertbare Information aus einer Stichprobe mit $\sum U_i \geq n/2$ ziehen kann. Dieses Argument ist analog zu dem, dass es kein Verfahren mit (finitem) Bruchpunkt $> 1/2$ geben kann. Als eine Folge von Theorem 2 in Hoeffding (1963) klingt

$$P\Big(\sum U_i \geq n/2 \Big) \leq \exp\big\{ -2n(0.5 - r/\sqrt{n}\,)^2 \big\}$$

exponentiell ab. Somit ist die obige Modifikation asymptotisch vernachlässigbar; d.h., alle Resultate zur schwachen Konvergenz über infinitesimalen Umgebungen bleiben unverändert gültig. Für weitere Details verweisen wir auf die Abschnitte 2.2–2.4 von Ruckdeschel (2004c).

Bemerkung Es ist ein Resultat von Ruckdeschel (2004b), dass mit dieser Modifikation der $1/\sqrt{n}$ Umgebungen der maximale mittlere quadratische Fehler (MSE) von asymptotisch linearen Schätzern mit beschränkten Influenzkurven ohne weiteres Stutzen der quadratischen Verlustfunktion konvergiert. Was das künstliche Stutzen unbeschränkter Verlustfunktionen betrifft, vergleiche Le Cam (1986), Rieder (1994), Bickel et al. (1998) oder van der Vaart (1998). ////

Schätzer

Als Schätzer wählen wir Mittelwert, Median und robuste Schätzer mit Influenzkurven (vgl. Definition 1.1.1) von Hampel-Gestalt[1]

$$\eta(x) = A[-c \vee x \wedge c] \qquad \text{with } A = [2\Phi(c) - 1]^{-1}$$

wobei $c \in (0, \infty)$ eine geeignete Stutzhöhe und Φ die Verteilungsfunktion von $\mathcal{N}(0, 1)$ sei.

Optimalitätseigenschaften Im Fall von normaler Lokation sind, als Erweiterung der Liste auf S. 285 in Huber (1981), Schätzer mit Influenzkurven von Hampel-Gestalt optimal in verschiedener Hinsicht. Sie minimieren:

(1) Die maximale asymptotische Varianz für symmetrische Konvexkontaminationen; siehe Huber (1964).

(2) Die maximale asymptotische Varianz unter einer Biasschranke für infinitesimale Kontaminations- und Totalvariationsumgebungen; vergleiche Lemma 5 von Hampel (1968), Abschnitt 2.5 von Hampel et al. (1986), Abschnitt 5.5 von Rieder (1994) und Unterabschnitt 1.3.3.

(3) Die maximale finite[2] Unter-/Überschusswahrscheinlichkeit für Kontaminations-/Totalvariationsumgebungen; vergleiche Huber (1968), Rieder (1989) und Teil V.

(4) Die maximale asymptotische Unter-/Überschusswahrscheinlichkeit für infinitesimale Kontaminations-/Totalvariationsumgebungen; vergleiche Rieder (1980) und Teil V.

(5) Den maximalen asymptotischen mittleren quadratischen Fehler (MSE) für infinitesimale Kontaminations- und Totalvariationsumgebungen; siehe Abschnitt 5.5 von Rieder (1994) und Unterabschnitt 1.3.4.

(6) Allgemeiner: Das maximale asymptotische Risiko für infinitesimale Kontaminations- und Totalvariationsumgebungen, wobei das Risiko irgendeine konvexe und isotone Funktion von asymptotischer Varianz und asymptotischem Bias sein kann; vergleiche Ruckdeschel and Rieder (2004).

(7) Die Entwicklung zweiter Ordnung des maximalen asymptotischen MSE für infinitesimale Kontaminationsumgebungen; siehe Ruckdeschel (2004b). ////

Für die Zwecke dieser Einleitung setzen wir $n = 16$ und Radius $r = 0.2$ (d.h., 5% gross errors) und wählen quadratischen Verlust (d.h., MSE). Wir betrachten den asymptotisch optimal-robusten Schätzer für $r = 0.2$ (d.h., $c = 1.492$) sowie die Radius–minimax Schätzer für $r \in [0.1, 0.4]$ (d.h., Radius bekannt bis auf den Faktor 2, $\varepsilon \in [0.025, 0.1]$), $r \in [0, 2.0]$ (d.h., $\varepsilon \in [0, 0.5]$) und $r \in [0, \infty]$. Die entsprechenden asymptotisch optimalen Stutzhöhen sind $c = 1.356$, $c = 0.824$ und $c = 0.718$. Die Definition des Radius–minimax Schätzers findet sich in Abschnitt 2.2.

[1] in Anspielung auf die Lösung, die in Lemma 5 von Hampel (1968) hergeleitet ist
[2] d.h., im Kontext von finiten Stichproben

Wahl der Stutzhöhen

(1) Wir verwenden die asymptotisch optimalen Schätzer, da wir demonstrieren wollen, dass diese Schätzer auch für kleine Stichproben gut funktionieren. Aber die Stutzhöhen könnten auch optimal im Bezug auf den finit maximalen MSE bzw. der Entwickung zweiter oder dritter Ordnung des asymptotischen MSE gewählt werden. Wie numerische Ergebnisse in Ruckdeschel (2004b) zeigen, sind die Unterschiede zwischen diesen verschiedenen Möglichkeiten allerdings klein.

(2) Zudem ergeben sich nur kleine Effizienzverluste, falls wir unterschiedliche asymptotische Risiken verwenden, um die optimale Stutzhöhe c zu bestimmen; siehe Abschnitt 7.2 von Ruckdeschel and Rieder (2004).

(3) Unter einer zusätzlichen Homogenitätsbedingung an die Verlustfunktion, welche zum Beispiel bei allen L_q Risiken mit $q \geq 1$ erfüllt ist, ist der Radius–minimax Schätzer für $r \in [0, \infty]$ unabhängig von der gewählten Verlustfunktion; vergleiche Abschnitte 6 und 7.3 von Ruckdeschel and Rieder (2004). Diese Risikounabhängigkeit des optimalen Verfahrens ist auch der Grund dafür, warum dieser Schätzer hier mit eingeschlossen wurde. Eigentlich führt $\varepsilon \in [0, 1]$ auf $r \in [0, 4]$ für den Stichprobenumfang $n = 16$; d.h., Radien $r > 4$ sind eigentlich nicht zugelassen. ////

Maximaler MSE für endliche Stichproben

Unsere Untersuchung endlicher Stichproben bestätigt, was in der robusten Statistik auf der Grundlage asymptotischer Resultate schon lange gängige Meinung ist: In der idealen Situation (d.h., $r = 0$) besitzen die geeignet gewählten asymptotisch optimal-robusten Schätzer einen etwas größeren finit[3] maximalen MSE als der Mittelwert. Jedoch verlieren diese nicht viel Effizienz und arbeiten im idealen Modell deutlich besser als der Median. Demgegenüber ist der finit maximale MSE des Mittelwertes für $r > 0$ unbeschränkt, wohingegen robuste Schätzer einen beschränkten finit maximalen MSE besitzen. Das heißt, bereits kleine Abweichungen vom idealen Modell können beim Mittelwert zu sehr großen Fehlern führen. Insbesondere arbeiten die asymptotisch optimal-robusten Schätzer wieder besser als der Median. Dies sind verbreitete Aussagen; vergleiche etwa die Abschnitte 1.1 und 1.2 von Huber (1981) oder die Abschnitte 1.1 und 1.2 von Hampel et al. (1986).

Die (numerisch) exakte finite Verteilung und die entsprechenden finiten Risiken im Fall von robusten Schätzern mit Hampel-Typ Influenzkurven, welche mit Hilfe des M Prinzips konstruiert sind, können mittels Algorithmen, die in Unterabschnitt 11.3.2 und in Ruckdeschel and Kohl (2005) hergeleitet werden, berechnet werden. Diese Verfahren verwenden in entscheidender Weise die schnelle Fourier Transformation (FFT). In Tabelle 1 finden sich die finit maximalen MSEs für $n = 16$ und Radius $r = 0, 0.2$. In diesen Situationen weist der Median einen Effizienzverlust von mehr als 22% ($r = 0$) bzw. 16% ($r = 0.2$) auf.

Finit versus asymptotisch optimale Stutzhöhen Unsere Untersuchung zeigt, dass die Stutzhöhe, die optimal für endlichen Stichprobenumfang ist, im Allgemeinen kleiner, das heißt, konservativer ist als die asymptotisch optimale Stutzhöhe.

[3] d.h., im Kontext finiter Stichproben

r	mean	$r = 0.2$	$r \in [0.1, 0.4]$	$r \in [0, 2.0]$	$r \in [0, \infty]$	median
0	1.000	1.035	1.049	1.145	1.176	1.446
0.2	∞	1.450	1.431	1.443	1.465	1.713

Table 1: Finit maximaler MSE für $n = 16$ und normale Lokation.

Dies ergibt sich aus der höheren Ordnungsasymptotik und numerischen Auswertungen; siehe Teil V und Ruckdeschel (2004b). Diese Tatsache spiegelt sich auch in den Resultaten in Tabelle 1 wider, in welcher der asymptotisch optimal-robuste Schätzer ($c = 1.492$) einen größeren finit maximalen MSE aufweist als die asymptotischen Radius–minimax Schätzer für $r \in [0.1,\, 0.4]$ ($c = 1.356$) und $r \in [0,\, 2.0]$ ($c = 0.824$). Tatsächlich ist die numerisch bestimmte finit optimale Stutzhöhe gleich $c = 1.130$ und führt auf einen finiten minimax MSE von 1.418. ////

Cniper Kontamination

Der Begriff "cniper" enthält die Bestandteile "nice" (nett) und "pernicious" (schädlich); eine solche Kontamination bedroht die Genauigkeit von Schätzern auf eine unerwartete und gefährliche Weise wie dies ein Heckenschütze (sniper) tut; vergleiche Abschnitt 5 von Ruckdeschel (2004a). Wir lassen nun nicht mehr beliebige $H_n \in \mathcal{M}_1(\mathbb{B}^n)$ zu, sondern nur Kontaminationen durch Diracmaße im Punkt $a \in \mathbb{R}$; d.h.,

$$Q_n(r, a) = \left[(1 - r/\sqrt{n}\,)P_\theta + r/\sqrt{n}\;\mathrm{I}_{\{a\}}\,\right]^{\otimes n}$$

Wir bestimmen den Kontaminationspunkt a minimal, so dass ein gegebener robuster Schätzer unter $Q_n(r, a)$ den klassisch optimalen Schätzer (in diesem Fall: den Mittelwert $\bar{X}_n$) ausschaltet; d.h.,

$$a = \sup\left\{z > 0 \,\middle|\, \mathrm{MSE}_{Q_n(r,z)}(\bar{X}_n) \le \mathrm{MSE}_{Q_n(r,z)}(S_n^c)\right\}$$

wobei S_n^c ein robuster Schätzer mit einer Hampel-Typ Influenzkurve zu einer vorgegebenen Stutzhöhe $c \in (0, \infty)$ sei. Eine Konsequenz hiervon ist, dass der robuste Schätzer S_n^c einen kleineren MSE für eine beliebige kontaminierende Verteilung H mit Träger $[a, \infty)$ (oder $(-\infty, -a]$) besitzt; vergleiche Proposition 5.1 von Ruckdeschel (2004a). Unter $Q_n(r, a)$ erhält er

$$n\,\mathrm{MSE}_{Q_n(r,a)}(\bar{X}_n) = (1 - r/\sqrt{n}\,) + a^2(r^2 + r/\sqrt{n} - r^2/n)$$

siehe Abschnitt 5.3 (ibid.). Daher erhalten wir für $M_c := n\,\mathrm{maxMSE}\,(S_n^c)$

$$a = \sqrt{\frac{M_c - (1 - r/\sqrt{n}\,)}{r^2 + r/\sqrt{n} - r^2/n}}$$

Für unsere robusten Schätzer, die in Tabelle 1 angegeben sind, führt dies auf $a = 2.391$ ($r = 0.2$), $a = 2.345$ ($r \in [0.1, 0.4]$), $a = 2.374$ ($r \in [0, 2.0]$)

und $a = 2.427$ ($r \in [0, \infty]$). Diese kleinen Kontaminationen liegen also deutlich weniger als 2.5 Standardabweichung von der Null entfernt. Man beachte, dass wir unter cniper Kontamination sogar auf weniger Ausreißer treffen, falls Ausreißer unter der Standardnormalverteilung als Beobachtungen definiert werden, die betragsmäßig größer als 2.5 sind; genauer werden so im idealen Modell die 1.24% größten Beobachtungen als Ausreißer identifiziert, hingegen sind dies unter cniper Kontamination $(1 - \varepsilon)1.24\% = 1.18\%$. Daher kann diese Situation, welche die Überlegenheit des Mittelwerts zerstört, sicherlich als harmlos bezeichnet werden.

Eine kleine Simulationsstudie

Als nächstes stellen wir die Ergebnisse einer kleinen Simulationsstudie in dem oben eingeführten Teilmodell vor, in der wir den empirischen Fehler zweiter Art von Anpassungstests und den empirischen MSE von Lokationsschätzern berechnet haben. Zu diesem Zweck simulierten wir $M = 1e05 = 10^5$ Stichproben der Größe $n = 16$ mit Radius $r = 0.2$ (d.h., $\varepsilon = 0.05$ und $P(\sum U_i \geq 8) = 3.50e-07$). Im Hinblick auf die obigen Resultate sollte $a = 2.45$ ($\Phi(-2.45) \approx 0.71\%$) ausreichen, damit unsere robusten Schätzer den Mittelwert übertreffen. Um Replikationen zu vermeiden, verwendeten wir $H = \text{Unif}([2.45, 2.46])$ anstelle von $H = \text{I}_{\{2.45\}}$.

Es ist naheliegend, zuerst etwas Diagnostik zu versuchen. Das heißt, wir haben unter Verwendung des `R` Pakets `fBasics` (vgl. Wuertz et al. (2005)) bekannte Tests auf Normalität berechnet. In Tabelle 2 finden sich die empirischen Fehler zweiter Art (die Nullhypothese wird nicht abgelehnt, obwohl sie falsch ist) der betrachteten Tests, wobei wir ein Signifikanzniveau von 5% verwendeten. Wie man sieht, sind die Resultate für die ausgewählten Tests sehr ähnlich. Diese deuten darauf hin, dass die Macht (Fähigkeit die Nullhypothese abzulehnen, falls sie wirklich falsch ist) von Anpassungstests im Fall solch harmloser Kontaminationen sehr klein ist. Daher sollten Schätzer auch unter solch harmlosen Abweichungen von der Normalverteilung ausgewertet und verglichen werden.

Test auf Normalität	Fehler zweiter Art
Anderson-Darling	93.3%
Cramér-von Mises	93.7%
Kolmogorov-Smirnov (Lilliefors)	94.2%
Shapiro-Wilk	93.4%

Table 2: Empirischer Fehler zweiter Art von Tests auf Normalität unter cniper Kontamination.

Bemerkung Diese empirischen Resultate legen nahe, die Nullhypothese von exakter Normalität auf approximative Normalität abzuändern. Dies ist im Sinn von Abschnitt 3 in Rieder (1981b), in dem er die Nullhypothese exakter Symmetrie auf approximative Symmetrie erweitert und einen nichtparametrischen asymptotischen maximin Test herleitet. Die entsprechende Modifikation von Anpassungstest ist

unseres Wissens nach noch offen. ////

Als zweites berechneten wir den empirischen MSE basierend auf den 1e05 Stichproben der Größe 16 und entsprechende 95% Konfidenzintervalle (basierend auf dem zentralen Grenzwertsatz) von Mittelwert, Median und unseren robusten Schätzern; siehe Tabelle 3. Diese Studie ist ähnlich angelegt wie die Studie in Abschnitt 5 von Ruckdeschel (2004b). Sie ist im Geiste der Princeton Robustheitsstudie; vergleiche Andrews et al. (1972). Im Unterschied dazu wählen wir besondere asymptotisch optimale Schätzer aus, vergleichen diese Schätzer bezüglich ihres finiten MSE und betrachten nur cniper Kontamination. Unter der ausgewählten cniper Kontamination besitzt der Mittelwert einen (numerisch) exakten finiten MSE von 1.477 , welcher deutlich innerhalb des angegebenen empirischen Konfidenzintervalls liegt. Im Fall unserer robusten Schätzer wurden die entsprechenden Schätzer als M Schätzer bzw. als ein-Schritt Schätzer ausgehend vom Median bestimmt. Im Hinblick auf das allgemeine Konstruktionsproblem haben wir auch die entsprechenden ein-Schritt Schätzer mit eingeschlossen. Wie man sieht, schlagen unsere robusten Schätzer tatsächlich den Mittelwert und den Median, wobei die Resultate für das M Prinzip und die ein-Schritt Methode sehr ähnlich sind.

Estimator		$n \times$ Emp. MSE	95% conf. interval
mean		1.480	[1.467, 1.493]
$r = 0.2$:	M Prinzip	1.445	[1.431, 1.458]
	ein-Schritt Konstruktion	1.434	[1.420, 1.447]
$r \in [0.1, 0.4]$:	M Prinzip	1.428	[1.414, 1.441]
	ein-Schritt Konstruktion	1.423	[1.410, 1.436]
$r \in [0, 2.0]$:	M Prinzip	1.441	[1.428, 1.454]
	ein-Schritt Konstruktion	1.448	[1.435, 1.461]
$r \in [0, \infty]$:	M Prinzip	1.462	[1.449, 1.476]
	ein-Schritt Konstruktion	1.468	[1.455, 1.481]
median		1.712	[1.696, 1.727]

Table 3: Empirischer MSE für normale Lokation, Stichprobenumfang $n = 16$ und Radius $r = 0.2$ unter cniper Kontamination.

Bemerkung

(1) M Prinzip und ein-Schritt Konstruktion funktionieren sogar noch gleich gut für kleinere Stichprobenumfänge. Aus den Arbeiten von Ruckdeschel (2004b) und Ruckdeschel (2005e) über höhere Ordnungsasymptotik des MSE von robusten Schätzern mit Hampel-Typ Influenzkurven ergibt sich, dass im Fall von normaler Lokation die M Schätzer und die ein-Schritt Schätzer die gleiche asymptotische Entwicklung bis zur zweiten Ordnung besitzen.

(2) Der Median erfüllt die Voraussetzungen an einen Startschätzer ($\sqrt{n}$ konsistent auf vollen $1/\sqrt{n}$ Kolmogorov-Umgebungen). Dies wird in Unterabschnitt 2.3.4

gezeigt werden. Für weitere Einzelheiten über ein-Schritt Konstruktionen verweisen wir auf Abschnitt 6.4 in Rieder (1994) und Abschnitt 2.3.

(3) In seinem Theorem 3.4 (b) zeigt Ruckdeschel (2004b), dass Kontamination rechts von $a_n := c(1 + A\sqrt{2\log(n)/n})$ mit $A = [2\Phi(c) - 1]^{-1}$ im wesentlichen ausreicht, damit ein robuster Schätzer mit einer Hampel-Typ Influenzkurve zu einer vorgegebenen Stutzhöhe $c \in (0, \infty)$ seinen maximalen asymptotischen MSE bis zur dritten Ordnung annimmt. Im Fall unserer robusten Schätzer führt dies auf $a_n = 2.508$ $(r = 0.2)$, $a_n = 2.324$ $(r \in [0.1, 0.4])$, $a_n = 1.645$ $(r \in [0, 2.0])$ und $a_n = 1.520$ $(r \in [0, \infty])$. Daher ist es nicht überraschend, dass die vorherigen empirischen MSEs unter cniper Kontamination (vgl. Table 3) bereits sehr nahe bei den finit maximalen MSEs, die in Tabelle 1 ausgewertet wurden, liegen. ////

Fazit

Gehen wir wie üblich von $1 - 10\%$ gross errors in den Beobachtungen aus, so führt dies zu folgendem Fazit:

(1) Unter cniper Kontamination übertreffen unsere asymptotisch optimal-robusten Schätzer den Mittelwert und den Median.

(2) Solche kleinen Abweichungen können nicht mit Sicherheit von Anpassungstests entdeckt werden.

(3) Unsere asymptotisch optimal-robusten Schätzer funktionieren gut bis hinunter zu kleinen Stichprobenumfängen; insbesondere, scheinen die Radius–minimax Schätzer für $r \in [0.1, 0.4]$ (d.h., $\varepsilon \in [0.025, 0.1]$) und $r \in [0, 2.0]$ (d.h., $\varepsilon \in [0, 0.5]$) eine gute Wahl für regelmäßig erhobene Daten zu sein, falls der Umgebungsradius nur ungefähr bekannt ist.

(4) M Prinzip und ein-Schritt Konstruktion funktionieren gleich gut.

Vorschlag

Im Fall von regelmäßig erhobenen Daten, die von einem idealen normalen Lokationsmodell stammen, legt das vorangegangene Fazit die folgende Vorgehensweise nahe:

Schritt 1: Versuche in Abhängigkeit von der Qualität der Daten, eine ungefähre Schätzung für den Anteil $\varepsilon \in [0, 1]$ an gross errors zu finden, so dass $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$.

Schritt 2: Berechne die Influenzkurve unseres asymptotisch optimalen Radius–minimax Schätzers für $r \in [\sqrt{n}\,\underline{\varepsilon}, \sqrt{n}\,\overline{\varepsilon}]$ unter Verwendung der generischen `S4` Funktion `radiusMinimaxIC` aus dem `R` Paket `ROptEst`, welches Bestandteil unseres `R` bundle's `RobASt` ist; siehe Anhang D.

Schritt 3: Wähle und berechne einen geeigneten Startschätzer. Mögliche Kandidaten sind der Median, der MAD oder der Kolmogorov(–Smirnov) MD Schätzer (vgl. die generische `S4` Funktion `ksEstimator` im Paket `ROptEst`).

Schritt 4: Schätze den interessierenden Parameter mit Hilfe der ein-Schritt Konstruktion unter Verwendung der generische `S4` Funktion `oneStepEstimator` aus dem Paket `ROptEst`. ////

In dieser Dissertation werden wir zeigen, dass der obige Vorschlag nicht nur im Fall

normaler Lokation funktioniert, sondern im Fall allgemeiner glatt parametrischer idealer Modelle wie Exponentialfamilien oder lineare Regressionsmodelle. Zusätzlich stellen wir die Implementation dieser Modelle und der entsprechenden optimal-robusten Schätzer in Form unseres `R` bundle's `RobASt` zur Verfügung.

Ausführliche Zusammenfassung

Teil I: Asymptotische Theorie der Robustheit

Kapitel 1 enthält eine Beschreibung und Zusammenfassung der asymptotischen Theorie, welche die Grundlage dieser Dissertation bildet. Im Kontext von Robustheit ist diese verbunden mit den Namen von Bickel und Rieder; vergleiche Bickel (1981) und Rieder (1994). Unsere Darstellung basiert auf den Kapiteln 4 und 5 von Rieder (1994). Wir beschränken uns auf die Schätzung eines endlich dimensionalen Parameters im Einstichprobenfall mit u.i.v. Beobachtungen. Genauer betrachten wir eine glatt parametrisierte Familie

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A})$$

von Wahrscheinlichkeitsmaßen auf einem meßbaren Raum $(\Omega, \mathcal{A})$, dessen Parameterraum Θ eine offene Teilmenge eines endlich dimensionalen $\mathbb{R}^k$ ist. Die Familie $\mathcal{P}$ sei L_2 differenzierbar für fixiertes $\theta \in \Theta$,

$$\|\sqrt{dP_{\theta+t}} - \sqrt{dP_\theta}\,(1 + \tfrac{1}{2}t^\tau \Lambda_\theta)\| = o(|t|)$$

mit L_2 Ableitung $\Lambda_\theta \in L_2^k(P_\theta)$ und besitze eine Fisher-Information von vollem Rang k,

$$\mathcal{I}_\theta = \mathrm{E}_\theta\, \Lambda_\theta \Lambda_\theta^\tau$$

Für weitere Einzelheiten zu L_2 bzw. L_r $(r \geq 1)$ Differenzierbarkeit verweisen wir auf Abschnitt 2.3 in Rieder (1994), Abschnitt 1.8 in Witting (1985) sowie auf Rieder and Ruckdeschel (2001).

In Abschnitt 1.1 führen wir (partielle) quadrat-integrierbare Influenzkurven ein (beinhalten eine Matrix $D \in \mathbb{R}^{p\times k}$ von vollem Rang $p \leq k$) und zeigen die folgende notwendige und hinreichende Bedingung für deren Existenz

$$\exists A \in \mathbb{R}^{p\times k} \colon D = A\mathcal{I}_\theta$$

Anschließend werden asymptotisch lineare Schätzer definiert und es wird die Cramér-Rao Schranke für diese Schätzerklasse hergeleitet.

Der Aufbau der infinitesimalen Robustheit, welche mit Umgebungen um das ideale Modell $\mathcal{P}$ arbeitet, die mit der Rate $\sqrt{n}$ schrumpfen, wird in Abschnitt 1.2 vorgestellt. In dieser Dissertation betrachten wir Kontaminations- und Totalvariationsumgebungen sowie gelegentlich Kolmogorov-Umgebungen. Rieder (1994) verwendet zusätzlich noch Hellinger- und Cramér von Mises-Umgebungen.

Im Anschluss daran werden Tangentenklassen für diese Umgebungen definiert und es werden einfache Perturbationen entlang solcher Tangenten anstelle der vollen

Umgebungen betrachten. Als eine Folge des dritten Lemma von Le Cam erhält man die asymptotische Normalität asymptotisch linearer Schätzer unter solchen einfachen Perturbationen. Arbeitet man mit quadratischem Verlust, so führt dies auf das Problem den asymptotischen mittleren quadratischen Fehler (MSE) zu minimieren; vergleiche Unterabschnitt 1.3.1. Dieses konvexe Optimierungsproblem beinhaltet gewisse Biasterme (abhängig vom Umgebungstyp), welche mehr oder weniger explizit berechnet werden können; siehe Unterabschnitt 1.3.2.

Die Lösung zu diesem Optimierungsproblem wird detailiert in Abschnitt 5.5 von Rieder (1994) hergeleitet. Die hierfür benötigten Aussagen über Lagrangemultiplikatoren sind in Anhang B (ibid.) enthalten. Um die entsprechende MSE Lösung zu erhalten, wird zuerst die Spur der asymptotischen Kovarianz unter einer Schranke an die verschiedenen Biasterme minimiert. Aus diesem Grund geben wir auch die Lösung (optimalen Influenzkurven) für dieses Hilfsproblem an; siehe Unterabschnitt 1.3.3. Zusätzlich wird der minimale asymptotische Bias und die Influenzkurve, die diesen minimalen Bias annimmt, spezifiziert. Die optimale Influenzkurve für das ursprüngliche minimax MSE Problem ist von der gleichen Form wie im Fall des minimum Spur Problems für eine geeignet gewählte Biasschranke. Im Fall des MSE ist diese Schranke durch eine zusätzliche implizite Gleichung festgelegt; siehe Unterabschnitt 1.3.4. Die MSE Lösung ist hierbei immer von Hauptfall-Form; vergleiche Theorem 1.3.9 (a).

Kapitel 2 beinhaltet Ergänzungen zur asymptotischen Theorie der Robustheit, welche sich für diese Dissertation als notwendig erwiesen.

Zuerst zeigen wir in Unterabschnitt 2.1.1, dass der Lagrange Multiplikator A, der in den optimalen Influenzkurven auftritt und mit Hilfe eines Optimierungsproblems unter Verwendung von Lagrange Argumenten gewonnen wurde, eine statistische Interpretation besitzt:

$$\text{minimaxMSE} = \operatorname{tr} A$$

Diese Identität stellt eine Erweiterung der klassischen Cramér-Rao Schranke für quadratischen Verlust dar und ist bemerkenswert, da zusätzlich zur Varianz, der Bias Berücksichtigung findet.

Als nächstes behandeln wir diskrete Modelle, welche bisher nur selten in der Robustheitsliteratur betrachten wurden; siehe Unterabschnitt 2.1.2. Diese Modelle zeigen gewisse Besonderheiten: Unter einer zusätzlichen "Lückenbedingung", fällt die MSE Lösung (immer von Hauptfall-Form) mit der minimum Bias Lösung zusammen. Dies geschieht für Radien r größer als ein endlicher Radius $\bar{r} \in [0, \infty)$, der sogenannte "lower case" Radius. Eine weitere Besonderheit, die bisher in der Literatur nicht untersucht wurde, ist die Nicht-Eindeutigkeit der Lagrange Multiplikatoren, welche Bestandteil der (eindeutigen) optimalen Influenzkurven sind.

Im verbleibenden Teil des Abschnitts 2.1 leiten wir technische Eigenschaften der in der MSE Lösung enthaltenen Lagrange Multiplikatoren her: Beschränktheit (siehe Unterabschnitt 2.1.3), Eindeutigkeit (siehe Unterabschnitt 2.1.4) und Stetigkeit (siehe Unterabschnitt 2.1.5). Diese Eigenschaften sind für die folgenden Zwecke von Bedeutung: Bestimmung eines unbekannten Umgebungsradius gemäß eines minimax Kriteriums (siehe Abschnitt 2.2), Schätzerkonstruktion (siehe Abschnitt 2.3) und Konvergenz robuster Modelle (siehe Abschnitt 2.4).

In Abschnitt 2.2 betrachten wir die Begriffe ungünstigster Radius und Radius–minimax Schätzer, welche von Rieder et al. (2001) eingeführt wurden. Dieses Konzept liefert eine Strategie, falls der wahre Umgebungsradius unbekannt ist bzw. nur bekannt ist, dass er innerhalb eines gewissen Intervalls liegt. Wir liefern die mathematischen Beweise für den ungünstigsten Radius, welche die rein numerische Bestimmung in Rieder et al. (2001) absichern und ergänzen.

Ein weiteres wichtiges Problem ist die Konstruktion optimal-robuster Schätzer. Bis hierher betreffen die Ergebnisse die MSE optimalen Influenzkurven, deren Herleitung nur auf den einfachen Perturbationen basiert. Für eine gegebene Familie von Influenzkurven $(\psi_\theta)_{\theta\in\Theta}$ ist es das Ziel, einen asymptotischen Schätzer S zu konstruieren, ohne den Parameter $\theta \in \Theta$ zu kennen, wobei S asymptotisch linear mit Influenzkurve ψ_θ in P_θ sein muss. Hinzu kommt, dass das Risiko dieses Schätzers nicht ansteigen soll, wenn wir von den einfachen Perturbationen zu den vollen Umgebungen um P_θ übergehen. Diese Ziele können (unter zusätzlichen Voraussetzungen) mit Hilfe der ein-Schritt Konstruktion erreicht werden. Hinreichende Bedingungen sind in Unterabschnitt 2.3.1 angegeben.

Ausgehend von den allgemeinen Bedingungen aus Abschnitt 6.4 von Rieder (1994) spezialisieren wir diese in Unterabschnitt 2.3.2 für MSE optimalen Influenzkurven. Diese spezialisierten Bedingungen sind unter anderem für Exponentialfamilien mit vollem Rang erfüllt; vergleiche Unterabschnitt 2.3.3. Folglich können wir die ein-Schritt Methode in etlichen wichtigen Modellen anwenden, welche häufig in der parametrischen Statistik verwendet werden. Insbesondere sind diese Resultate auf die meisten in dieser Dissertation betrachteten Modelle anwendbar.

Die ein-Schritt Konstruktion erfordert einen geeigneten Startschätzer. Nach Theorem 6.3.7 in Rieder (1994) besitzt der Kolmogorov minimum Distanz Schätzer die notwendigen Eigenschaften, falls wir $1/\sqrt{n}$ Umgebungen von Kolmogorov Typ betrachten. Folglich gilt dies auch, falls wir kleinere $1/\sqrt{n}$ Umgebungen, wie Kontaminations- oder Totalvariationsumgebungen, betrachten. Jedoch werden in der Robustheitsliteratur meist der einfachere Median bzw. der Median der absoluten Abweichungen vom Median (MAD) als geeignete Startschätzer vorgeschlagen. Da wir keine Referenz dafür gefunden haben, dass diese Schätzer auch die behaupteten Eigenschaften besitzen, beweisen wir deren uniforme $\sqrt{n}$ Konsistenz auf $1/\sqrt{n}$ Kolmogorov-Umgebungen ohne dabei eine Lokations- oder Skalenstruktur vorauszusetzen; siehe Unterabschnitt 2.3.4.

Im verbleibenden Teil des aktuellen Kapitels leiten wir einige Resultate her, welche als Konvergenz robuster Modelle interpretiert werden können; vergleiche Abschnitt 2.4. Wir zielen dabei aber nicht auf den abstrakten Rahmen von Le Cam (1986) ab, der beliebige Entscheidungsregeln beinhaltet. Vielmehr basiert unser Konzept von Anfang an allein auf den optimal-robusten Schätzern. Wir beweisen, dass unter gewissen schwachen Voraussetzungen und mit geeigneten Standardisierungen die Lagrange Multiplikatoren der MSE optimalen Influenzkurven eines Modells gegen die Lagrange Multiplikatoren eines anderen Modells konvergieren. Daher konvergieren auch der minimax asymptotische MSE, der standardisierte asymptotische Bias und die asymptotische Varianz. Falls es also ein infinitesimal robustes Modell gibt, in dem die optimal-robusten Influenzkurven schwer zu bestimmen sind, können wir versuchen, ein anderes robustes Modell zu finden, welches als Approx-

imation dienen kann und in dem die Berechnung der entsprechenden optimalen Influenzkurven viel einfacher ist. Unter Verwendung dieser Influenzkurven sind wir in der Lage, Approximationen für die optimal-robusten Influenzkurven für das interessierende Modell zu konstruieren, was auch im Sinne von Le Cam (1986) ist. Überzeugende Beispiele sind in den Kapiteln 3 und 4 gegeben. Das Konzept — Konvergenz robuster Modelle — kann sicherlich in abstrakterer Weise erweitert werden.

Teil II: Nicht-standard Modelle

Das Thema dieses zweiten Teils sind bekannte parametrische Modelle, die in der Robustheitsliteratur bisher nur selten betrachtet wurden. Falls übliche robuste Modelle wie Lokation und Skala eingeschlossen sind, so entsprechen unsere Verteilungsannahmen an das ideale Modell nicht dem Standard.

Binomial- und Poissonmodell

Die robuste Schätzung im Binomial- und Poissonmodell hat bisher nur wenig Aufmerksamkeit in der Robustheitsliteratur erhalten. Sie wurde zum ersten Mal in Abschnitt F.3 von Hampel (1968) erwähnt, in dem er die Scorefunktion Λ_θ berechnet und sein Lemma 5 auf das Binomial- und Poissonmodell anwendet. Seine optimale ψ Funktion $\tilde{\psi}_\theta$ minimiert im Allgemeinen die asymptotische Varianz $\mathrm{E}_\theta\, \psi_\theta^2/(\mathrm{E}_\theta\, \psi_\theta \Lambda_\theta)^2$ unter der Schranke $b = c/\,\mathrm{E}_\theta\, \tilde{\psi}_\theta \Lambda_\theta$ an die gross error Empfindlichkeit $\sup |\psi_\theta/\,\mathrm{E}_\theta\, \psi_\theta \Lambda_\theta|$ für beliebiges $c \in (0, \infty)$. Hampels Lösung ist von der selben Form wie unsere optimal-robusten Influenzkurven im Fall von infinitesimalen Kontaminationsumgebungen, die in Unterabschnitt 1.3.3 spezifiziert sind.

In seiner Behandlung des Binomial- und Poissonmodels, wie im Fall allgemeiner glatt parametrisierter Modelle, besitzt Hampel (1968) kein Kriterium für die Wahl der Empfindlichkeitsschranke b. Indem wir das MSE Problem betrachten, erhalten wir eine zusätzliche Gleichung mit deren Hilfe wir b in Abhängigkeit vom Startradius $r \in (0, \infty)$ der infinitesimalen Umgebungen auf eindeutige und optimale Weise bestimmen können; vergleiche Unterabschnitt 1.3.4.

Es gibt noch weitere Artikel über die robuste Schätzung in diskreten Modellen, welche jedoch nur spezielle Aspekte betrachten. Ruckstuhl and Welsh (2001) schlagen zum Beispiel einen robusten Schätzer vor, der einen hohen Bruchpunkt und zugleich eine beschränkte Influenzkurve besitzt. Simpson et al. (1987) beweisen die asymptotische Nicht-Normalität über Umgebungen von Hampel's optimalen M Schätzern, falls die zugrunde liegende Verteilung diskret ist. Das Resultat, eine Grenzverteilung, die aus zwei Normalverteilungen mit unterschiedlichen Streuungen zusammengesetzt ist, entspricht dem Resultat auf S. 78 von Huber (1964) bzw. auf S. 51 von Huber (1981). Außerdem wurde ein ähnliches Resultat über die asymptotische Nicht-Normalität des getrimmten Mittels von Stigler (1973) bewiesen. Als Ausweg schlagen Simpson et al. (1987) vor, $\tilde{\psi}_\theta$ durch eine glatte Approximation zu ersetzen, um die asymptotische Normalität beizubehalten.

Im Gegensatz dazu sind unsere optimal-robusten Influenzkurven die Lösungen zu gewissen Optimierungsproblemen, welche auf dem MSE Kriterium basieren. Die asymptotische Normalität unserer allgemeineren AL Schätzer auf vollen $1/\sqrt{n}$

Umgebungen wird durch die Glattheit des zugrundeliegenden parametrischen Modells sowie eine geeignete Schätzerkonstruktion erreicht. Am häufigsten werden in der Literatur robuste Schätzer mittels dem M Prinzip konstruiert. Wir ziehen es vor, die entsprechenden optimal-robusten Schätzer mit Hilfe der ein-Schritt Methode zu konstruieren, welche, bei Vorhandensein eines geeigneten Startschätzer, schneller zu berechnen ist und immer eine eindeutige Lösung liefert. Für mehr Einzelheiten über die ein-Schritt Konstruktion verweisen wir auf Abschnitt 6.4 von Rieder (1994) und Abschnitt 2.3.

In den Kapiteln 3 und 4 betrachten wir detailliert das Binomial- und Poissonmodell, wobei wir zu Beginn kurz die idealen Modelle einführen; siehe Abschnitte 3.1 und 4.1. In den Unterabschnitten 3.2.1.1 und 3.2.2.1 bzw. 4.2.1.1 und 4.2.2.1 spezifizieren wir die MSE optimalen Influenzkurven im Fall von Kontaminations- $(* = c)$ sowie Totalvariationsumgebungen $(* = v)$ und geben einige numerische Ergebnisse für den "lower case" Radius an, der in Unterabschnitt 2.1.2 eingeführt wurde.

Anschließend führen wir eine numerische Untersuchung von technischen Eigenschaften (Stetigkeit und Eindeutigkeit) der Lagrange Multiplikatoren, die in den optimalen Lösungen enthalten sind, durch. Diese Eigenschaften sind nützlich für: Die Bestimmung ungünstigster Radien (vgl. Abschnitt 2.2), die ein-Schritt Konstruktion (vgl. Abschnitt 2.3) und die Konvergenz robuster Modelle (vgl. Abschnitt 2.4); siehe Unterabschnitte 3.2.1.2 und 3.2.2.2 bzw. 4.2.1.2 und 4.2.2.2.

Zuerst untersuchen wir die Abhängigkeit vom Umgebungsradius r. Die numerischen Ergebnisse deuten darauf hin, dass die Standardisierungskonstante A_r glatt in r ist, hingegen können der standardisierte Bias b_r, die untere Stutzhöhe c_r $(* = v)$ und die asymptotische Varianz $A_r - r^2 b_r^2$ an einigen Werten von r nicht differenzierbar sein. Zusätzlich betrachten wir diejenigen Parameterwerte, für die $\mathrm{med}(\Lambda_\theta)$ nicht eindeutig ist. Als eine Folge von Proposition 2.1.3 ist die optimale Zentrierungskonstante a_r nicht eindeutig für $r \geq \bar{r}$. Genauer gesagt existiert ein ganzes Intervall von gültigen Zentrierungskonstanten für $r \geq \bar{r}$.

Als zweites behandeln wir die Stetigkeit bezüglich dem Parameter θ. Die numerischen Ergebnisse weisen darauf hin, dass die Standardisierungskonstante A_r, der standardisierte Bias b_r, die untere Stutzhöhe c_r $(* = v)$ und die asymptotische Varianz $A_r - r^2 b_r^2$ stetige, aber nicht notwendig glatte Funktionen in θ sind. Außerdem ist die Zentrierungskonstante a_r $(* = c)$ sogar unstetig für Radien $r \geq \bar{r}$ und solche Werte von θ, für die $\mathrm{med}(\Lambda_\theta)$ nicht eindeutig ist.

Diese numerischen Ergebnisse bestätigen die Stetigkeits- und Eindeutigkeitsresultate, die in den Unterabschnitten 2.1.4 und 2.1.5 hergeleitet wurden und deuten darauf hin, dass die Lagrange Multiplikatoren im Allgemeinen weder im Radius noch im Parameter glatte Funktionen sind.

Wir verwenden das Binomial- und Poissonmodell auch, um die Konvergenz robuster Modelle, die in Abschnitt 2.4 hergeleitet wurde, zu demonstrieren; vergleiche Unterabschnitte 3.2.1.3 und 3.2.2.3 bzw. Unterabschnitte 4.2.1.3, 4.2.2.3, 4.2.1.4 und 4.2.2.4.

Zu diesem Zweck geben wir einen Beweis dafür an, dass die geeignet standardisierten Lagrange Multiplikatoren im Fall des Binomial- und Poissonmodells gegen

die entsprechen Lagrange Multiplikatoren aus der eindimensionalen normalen Lokation konvergieren. Außerdem zeigen wir, dass die Lagrange Multiplikatoren im Fall des Poissonmodells durch die entsprechenden Lagrange Multiplikatoren aus dem Binomialmodell approximiert werden können.

Auf der Basis dieser Resultate können wir numerisch den "Abstand" zwischen der optimalen IC und ihrer Approximation in Termen der MSE–Ineffizienz berechnen. Im Fall von Kontaminationsumgebungen funktionieren diese Approximationen gut für kleine Radien ($r \leq 0.5$). Im Fall von Totalvariationsumgebungen schneiden diese Approximationen sogar noch viel besser ab und wir scheinen eine sehr gute Approximation unabhängig vom betrachteten Umgebungsradius zu erhalten.

In den Abschnitten 3.3 und 4.3 nehmen wir an, dass der Startradius der infinitesimalen Umgebungen unbekannt ist. Wir geben einige numerische Resultate für die ungünstigsten Radien und die entsprechenden MSE–Ineffizienzen im Fall des Binomial- und Poissonmodells an. In beiden Modellen und allen betrachteten Fällen bleibt der Effizienzverlust unter 30% und in den meisten Fällen ist er sogar deutlich kleiner.

Das Konstruktionsproblem im Fall des Binomial- und Poissonmodells ist in den Abschnitten 3.4 und 4.4 gelöst. Das heißt, wir verifizieren, indem wir die Resultate aus Unterabschnitt 2.3.3 anwenden, dass wir den optimal-robusten Schätzer mit Hilfe der ein-Schritt Methode konstruieren können. Insbesondere untersuchen wir diejenigen Parameterwerte, für welche die Zentrierungskonstante a_r $(* = c)$ im Fall $r \geq \bar{r}$ nicht eindeutig ist. Als Startschätzer schlagen wir den Kolmogorov(–Smirnov) minimum Distanz Schätzer vor, den wir auch implementiert haben (vgl. die generische `S4` Funktion `ksEstimator` in unserem `R` Paket `ROptEst`).

Die Implementation des Binomialmodells mit Hilfe von `S4` Klassen und Methoden (vgl. Chambers (1998)) unter Verwendung von `R` (vgl. R Development Core Team (2005)) ist detailliert in Abschnitt 3.5 beschrieben. Da die Implementation des Poissonmodells sehr ähnlich ist, geben wir in Abschnitt 4.5 nur eine sehr kurze Beschreibung. Beide Modelle sind in unserem `R` Paket `ROptEst` (vgl. Anhang D.3) eingeschlossen, welches Bestandteil unseres `R` bundle's `RobASt` ist.

Um die Notwendigkeit von robusten Schätzern in diesen zwei einfachen diskreten Modellen zu demonstrieren, führten wir einige kleine Simulationsstudien durch; siehe Abschnitte 3.6 und 4.6. Die Ergebnisse weisen darauf hin, dass der klassisch optimale Schätzer (Mittelwert) zu empfindlich ist und bereits sehr kleine Abweichungen vom idealen Modell zu sehr hohen Effizienzverlusten im Vergleich zu den optimal-robusten Schätzern führen können. Zusätzlich deuten die Resultate dieser Studien darauf hin, dass der Radius–minimax Schätzer ein gute Wahl darstellt, falls der tatsächliche Umgebungsradius unbekannt ist.

Exponentiale Skala und Gumbel Lokation

Hampel (1968) (vgl. Abschnitt F.1) diskutiert die robuste Schätzung im Fall des Exponentialmodells, wobei sich dieses Modell als ein wichtiger Spezialfall des Gammamodells ergibt. Er schlägt vor, ein getrimmtes Mittel zu verwenden und weist darauf hin, dass das getrimmte Mittel denselben Bruchpunkt wie das allgemein verwendete Winsorisierte Mittel besitzt (vgl. Feller (1971), Problem 17, S. 41), aber zusätzlich eine kleinere Sensitivität aufweist.

Gather and Schultze (1999) betrachten den standardisierten Median als robusten Schätzer für das exponentiale Skalenmodell. Sie zeigen (cf. Theorem 2.1, ibid.), dass dieser Schätzer am B-robustesten im Sinne von Hampel et al. (1986) ist; d.h., die minimale Sensivitität gegenüber gross errors besitzt. Zusätzlich führen Gather and Schultze (1999) zwei andere robuste Schätzer ein (RCS und Q Schätzer), die von Rousseeuw and Croux (1993) vorgeschlagen wurden. Alle drei Schätzer besitzen den höchst möglichen Bruchpunkt, der in diesem Setup bei 0.5 liegt. Jedoch sind ihre Biaskurven und asymptotischen relativen Effizienzen unterschiedlich.

Wie bereits oben angemerkt, sind unsere optimal-robusten Influenzkurven die Lösungen von wohlgestellten konvexen Optimierungsproblemen und wir erhalten die asymptotische Normalität unserer allgemeineren AL Schätzer auf vollen $1/\sqrt{n}$ Umgebungen durch die Glattheit des zugrunde liegenden parametrischen Modells und eine geeignete Schätzerkonstruktion. Außerdem können im Fall der ein-Schritt Konstruktion globale Eigenschaften wie Bruchpunkt auf den Startschätzer delegiert werden. Neben diesen (lokalen und globalen) Eigenschaften ist die Zielsetzung von Kapitel 5 eher der Zusammenhang zwischen Lokations- und Skalenmodellen als die Modelle selbst.

In Kapitel 5 zeigen wir, dass bestimmte Skalen- und Lokationsmodelle durch die Transformationen $\pm\log|\cdot|$ verbunden sind. Huber (1981) verwendet diese Tatsache bei seiner Behandlung des normalen Skalenmodells; siehe Abschnitt 5.6 (ibid.). Einen solchen Zusammenhang gibt es zum Beispiel auch zwischen dem exponentialen Skalen- und dem Gumbel Lokationsmodell.

Wir beginnen mit einer kurzen Einführung des eindimensionalen Skalen- und des eindimensionalen Lokationsmodells; siehe Unterabschnitte 5.1.1 und 5.1.2. Anschließend leiten wir die erwähnte Beziehung her (vgl. Unterabschnitt 5.1.3) und zeigen, dass diese zu einem engen Zusammenhang zwischen den Lagrange Multiplikatoren, die in den entsprechenden MSE optimalen ICs enthalten sind, führt. Um unsere Resultate zu demonstrieren, verwenden wir das exponentiale Skalenmodell, welches über die Transformation $-\log|\cdot|$ mit dem Gumbel Lokationsmodell in Beziehung steht.

Die optimal-robusten ICs für diese zwei Modelle sind in Abschnitt 5.2 sowohl für Kontaminations- $(*=c)$ als auch für Totalvariationsumgebungen $(*=v)$ spezifiziert. In beiden Fällen $(*=c,v)$ können die optimalen ICs so umgeschrieben werden, dass die darin enthaltenen Lagrange Multiplikatoren für beide Modelle identisch sind.

Als eine Folge dieses Zusammenfallens der Lagrange Multiplikatoren sind die ungünstigsten Radien und die entsprechenden MSE–Ineffizienzen, für beide Modelle identisch; siehe Abschnitt 5.3. Im Fall dass der Radius gänzlich unbekannt ist, beträgt der maximalen Effizienzverlust ungefähr 38% $(*=c)$ bzw. 22% $(*=v)$. Das bedeutet, dass der Verlust größer ist als im Fall der normalen Lokation bzw. der lognormalen Skala, bei der wir ca. 18% $(*=c,v)$ erhalten; vergleiche Bemerkung 5.1.9 (b). Aber er ist kleiner als im Fall der normalen Skala, bei der die Subeffizienz etwa 50% $(*=c)$ bzw. 25% $(*=v)$ beträgt; siehe Abschnitt 5.2 in Rieder et al. (2001).

Das Konstruktionsproblem für eindimensionale Lokations- bzw. Skalenmodelle

ist in Abschnitt 5.4 behandelt. Falls das betrachtete Lokations- bzw. Skalenmodell eine Exponentialfamilie bildet, können wir die optimal-robusten Schätzer mit Hilfe der ein-Schritt Methode konstruieren; vergleiche Lemma 2.3.6. Als Startschätzer schlagen wir den Kolmogorov(–Smirnov) minimum Distanz Schätzer vor, der die erforderlichen Eigenschaften (strikt und $\sqrt{n}$ konsistent) besitzt.

Eine kurze Beschreibung der Implementation von verschiedenen eindimensionalen Skalen- (exponential, normal, lognormal) bzw. eindimensionalen Lokationsmodellen (Gumbel, normal) ist in Abschnitt 5.5 gegeben. Alle diese Modelle sind in unserem `R` Paket `ROptEst` (vgl. Anhang D.3) eingeschlossen.

Gammamodell

In Abschnitt F.1 behandelt Hampel (1968) die robuste Schätzung im Gammamodell. Jedoch betrachtet er nur die Schätzung des Skalenparameters σ für bekannten Shapeparameter α bzw. die Schätzung des Shapeparameters α für bekannten Skalenparameter σ und nicht die simultane Schätzung von Skala und Shape.

Hampel et al. (1986) (Abschnitt 4.4, S. 256) betrachten die robuste Schätzung des Shapeparameters α, wobei die Skala σ als Nebenparameter angesehen wird. Anstelle von σ verwenden sie die Umparametrisierung $\nu = \log(\sigma)$, welche in Beispiel 1 von Unterabschnitt 4.3d (ibid.) eingeführt ist. Diese Umparametrisierung stattet das Gammamodell mit einer gewissen Invarianzstruktur aus; siehe Abschnitt 6.1.

Marazzi and Ruffieux (1996) diskutieren die Implementation der M Schätzer, die von Hampel et al. (1986) für das Gammamodell vorgeschlagen wurden. Sie arbeiten ebenfalls mit der Umparametrisierung ν. Zusätzlich betrachten sie die Parametrisierung $\kappa = \log(\alpha) + \nu$, da ihr Hauptinteresse die Schätzung des Erwartungswerts $\alpha\sigma = e^{\kappa}$ der Gammaverteilung ist.

Solche differenzierbaren Parametertransformationen mit Jacobi-Matrix von vollem Rang sind auch im Fall unserer optimalen Lösungen, die in Abschnitt 1.3 dargestellt sind, erlaubt. Wir verwenden das Gammamodell, um zu demonstrieren, wie man solche Transformationen in unserem Setup schätzen kann. Hinzu kommt, dass das Optimalitätsresultat, welches in Theorem 1.3.11 gegeben ist, deutlich stärker ist als die Optimalität, die durch Theorem 4.3.1 von Hampel et al. (1986) geliefert wird (vgl. auch die Diskussion vor Theorem 4.3.1, ibid.).

In Kapitel 6 führen wir zuerst kurz das Gammamodell als ideales Modell ein, wobei wir auch die oben zitierten Parametertransformationen berücksichtigen; siehe Abschnitt 6.1.

Die MSE optimale IC im Fall von Kontaminationsumgebungen ($* = c$) ist in Abschnitt 6.2 spezifiziert. Wir zeigen weiter, wie die Umparametrisierung $\nu = \log(\sigma)$ mit Hilfe von Theorem 2.4.1 auch in unserem Setup zu einer Vereinfachung führt. Jedoch im Gegensatz zu Abschnitt 4.4 in Marazzi and Ruffieux (1996), in dem die standardisierenden Matrizen für bijektive und differenzierbare Parametertransformationen immer mit Hilfe der entsprechenden Jacobi-Matrizen erhalten werden können, ist dies für die Lagrange Multiplikatoren, die in unseren MSE Lösungen enthalten sind, im Allgemeinen nicht möglich. Wir können mit Hilfe der entsprechenden Jacobi-Matrizen durchaus zulässige ICs definieren, aber diese ICs führen auf suboptimale robuste Schätzer, die einen ziemlich großen Effizienzverlust

($> 100\%$) aufweisen können; siehe Tabelle 6.1.

In Abschnitt 6.3 geben wir einige numerische Ergebnisse für die ungünstigsten Radien und die entsprechenden MSE–Ineffizienzen an. Im Fall dass der wahre Umgebungsradius vollständig unbekannt ist, liegen die maximalen Subeffizienzen in allen betrachteten Beispielen bei etwa 50%.

Da das Gammamodell eine Exponentialfamilie von vollem Rang bildet, können wir die Resultate aus Unterabschnitt 2.3.3 anwenden; vergleiche Abschnitt 6.4. Das heißt, wir können die optimal-robusten Schätzer mit Hilfe der ein-Schritt Methode konstruieren, wobei wir den Kolmogorov(–Smirnov) minimum Distanz Schätzer als Startschätzer verwenden.

Eine kurze Beschreibung der Implementation des Gammamodells wird in Abschnitt 6.5 gegeben. Wiederum lassen sich die entsprechenden optimal-robusten Schätzer mittels unseres `R` Pakets `ROptEst` (vgl. Anhang D.3) berechnen. Bis jetzt (Version 0.3-9) kann das Paket `ROptEst` verwendet werden, um MSE optimale ICs und Schätzer für beliebige L_2 differenzierbare parametrische Familien, die auf einer univariaten Verteilung basieren, zu berechnen.

Teil III: Robuste Regression und Skala

Die Behandlung robuster linearer Regression mit unbekannter Fehlerskala ist in den Theorien von Huber (1981) und Hampel et al. (1986) in verschiedener Hinsicht unvollständig, und dies gilt bereits für das einfachere Lokations- und Skalenmodell. Daher ist eine systematischere Untersuchung erforderlich.

Hubers (1981) Ansatz

Weder Hubers minimax Theorie für die asymptotische Varianz von Lokations-M-Schätzern (vgl. Huber (1964) bzw. Huber (1981)), welche auf der minimalen Fisher-Information und symmetrischer Fehlerverteilung sogar unter Kontamination basiert, noch Hubers minimax Intervallschätzer für endliche Stichproben (vgl. Huber (1968)), welcher auf robusten Tests und ungünstigsten Paaren basiert — beide Ansätze verwenden Umgebungen von fester Größe — wurden von Lokation auf gemeinsame Lokation und Skala erweitert.

Bereits für die Skala allein, im Falle eines symmetrisch kontaminierten Modells um eine Normalverteilung, bleibt der minimax Varianz Ansatz unvollständig, da der Sattelpunkt für die (relative) asymptotische Varianz der Skalen-M-Schätzer nur für Umgebungen mit einer maximalen Größe von 4% verifiziert werden konnte; siehe Abschnitt 5.7 in Huber (1981). Daneben besitzen die ungünstigsten Verteilungen, welche die Fisher-Information bei mehr als 20.5% Kontamination minimieren, "pathologische" Dichten mit einer Singularität in der Null; vergleiche Abschnitt 5.6 von Huber (1981).

Unter der Annahme, die Skala der Fehler ist bekannt, besitzt die Theorie von der minimax asymptotischen Varianz jedoch eine unmittelbare Erweiterung von Lokation auf lineare Regression, falls Regressor und Fehler auch unter Kontamination stochastisch unabhängig bleiben. Dies gilt, falls die Fehlerverteilung wie im Lokationsfall gestört wird, hingegen die Regressorverteilung fixiert (ideal) bleibt und

falls die Regressions-M-Schätzer aus Abschnitt 7.3 von Huber (1981) verwendet werden, welche die Residuen ohne Berücksichtigung des Regressors modifizieren.

Hubers minimax Intervallschätzer für endliche Stichproben und Lokation (vgl. Huber (1968)) wurde von Rieder (1989) auf einfache Regression (eindimensional, durch den Ursprung) erweitert, wobei wiederum die Skala der Fehler als bekannt vorausgesetzt ist. Diese Erweiterung wird ausführlich in Teil V dieser Dissertation untersucht.

Seine Definition von Schätzern für die gemeinsame Lokation und Skala beginnt Huber (1981) (vgl. Abschnitt 6.4, ibid.) mit einem Paar von M Gleichungen, wobei er verallgemeinerte Lokationsscores in die zwei Maximum Likelihood Schätzer (MLE) Gleichungen einführt. Er verallgemeinert dies, indem er dann in diesen zwei Gleichungen auf die Verbindung von Lokations- und Skalenscores verzichtet. Diese Unterscheidung wird in diesem Teil der Dissertation den Unterschied zwischen M und AL Schätzern ausmachen.

Neben der Bestimmung von Bruchpunkten in Abschnitt 6.6 (ibid.) verfolgt Huber (1981) keine quantitative, geschweige denn optimale, Robustheit von seinen Schätzern für die gemeinsame Lokation und Skala. Seine Beispiele bestehen einfach aus Kombinationen von Schätzern, welche separat optimal sind: Ein minimax Lokationsschätzer mit einem minimax (eingeschränkt, wie oben erwähnt) Skalenschätzer in Beispiel 4.1 (ibid.), dies erweitert den Vorschlag 2 von Huber (1964), und, in Beispiel 4.2 (ibid.), den Median zusammen mit dem Median der absoluten Abweichungen vom Median (MAD), wobei sich ersterer durch seinen minimax Bias im Fall reiner Lokation auszeichnet (vgl. Abschnitt 4.2 in Huber (1981)). Sogar im reinen Skalenproblem bleibt die entsprechende Eigenschaft des MAD, neben seinem Bruchpunkt von 50%, offen. In Abschnitt 6.5 (ibid.) und andernorts betrachtet Huber (1981) die Skala nachrangig zur Lokation und ruft die Symmetrie der Fehlerverteilung des Lokations- und Skalenmodells an, damit die Influenzkurve des Lokationsschätzers nicht vom Skalenschätzer abhängt, außer von dessen Grenzwert, wobei er einen beliebigen $\sqrt{n}$ -konsistenten asymptotisch linearen Schätzer für die Skala voraussetzt.

Hampels et al. (1986) Ansatz

Auf der einen Seite ist das Model tatsächlich durch die lokale und asymptotische, infinitesimale Robustheitstheorie von Hampel et al. (1986) und Rieder (1994) abgedeckt, da diese zwei verwandten Ansätze für einen allgemeinen (endlich dimensionalen, glatten) Parameter geeignet sind. Aber auf der anderen Seite wurden in diesen Setups die Ergebnisse für Lokation (Regression) und Skala bisher noch nicht sehr explizit aufgeschrieben.

Hampel et al. (1986) bestimmen die ICs von Gateaux-differenzierbaren Fisherkonsistenten Funktionalen für Lokation und Skala unter der Voraussetzung einer symmetrischen Fehlerverteilung, was ein Spezialfall robuster Adaptivität ist, wie diese in Teil IV dieser Dissertation definiert ist, und geben einige Beispiele an (vgl. Abschnitt 4.2.d, S. 232–237, ibid.). Das Modell von Lokation und Skala wird erneut herangezogen, um Modelle mit partitionierten Parametern einzuführen (vgl. Abschnitt 4.4.a, S. 253, ibid.). Unter der abermaligen Voraussetzung symmetrischer Fehler wird die Schätzung der Skala untergeordnet zur Lokation betrachtet, was aus

praktischer Sicht richtig sein könnte. Mit dem Verweis auf die Intuition und die Simulationsstudie von Andrews et al. (1972) wird empfohlen, einen robustesten Skalenschätzer (im Hinblick auf minimalem Bias bzw. maximalem Bruchpunkt?) zusammen mit einem robusten Lokationsschätzer zu verwenden.

Anschließend, in Abschnitt 4.4.b, S. 253–256, minimieren Hampel et al. (1986) jedes Diagonalelement (oder Block) der asymptotischen Kovarianz unter separaten Schranken in Supnorm an die entsprechenden Komponenten der ICs (4.4.b Theorem 1, S. 255, ibid.). Dabei wird keine Anordnung der Komponenten vorgenommen und kein Kriterium für die Wahl der Biasschranken angegeben. Wie im Fall von Lokation und Skala erwähnen die Autoren die Erfahrung, dass die Schranken an Lokations- und Skalenkomponente möglichst klein zu wählen sind (4.3.d Bemerkung 4, 4.4.b Bemerkung 4, ibid.).

In Rieder (1994) wird die robuste Lokation (Regression) und Skala überhaupt nicht erwähnt. Jedoch sollte beachtet werden, dass 4.4 Theorem 1 von Hampel et al. (1986) und 4.3 Theorem 1 (verwendet für 4.4 Theorem 1) in verschiedener Hinsicht durch Theorem 5.5.1, Bemerkung 5.5.4 und dem Paragraphen über "one-at-a-time" Optimalität (Ende von S.197) in Rieder (1994) verallgemeinert werden. Darüber hinaus kann die Sensitivitätsschranke gemäß dem MSE Kriterium als ein Funktion des (Start-)Radius $r \in (0, \infty)$ (der $r/\sqrt{n}$-Umgebungen, welche mit der Stichprobengröße n schrumpfen) bestimmt werden; siehe Rieder (1994), Abschnitt 5.5.2, Theorem 5.5.7.

Es verbleiben also einige Bemühungen, um die infinitesimale Robustheit für dieses Modell expliziter zu machen.

Unser Ansatz

In Kapitel 7 spezifizieren wir zuerst das ideale Modell; siehe Unterabschnitt 7.1.1. Zusätzlich zu der allgemeinen Klasse von AL Schätzern, die in Unterabschnitt 7.1.3.1 angegeben ist, führen wir verschiedene engere Klassen von M Schätzern ein; vergleiche Unterabschnitt 7.1.3.2. Die allgemeinen Regressions- und Skalen-M-Schätzer sind motiviert durch die Gleichungen (4.3) und (4.4) in Unterabschnitt 6.4 von Huber (1981), in dem er Lokation und Skala betrachtet. Als einen Spezialfall präsentieren wir M-Schätzer, die von Bednarski and Müller (2001) vorgeschlagen wurden und welche wir daher BM Schätzer nennen. Diese Schätzer sind eingeschränkt auf Regressorverteilungen mit endlichem Träger wie dies in dem von ihnen betrachteten Kontext von Versuchsdesigns üblich ist.

Als nächstes betrachten wir die Frage der Äquivarianz für diese Schätzerklassen; siehe Unterabschnitt 7.1.4. Im Gegensatz zur bisherigen Literatur ergibt sich Äquivarianz nicht durch eine Einschränkung an die Schätzer, sondern als eine Folge der Optimierungsprobleme.

In Kapitel 7 verwenden wir unbedingte ($t = 0$) bzw. gemittelte bedingte ($t = \alpha = 1$) Kontaminationsumgebungen ($* = c$).

In Abschnitt 7.2 leiten wir optimal-robuste ICs für die simultane Schätzung von Regression und Skala her. Zuerst betrachten wir die allgemeinen AL Schätzer; siehe Unterabschnitt 7.2.1. Die optimalen Lösungen werden durch Spezialisierung der konvexen Optimierungsresultate aus Kapitel 5 und 7 von Rieder (1994) erhalten. Jedoch treten neue Aspekte auf. Zum Beispiel ist die Regressionskoordinate der

optimalen IC aufgrund der zusätzlichen Skala “redescending”. In einem weiteren Schritt spezialisieren und vereinfachen wir die Lösungen, indem wir elliptische bzw. sphärische Symmetrie der idealen Regressorverteilung annehmen.

Als zweites betrachten wir in Unterabschnitt 7.2.2 die allgemeinen M Schätzer. Für diese Unterklasse ist das Optimierungsproblem für die simultane Schätzung von Regression und Skala nicht konvex. Daher lösen wir ein restringiertes Problem und benötigen eine zusätzliche äußere Optimierung, um die MSE Lösung zu erhalten; vergleiche Unterabschnitte 7.2.2.1 und 7.2.2.2. Das äußere Problem kann nur durch numerische Optimierung gelöst werden. Wir geben wiederum Spezialisierungen für elliptisch bzw. sphärisch symmetrische ideale Regressorverteilungen an.

In Abschnitt 7.3 leiten wir dann die optimal-robusten ICs für die separate Schätzung von Regression und Skala her. Dies bedeutet, dass wir zu einem Zeitpunkt die Regression schätzen wollen, wobei die Skala als Nebenparameter betrachtet wird, zu einem anderen Zeitpunkt hingegen sind wir an der Schätzung der Skala interessiert und die Regression ist nur ein Nebenparameter. Wir schließen dieses Problem mit ein, da es möglich ist, die BM Schätzer als Schätzer für die separate Schätzung von Regression und Skala zu motivieren.

Zuerst spezifizieren wir die AL Lösungen für die separate Schätzung; siehe Unterabschnitt 7.3.1. Aufgrund robuster Adaptivität, welche in Unterabschnitt 9.2.1 verifiziert ist, fällt die Lösung mit denen für die separaten Probleme zusammen.

Als zweites leiten wir die M Lösungen für die separate Schätzung von Regression und Skala her; vergleiche Unterabschnitt 7.3.2. In Analogie zu Bednarski and Müller (2001) betrachten wir nur M Schätzer mit bedingt zentrierten ICs und setzen als ideale Fehlerverteilung $F = \mathcal{N}(0,1)$ voraus. Wie im Fall der simultanen Schätzung ist das Optimierungsproblem nicht konvex und wir lösen ein restringiertes Problem. Das heißt, eine zusätzliche äußere Optimierung wird benötigt, um die MSE Lösung zu erhalten.

Als drittes behandeln wir die BM Schätzer; siehe Unterabschnitt 7.3.3. Wir entschieden uns, für die Herleitung der optimalen BM Lösung unsere eigenen Beweise anzugeben, da Bednarski and Müller (2001) eine innerer Punkt Voraussetzung benötigen, um die optimalen Lösungen herzuleiten und zudem nur unvollständige Lagrange Argumente angeben. Im Fall der BM Schätzer fallen die Probleme für die simultane und separate Schätzung von Regression und Skala zusammen (im Gegensatz zu AL und M Schätzern).

Ausgehend von den optimalen ICs müssen optimale Schätzer konstruiert werden. Wir betrachten dieses Problem für die optimalen AL Schätzer im Fall von linearer Regression mit unbedingten Kontaminationsumgebungen ($* = c,\ t = 0$); siehe Abschnitt 7.4. Aufgrund von Theorem 2.3.3 ist es möglich, abhängig von einem $\sqrt{n}$-konsistenten Startschätzers, den optimal-robusten Schätzer mit Hilfe der ein-Schritt Methode zu konstruieren. Jedoch gilt unser Resultat nur im Fall der idealen Fehlerverteilung $F = \mathcal{N}(0,1)$ und beschränkter Regressoren; vergleiche Unterabschnitte 7.4.1 und 7.4.2. Darüber hinaus betrachten wir nur die optimalen AL Schätzer für die simultane Schätzung von Regression und Skala.

Numerische Berechnungen werden bereits für die äußere Optimierungsschleife bei der Bestimmung der M Lösungen benötigt. Wir berechnen sowohl AL als auch M Lösungen in einigen einfachen Beispielen und bestimmen den Effizienzverlust

von M zu AL; siehe Abschnitt 7.5. Für die Zwecke dieser Einleitung wählen wir einige Effizienzverluste aus.

Zuerst betrachten wir die simultane Schätzung von Regression und Skala; vergleiche Unterabschnitt 7.5.1. Im Fall von unbedingten Kontaminationsumgebungen ($* = c,\ t = 0$), ist die Subeffizienz der allgemeinen M Schätzer in allen betrachteten Situationen klein ($< 10\%$). Jedoch im Fall von gemittelten bedingten Kontaminationsumgebungen ($* = c,\ t = \alpha = 1$) kann der Effizienzverlust der allgemeinen M Schätzer bezogen auf die optimalen AL Schätzer, welche wir ALc Schätzer nennen, recht groß werden. Die maximale Subeffizienz liegt in den betrachteten Beispielen bei etwa 300%. Im Fall der BM Schätzer steigt dieser Effizienzverlust weiter an und erreicht annähernd 425% bezogen auf die optimalen ALc Schätzer.

Als zweites geben wir einige Resultate für die separate Schätzung von Regression und Skala an; siehe Unterabschnitt 7.5.2. Wiederum arbeiten die optimalen AL Schätzer, welche wir in diesem Kontext ALs Schätzer nennen, viel besser als die Ms und BM Schätzer. Der maximale Effizienzverlust dieser Schätzer liegt bei ungefähr 315% bzw. 360%. In allen betrachteten Beispielen ist die Subeffizienz von BM bezogen auf Ms moderat und bleibt unter 15%.

Diese numerischen Vergleiche können mit Hilfe unserer `R` Pakete `ROptRegTS` und `RobRex` durchgeführt werde, welche Bestandteil unseres `R` bundle's `RobASt` sind; vergleiche Anhang D. Eine Beschreibung dieser Pakete findet sich in den Unterabschnitten 7.6.1 und 7.6.2. Das Paket `ROptRegTS` stellt eine Erweiterung unseres Pakets `ROptEst` auf Regressions-Typ Modelle, wie sie in Anhang A definiert sind, dar. Es verwendet Objektorientierung in Form von `S4` Klassen und Methoden; siehe Chambers (1998). Das Paket `RobRex` beinhaltet `R` Funktionen, welche dafür vorgesehen sind, die optimalen ICs von allen Schätzern zu berechnen, die im Verlauf dieses Kapitels betrachtet wurden.

Ein klarerer Vergleich zwischen diesen Schätzern ist im Fall von Lokation und Skala möglich, da keine Regressorverteilung ausgewählt werden muss. Zusätzlich berücksichtigen wir verschiedene bekannte robuste Schätzer, welche in der Literatur für Lokation und Skala vorgeschlagen wurden.

In Kapitel 8 vergleichen wir 18 verschiedene Schätzer für die robuste Schätzung von normaler Lokation und Skala.

Zuerst stellen wir das normale Lokations- und Skalenmodell in Verbindung mit infinitesimalen Kontaminationsumgebungen mit (Start-)Radius $r \in (0, \infty)$ vor; vergleiche Abschnitt 8.1. Die entsprechenden optimalen MSE Lösungen im Fall von AL, M und BM Schätzern sind in den Abschnitten 8.2–8.4 angegeben. Anschließend spezifizieren wir andere bekannte robuste Lokations- und Skalenschätzer, welche auf Vorschlägen von Huber (vgl. Unterabschnitt 8.5.1), Hampel (vgl. Unterabschnitt 8.5.2), Andrews (vgl. Unterabschnitt 8.5.3), Tukey (vgl. Unterabschnitt 8.5.4) und Yohai (vgl. Unterabschnitt 8.6) beruhen. Alle diese Schätzer sind asymptotisch linear.

In Abschnitt 8.1 führen wir zudem die Begriffe absolute und relative Information in Termen der Norm der IC ein. Diese Begriffe werden verwendet, um die verschiedenen Schätzer im Bezug auf die Menge an absoluter und relativer Information, welche sie mit einer gegebenen Beobachtung assoziieren, zu vergleichen.

Im Gegensatz zur klassisch optimalen IC, welche im Fall von Lokation und Skala unbeschränkt ist (d.h., ihre absolute Information ist unbeschränkt), erwarten wir von der IC eines robusten Schätzers, dass sie beschränkt ist (d.h., ihre absolute Information ist beschränkt). In der Tat gilt dies für alle Schätzer, die in diesem Kapitel betrachtet werden. Im Fall der relativen Information erhalten wir jedoch ein anderes Bild. Es zeigt sich, dass die relative Information der IC der optimalen AL Schätzer sehr ähnlich zur relativen Information der klassisch optimalen IC ist. Außerdem sind die Abweichungen im Fall der allgemeinen M Schätzer nur geringfügig größer. Im Fall der verbleibenden Schätzer sind die Abweichungen deutlich größer und in einigen Fällen besitzt die relative Information sogar eine völlig andere Form.

Wir ergänzen diese qualitativen Vergleiche durch weitere numerische Berechnungen. In Abschnitt 8.7 vergleichen wir den (numerischen) minimax asymptotischen MSE der betrachteten Schätzer und bestimmen den Effizienzverlust im Vergleich zu den optimalen AL Schätzern. Für die Zwecke dieser Einleitung erwähnen wir nur einige Effizienzvergleiche: Unter den übrigen Schätzern schneiden die allgemeinen M Schätzer am besten ab und verlieren nur wenige Promille Effizienz bezogen auf die optimalen AL Schätzer. Das Proposal 2 von Huber (1964) und der Schätzer, der in Beispiel 6.4.1 von Huber (1981) vorgeschlagen wird, funktionieren gut für kleine Startradien ($r \leq 0.5$) und besitzen Subeffizienzen von bis zu etwa 21% bzw. 12% im Grenzfall $r \to \infty$. Des Weiteren schneidet die Kombination von Hubers ψ-Funktion (vgl. Huber (1964)) bzw. von Hampels dreiteiliger "redescending" ψ-Funktion (vgl. Unterabschnitte 2C3, 3C3 von Andrews et al. (1972)) mit dem MAD, wie dies von Andrews et al. (1972) vorgeschlagen wurde, recht gut für große Startradien r ab. Deren Effizienzverluste im Fall $r \geq 1$ variieren zwischen ungefähr 10% und 12%.

Für eine gegebene optimale IC müssen wir den entsprechenden Schätzer konstruieren. Zumindest die optimalen AL Schätzer können mit Hilfe der ein-Schritt Methode erhalten werden. Dies ist in Abschnitt 8.8 mit Hilfe einer Anwendung von Lemma 2.3.6 nachgewiesen.

Für die numerische Bestimmung der MSE-optimalen ICs stellen wir unser `R` Paket `RobLox` zur Verfügung. Dieses Paket enthält `R` Funktionen für die Bestimmung der MSE-optimalen ICs von allen Schätzern, die im Verlauf diese Kapitels betrachtet werden; siehe Abschnitt 8.9. Außerdem können die optimalen ICs im Fall der AL Schätzer mit Hilfe unseres `R` Pakets `ROptEst` bestimmt werden. Beide `R` Pakete sind Teil unseres `R` bundle's `RobASt`.

Teil IV: Robuste Adaptivität

In seinem berühmten Artikel betrachtet Stein (1956) das Schätzen und Testen eines endlich dimensionalen euklidischen Parameters θ bei Vorhandensein eines unendlich dimensionalen Nebenparameters ν. Er leitet eine einfache notwendige Bedingung für Adaptivität her, nämlich die Diagonalform der Fisher-Information von eingebetteten endlich dimensionalen parametrischen Modellen. In Abhängigkeit von geeigneten Konstruktionen bedeutet klassische Adaptivität, dass das Schätzen (Testen) von θ bei unbekanntem ν asymptotisch nicht schwieriger ist als das

Schätzen (Testen) von θ bei bekanntem ν. Das Notwendigkeitsresultat von Stein (1956) wurde von Bickel (1982) aufgenommen, der hinreichende Bedingungen erhält unter denen adaptive Schätzer existieren (vgl. Theoreme 3.1 und 3.2, ibid.). Eine sehr ausführliche Behandlung von Adaptivität in semiparametrischen Modellen wird in Bickel et al. (1998) gegeben.

Da semiparametrische Modelle, die auf strikte Annahmen wie Symmetrie angewiesen sind, zu Umgebungsmodellen vergrößert werden können, ergibt sich die Frage der Adaptivität auch in der robusten Statistik. Da aber die klassische Scores in diesem Kontext nicht mehr optimal ist, muss man über die Bedeutung von robuster Adaptivität neu nachdenken.

Unserer Meinung nach ist es am überzeugendsten, die Definition robuster Adaptivität mit Hilfe des identischen Wertes zweier robuster Optimierungsprobleme auszudrücken. Mit dieser Definition ist Adaptivität nicht mehr länger nur ein dichotomes Kriterium, sondern besitzt, im Gegensatz zur bisherigen Literatur, nun auch eine quantitative Bedeutung. Allgemein gesprochen hängt die Tatsache, ob Adaptivität vorliegt oder nicht, nicht von einem gegebenen Startradius $r \in (0, \infty)$ ab. Jedoch kann sich die Situation für $r = 0$ (klassische Adaptivität) bzw. für den Grenzfall $r \to \infty$ von $r \in (0, \infty)$ unterscheiden.

In Abschnitt 9.1 definieren wir Adaption mit Hilfe zweier asymptotischer MSE Probleme. Das heißt, durch die Betrachtung der MSE–Ineffizienzen zwischen den entsprechenden Lösungen erhalten wir einen Ausdruck für die Größe der Nicht-Adaptivität.

In dieser Dissertation beschränken wir uns auf endlich dimensionale Parameter, jedoch kann der Begriff der robusten Adaptivität leicht auf Umgebungsmodelle mit unendlich dimensionalen Parametern ausgedehnt werden; siehe Abschnitt 6.1 von Rieder (2003) bzw. Abschnitt 2 von Shen (1995) (implizit verwendet).

Im aktuellen Kapitel treten verschiedene Kombinationen von klassischer und robuster Adaptivität auf. Zum einen gibt es Modelle, welche sowohl klassisch als auch robust adaptiv sind. Zum anderen geben wir Beispiele, in denen wir klassische aber keine robuste Adaptivität haben und schließlich behandeln wir Modelle, die weder klassisch noch robust adaptiv sind. Unser Studium der Adaptivität wird durch numerische Auswertungen der Größe der Nicht-Adaptivität unterstützt.

In Abschnitt 9.2 betrachten wir das lineare Modell mit zufälligen Regressoren, wobei wir zusätzlich Umgebungen um das ideale Modell betrachten. Im Einzelnen sind dies unbedingte $(* = c,\ t = 0)$, gemittelte bedingte $(* = c,\ t = \alpha = 1)$ und quadratisch gemittelte bedingte $(* = c,\ t = \alpha = 2)$ Kontaminationsumgebungen sowie gemittelte bedingte Totalvariationsumgebungen $(* = v,\ t = \alpha = 1)$.

Zuerst untersuchen wir robuste Adaptivität im Fall von linearer Regression mit Skala; vergleiche Unterabschnitt 9.2.1. Unter der Voraussetzung einer symmetrischen idealen Fehlerverteilung F, ist dieses Modell klassisch adaptiv bezogen auf die Skala. Falls wir unseren Blickwinkel ändern und Skala als Hauptparameter und den Regressionsparameter als Nebenparameter betrachten, erhalten wir erneut klassische Adaptivität. Nun wird das ideale Modell um die oben erwähnten Umgebungen erweitert. Aufgrund der Symmetrie der idealen Fehlerverteilung F bleibt die Adaptivität bezogen auf die Skala tatsächlich auch unter den Umgebungen er-

halten.

Im Fall von unbedingten ($* = c,\ t = 0$) und gemittelten bedingten ($* = c,\ t = \alpha = 1$) Kontaminationsumgebungen werden die Lösungen für die simultane Schätzung von Regression und Skala in Unterabschnitt 7.2.1 hergeleitet. Mit Hilfe dieser Lösungen können wir auch die robuste Adaptivität im Fall, dass die Skala den Hauptparameter und der Regressionsparameter den Nebenparameter darstellt, untersuchen. Es stellt sich heraus, dass das lineare Modell ausgestattet mit diesen Umgebungen robust adaptiv mit Bezug auf den Regressionsparameter ist.

Als zweites behandeln wir Regression mit Achsenabschnitt, wobei wir wieder eine symmetrische ideale Fehlerverteilung F voraussetzen; vergleiche Unterabschnitt 9.2.2. Zusätzlich nehmen wir eine asymmetrische ideale Regressorverteilung K an. Unter diesen Voraussetzungen ist das lineare Modell klassisch aber nicht notwendigerweise robust adaptiv bezogen auf den Achsenabschnitt. Im Fall quadratisch gemittelter bedingter Kontaminationsumgebungen ($* = c,\ t = \alpha = 2$) erhalten wir tatsächlich robuste Adaptivität mit Bezug auf den Achsenabschnitt. Jedoch trifft dies nicht in Verbindung mit den größeren unbedingten ($* = c,\ t = 0$) und gemittelten bedingten Kontaminationsumgebungen ($* = c,\ t = \alpha = 1$) sowie den gemittelten Totalvariationsumgebungen ($* = v,\ t = \alpha = 1$) zu.

In Unterabschnitt 9.2.2.5 geben wir einige numerische Resultate für die Größe der Nicht-Adaptivität im Fall des linearen Modells mit gemittelten bedingten Kontaminationsumgebungen ($* = c,\ t = \alpha = 1$) an. Für diese Berechnungen setzen wir als ideal Fehlerverteilung $F = \mathcal{N}(0, 1)$ voraus und betrachten einige einfache eindimensionale Regressorverteilungen K. Es zeigt sich, dass bereits in diesen beispielhaften Situationen das Ausmaß der Nicht-Adaptivität recht groß werden kann. Wir erhalten Subeffizienzen von bis zu 300%. Außerdem zeigen die numerischen Resultate, dass die Größe der Nicht-Adaptivität nicht notwendigerweise monoton wachsend im Startradius r ist. In einem Beispiel erhalten wir sogar robuste Adaptivität im Grenzfall $r \to \infty$; vergleiche Beispiel 9.2.26.

Diese Berechnungen sowie die numerische Bestimmung der optimalen ICs im Fall von linearer Regression können mit Hilfe unseres `R` Pakets `ROptRegTS` durchgeführt werden, welches Bestandteil unseres `R` bundle's `RobASt` ist; siehe Anhang D. In Unterabschnitt 9.2.3 geben wir eine kurze Beschreibung dieses Pakets. Die Implementation basiert auf `S4` Klassen und Methoden wie sie in Chambers (1998) eingeführt wurden. Weitere Details zur Implementation von linearen Regressionsmodellen können im Unterabschnitt 7.6.1 bzw. in Anhang D.4 gefunden werden.

In Abschnitt 9.3 untersuchen wir die Adaptivität von Zeitreihenmodellen. Wie in Anhang A zu sehen ist, können einige dieser Modelle, welche eine bestimmte Regressionsstruktur aufweisen, analog zu linearen Modellen behandelt werden. Wichtige Beispiele sind das ARMA(p, q) und das ARCH(p) Modell, welche wir in Verbindung mit gemittelten ($* = c,\ t = \alpha = 1$) und quadratisch gemittelten ($* = c,\ t = \alpha = 2$) Kontaminationsumgebungen der Übergangswahrscheinlichkeiten sowie gemittelten Totalvariationsumgebungen ($* = v,\ t = \alpha = 1$) der Übergangswahrscheinlichkeiten untersuchen.

Zuerst betrachten wir das ARMA(p, q) Modell inklusive Verschiebung; siehe

Unterabschnitt 9.3.1. In der Tat besitzt dieses Modell ein ähnliche Struktur wie das lineare Modell mit Achsenabschnitt, welches in Unterabschnitt 9.2.2 behandelt wird.

Unter der Voraussetzung, dass die ideale Verteilung der Innovationen F den Erwartungswert $\mu_F = 0$ besitzt, ist das ARMA(p, q) Modell klassisch adaptiv mit Bezug auf die Verschiebung. Die Adaptivität überträgt sich auf Umgebungen im Fall von quadratisch gemittelten Kontaminationsumgebungen $(* = c,\ t = \alpha = 2)$; vergleiche Unterabschnitt 9.3.1.1. Im Fall der größeren gemittelten Kontaminations- $(* = c,\ t = \alpha = 1)$ und Totalvariations- $(* = v,\ t = \alpha = 1)$ Umgebungen benötigen wir zusätzliche Voraussetzungen, um robuste Adaptivität mit Bezug auf die Verschiebung zu erhalten. Falls wir zusätzlich eine symmetrische ideale Innovationsverteilung F voraussetzen, überträgt sich die robuste Adaptivität tatsächlich; siehe Unterabschnitt 9.3.1.2 bzw. 9.3.1.3.

Anschließend ändern wir unseren Blickwinkel und betrachten die Verschiebung als Hauptparameter während die ARMA Parameter die Nebenparameter sind. Dieses Modell kann als eine Verallgemeinerung des klassischen u.i.v. Lokationsmodells angesehen werden, wobei die Fehler nun zusätzlich eine ARMA Struktur besitzen. Wiederum erhalten wir klassische Adaptivität als eine Konsequenz von $\mu_F = 0$. Außerdem ergibt sich robuste Adaptivität im Fall von quadratisch gemittelten bedingten Kontaminationsumgebungen $(* = c,\ t = \alpha = 2)$. Aber dieses Mal liegt robuste Adaptivität (ohne weitere Voraussetzungen) auch im Fall der größeren gemittelten bedingten Kontaminations- $(* = c,\ t = \alpha = 1)$ und Totalvariationsumgebungen $(* = v,\ t = \alpha = 1)$ vor.

Der Umfang der robusten Nicht-Adaptivität wird im Fall von AR(1) bzw. MA(1) mit Verschiebung und Gumbel verteilten Innovationen ausgewertet. Wie wir sehen, ist der Umfang in Termen des MSE Effizienzverlustes nur sehr klein. Für alle betrachteten Parameterwerte bleibt die Subeffizienz jeweils unterhalb von 3%; vergleiche Unterabschnitt 9.3.3.1.

Als zweites untersuchen wir die robuste Adaptivität des ARCH(p) Modells, welches von Engle (1982) eingeführt wurde; siehe Unterabschnitt 9.3.2. Wieder betrachten wir gemittelte $(* = c,\ t = \alpha = 1)$ und quadratisch gemittelte $(* = c,\ t = \alpha = 2)$ Kontaminationsumgebungen von Übergangswahrscheinlichkeiten sowie gemittelte Totalvariationsumgebungen $(* = v,\ t = \alpha = 1)$ von Übergangswahrscheinlichkeiten. Dieses Modell ist nicht klassisch adaptiv mit Bezug auf die Skala der Innovationen. Außerdem ist es in Verbindung mit den genannten Umgebungen auch nicht robust adaptiv bezogen auf die Skala der Innovationen.

Numerische Resultate im Fall des ARCH(1) Modells mit lognormal verteilten Innovationen sind in Unterabschnitt 9.3.3.2 angegeben. Wie sich herausstellt, kann der Umfang der Nicht-Adaptivität in diesem Setup sehr groß sein. Insbesondere falls wir Parameter auswählen, die in der Nähe der Grenze des Stationaritätsbereichs des betrachteten ARCH(1) Prozesses liegen, erhalten wir wirklich enorme Subeffizienzen $(> 1000\%)$; siehe Tabelle 9.8.

Im Fall der Zeitreihenmodelle sind bis jetzt nur die Spezialfälle, die in Unterabschnitt 9.3.3 betrachtet wurden, implementiert, wobei wir keine Objektorientierung verwenden. Die Implementation von AR(1) und ARCH(1) ist im Unterabschnitt 9.3.4 beschrieben. Jedoch passen diese speziellen Regressions-Typ Zeitrei-

henmodelle gut in den objektorientierten Aufbau unseres `R` Paketes `ROptRegTS` (vgl. Anhang D.4) und werden in naher Zukunft darin integriert werden.

Teil V: Finiter versus Asymptotischer Schätzer

In diesem letzten Teil kontrollieren wir die Asymptotik anhand von Resultaten für finite Stichproben.

In Fall eindimensionaler Lokation bzw. einfacher Regression gibt es von Huber (1968) bzw. Rieder (1989) eine exakte finite[4] minimax Theorie, die auf robusten Tests für spezielle Kapazitäten basiert. Dieser Ansatz erfordert eine spezielle Pseudo-Verlustfunktion in Gestalt von Unter-/Überschusswahrscheinlichkeiten und liefert minimax Optimalität unter beliebigen Schätzern. Jedoch scheint er auf einen reellen Lokations- bzw. Regressionsparameter beschränkt zu sein. Der entsprechende asymptotische minimax Schätzer für diese spezielle Pseudo-Verlustfunktion wurde in Rieder (1980) für die Klasse asymptotisch linearer Schätzer und in Rieder (1981a) für die Klasse beliebiger Schätzer hergeleitet. Daher kann die Asymptotik anhand von Resultaten für finite Stichproben und Kontaminations- sowie Totalvariationsumgebungen fester Größe überprüft werden. Der Vergleich zwischen asymptotischen und finiten Resultaten erfordert jedoch die Berechnung der exakten finiten Risiken. Die analytische Berechnung finiter Risiken erweist sich als sehr schwer oder sogar unmöglich für Stichprobenumfänge $n \geq 3$. Daher haben wir einen Algorithmus konzipiert, der auf der schnellen Fourier Transformation (FFT) beruht, und mit dem die exakte finite Verteilung (und die entsprechenden finiten Risiken) dieser in unterschiedlicher Weise robusten Schätzer bestimmt werden kann.

Zwei interessante Ergebnisse sind: Die (erste Ordnungs-) Asymptotik ist zu optimistisch und die Konvergenz gegen die asymptotischen Werte ist besser im Fall der Totalvariationsumgebungen als im Fall der Kontaminationsumgebungen.

Wir geben nun einen kurzen Überblick über die finite und die asymptotische minimax Theorie.

Hubers (1968) Ansatz

Huber (1968) betrachtet eindimensionale Lokation und setzt Kontaminations-/Totalvariationsumgebungen fester Größe ein. Auf der Basis einer minimax Testtheorie für spezielle Kapazitäten, die in Huber (1965) entwickelt wurde, leitet er eine finite minimax Lösung her, welche die maximale Wahrscheinlichkeit, dass der Schätzer den wahren Wert des Parameters um mehr als eine feste Konstante übersteigt bzw. unterschreitet, minimiert. Wir nennen diese Pseudo-Verlustfunktion *Unter-/Überschuss Konfidenzrisiko*. Eine Zusammenfassung dieser Resultate findet sich auch in Kapitel 10 von Huber (1981).
Im Fall eindimensionaler normaler Lokation und Stichprobenumfang $n = 2$ ist Huber (1964) in der Lage, das finite Unter-/Überschuss Konfidenzrisiko analytisch zu bestimmen. Seine Ergebnisse für die normale Lokation werden durch unsere analytischen und numerischen Berechnungen, die in den Abschnitten 11.3 und 11.4 angegeben sind, bestätigt und erweitert.

[4] d.h., im Kontext finiter Stichproben

Rieders (1989) Ansatz

Rieder (1989) dehnt die Resultate von Huber (1968) auf die einfache lineare Regression durch den Ursprung aus, wobei der Regressor zufällig ist und er zwei Arten von fixierten Umgebungen einsetzt: unbedingte ("Fehler in den Variablen") und bedingte ("fehlerfreie Variablen") Kontaminations-/Totalvariationsumgebungen. Für den Spezialfall von endlich vielen deterministischen Regressoren und bedingten Umgebungen wurde eine vorläufige Erweiterung von Huber (1983) angedeutet.
Im Fall normaler Fehler und unbedingter Umgebungen entspricht der finite minimax Schätzer dem Hampel-Krasker Schätzer. Dies widerspricht der Vermutung von Huber (1983) (vgl. Abschnitte 1, 7 und Erwiderung), dass der Hampel-Krasker Schätzer keine Optimalität besitzt in dem Fall, dass "Fehler in den Variablen" vorliegen. Außerdem treten im Fall von normalen Fehlern und bedingten Umgebungen sowohl der Hampel-Kraker als auch der Huber Schätzer als minimax Lösungen für geeignet gewählte Kontaminationskurven auf.

Rieders (1980) und (1981) Ansatz

Die asymptotische Entsprechung der Resultate für finite Stichprobenumfänge, die oben angegeben sind, wurde in Rieder (1980) unter Verwendung von Resultaten aus der robusten asymptotischen Testtheorie (vgl. Rieder (1978)) hergeleitet. Um diese Resultate zu erhalten, wird vorausgesetzt, dass die Größe der Kontaminations-/Totalvariationsumgebungen sowie die Breite des betrachteten Konfidenzintervalls mit der Rate $\sqrt{n}$ schrumpfen.
Im Gegensatz zu den finiten Resultaten sind die Resultate, die in Rieder (1980) angegeben sind, auf einen beliebigen Parameter anwendbar. Insbesondere lassen sich diese auf lineare Regression mit zufälligen Regressoren und unbedingten Kontaminations-/Totalvariationsumgebungen anwenden.
Indem man das asymptotische Unter-/Überschuss Konfidenzrisiko uniform auswertet, werden supereffiziente Schätzer nicht aufgrund von speziellen Annahmen, sondern durch ein hohes maximales Risiko ausgeschlossen. Auf diese Weise etabliert Rieder (1981a) eine lokale asymptotische minimax Schranke für beliebige Schätzer. In diesem Sinne ist der asymptotische minimax Schätzer, der in Rieder (1980) hergeleitet ist, minimax unter allen beliebigen Schätzern.
Im Fall von bedingten Kontaminationsumgebungen verifiziert Rieder (1994) (vgl. Unterabschnitt 7.5.2), dass der Huber Schätzer im Vergleich zum Hampel-Krasker Schätzer in einem ähnlichen aber etwas kleinerem Modell minimax ist. Nämlich im Fall von quadratisch gemittelten im Gegensatz zu gemittelten bedingten Umgebungen (vgl. Bemerkung 7.5.17 (a), ibid.). Da die Minimierung des asymptotischen Unter-/Überschuss Konfidenzrisikos äquivalent zur Minimierung der asymptotischen Varianz unter einer Biasschranke ist (vgl. Lemma 10.3.5), ist dieses Resultat auch in unserem Setup gültig.

Unser Ansatz

In Kapitel 10 führen wir zuerst den finiten und den asymptotischen Setup ein; siehe Abschnitt 10.1. Als zweites präsentieren wir die Herleitung des finiten minimax Regressionsschätzers, welche zum größten Teil auf Rieder (1989) und Abschnitt 1 in Rieder (1995) basiert; vergleiche Abschnitt 10.2. In Ergänzung zu

Rieder (1989), beweisen wir die Zulässigkeit des finiten minimax Schätzers (vgl. Theorem 10.2.10 (b)) und geben einige formale Argumente für die Herleitung der Schätzer in den Randfällen an (vgl. Unterabschnitt 10.2.4).

Als drittes wenden wir die asymptotischen Resultate von Rieder (1980) auf lineare Regression an und erweitern die asymptotische Theorie auf bedingte Regressionsumgebungen; siehe Abschnitt 10.3. Wir leiten den asymptotischen minimax Schätzer ohne jede Verbindung zum entsprechenden robusten Testproblem von Rieder (1978) her, indem wir stattdessen Optimierungsargumente wie in Ruckdeschel and Rieder (2004) verwenden; vergleiche Unterabschnitt 10.3.3.

Wie oben erwähnt, sind diese Resultate die asymptotischen Gegenstücke zum Fall finiter Stichproben. Daher sind wir in der Lage, die Asymptotik mit Hilfe von Resultaten, die für finite Stichproben und fixierte Umgebungen hergeleitet wurden, zu überprüfen. Dies gilt zumindest für den Fall dieser speziellen Pseudo-Verlustfunktion und eines eindimensionalen Lokations- bzw. Regressionsparameters. Vorausgesetzt natürlich, dass wir in der Lage sind, die finiten Risiken mit hoher Genauigkeit numerisch zu berechnen.

In Kapitel 11 spezialisieren wir die Resultate aus Kapitel 10 von einfacher Regression auf eindimensionale Lokation und von Fehlerverteilungen mit finiter Fisher-Information auf normalverteilte Fehler. Wir betrachten Kontaminations- ($* = c$) sowie Totalvariationsumgebungen ($* = v$); siehe Abschnitt 11.1 bzw. 11.2. In diesem Setup sind der finite und der asymptotische minimax Schätzer von der gleichen Gestalt und können anhand der entsprechenden optimalen Stutzhöhen identifiziert werden. Wir vergleichen daher zuerst die finit und die asymptotisch optimalen Stutzhöhen unter Verwendung von Taylorentwicklungen. Es zeigt sich, dass die Konvergenz gegen die asymptotischen Werte von unten erfolgt; d.h., die (erste Ordnungs-) Asymptotik ist zu optimistisch. Außerdem ist die Konvergenzgeschwindigkeit im Fall von Kontaminationsumgebungen ($* = c$) von der Ordnung $n^{-1/2}$ während sie im Fall von Totalvariationsumgebungen ($* = v$) von der Ordnung n^{-1} ist. Deshalb besteht im Fall der Stutzhöhen ein deutlicher Unterschied zwischen diesen zwei Typen von Umgebungen, der unserer Vermutung nach von der größeren Symmetrie der Totalvariationsumgebungen verursacht wird. Im Verlauf dieser Untersuchungen, erhalten wir auch $\mathrm{O}(n^{-1/2})$-korrigierte ($* = c$) bzw. $\mathrm{O}(n^{-1})$-korrigierte ($* = v$) asymptotisch optimale Stutzhöhen.

In Abschnitt 11.3 untersuchen wir direkt die exakten finiten Risiken (vgl. Unterabschnitt 11.3.1) und führen zwei Algorithmen ein, die auf der schnellen Fourier Transformation (FFT) beruhen (vgl. Unterabschnitt 11.3.2). Diese stellen numerische Approximationen zur Verfügung, die sehr genau sind. Wir überprüfen die Genauigkeit dieser Algorithmen mittels analytischer Berechnungen, die für Stichprobenumfang $n = 2$ verfügbar sind (der einfachere Fall $n = 1$ wird nicht betrachtet!) sowie mittels numerischer Simulationen; siehe Unterabschnitt 11.3.2.3. Wir verwenden diese Algorithmen auch dazu, die Resultate von anderen Approximationen, nämlich von Edgeworth Entwicklungen bis zur ersten und zweiten Ordnung (vgl. Ibragimov (1967)) sowie von Sattelpunktapproximationen (vgl. Field and Ronchetti (1990)), zu überprüfen; siehe Unterabschnitt 11.3.3. Es stellt sich heraus, dass in diesem Konfidenz-Setup für robuste normale Lokation sowohl die

Edgeworth Entwicklungen bis zur zweiten Ordnung als auch die Sattelpunktsapproximationen gute Näherungen bis hinunter zu Stichproben mit 5 Beobachtungen liefern.

Die Hauptanwendung unserer Algorithmen ist der numerische Vergleich zwischen finiten und asymptotischen Resultaten; siehe Abschnitt 11.4. Zuerst vergleichen wir die finit, die asymptotisch und die O-korrigierten asymptotisch optimalen Stutzhöhen; siehe Unterabschnitt 11.4.1.1 bzw. 11.4.2.1. Es stellt sich heraus, dass es deutliche Unterschiede zwischen den finit und den asymptotisch optimalen Stutzhöhen gibt. Jedoch liegen die O-korrigierten asymptotisch optimalen Stutzhöhen für mittelgroße Stichproben mit ungefähr 10 ($* = v$) bzw. 20 ($* = c$) Beobachtungen bereits sehr nahe bei den finit optimalen.

Als zweites vergleichen wir die finiten Risiken des finiten und des asymptotischen minimax Schätzers sowie des Schätzers, der auf der O-korrigierten asymptotisch optimalen Stutzhöhe basiert; siehe Unterabschnitt 11.4.1.2 bzw. 11.4.2.2. Obwohl es deutliche Unterschiede zwischen den optimalen Stutzhöhen gibt, sind die Unterschiede (in absoluten Zahlen) zwischen den entsprechenden finiten Risiken nur klein. Zusätzlich deuten die Resultate der Box-Cox Potenztransformation, die vom `MASS` Paket von Venables and Ripley (2002) zur Verfügung gestellt wird, darauf hin, dass die Konvergenzgeschwindigkeit der finiten Risiken gegen die asymptotischen Risiken von derselben Ordnung ist wie die der optimalen Stutzhöhen. Das heißt, im Fall von Kontaminationsumgebungen ($* = c$) scheint die Konvergenzgeschwindigkeit von der Ordnung $n^{-1/2}$ hingegen im Fall von Totalvariationsumgebungen ($* = v$) von der Ordnung n^{-1} zu sein.

Da es (bereits a priori) nur kleine Unterschiede zwischen den absoluten Werten gibt, untersuchen wir auch relative Werte; d.h., wie viel Effizienz verlieren wir, falls wir die asymptotisch optimale bzw. die O-korrigierte asymptotisch optimale Stutzhöhe anstelle der finit optimalen Stutzhöhe verwenden. Die numerischen Ergebnisse zeigen, dass die maximalen relativen Risiken des Schätzers, der auf der O-korrigierten asymptotisch optimalen Stutzhöhe basiert, sehr klein sind. Außerdem gibt es bereits für mittelgroße Stichproben ($* = c$: $20 - 50$, $* = v$: 10) keine großen Unterschied mehr zwischen allen drei Schätzern (maximaler Effizienzverlust $< 5\%$).

Da die Größe der Kontamination in den meisten Anwendungen unbekannt bzw. bis auf ein Intervall unbekannt ist, bestimmen wir auch die ungünstigsten Radien und die entsprechenden Ineffizienzen. Diese sind analog zu Abschnitt 2.2 definiert. Die Effizienzverluste bleiben in allen betrachteten Fällen deutlich unter 40% ($\rho = 0$), 20% ($\rho = 1/2$) und 10% ($\rho = 1/3$), wobei die größten Ineffizienzen im asymptotischen Fall auftreten. Da die Normalverteilung symmetrisch zur Null ist, sind im asymptotischen Setup die Subeffizienzen für Kontaminations- ($* = c$) und Totalvariationsumgebungen ($* = v$) identisch. Im finiten Setup gibt es jedoch Unterschiede und die Ineffizienzen sind im Fall von Totalvariationsumgebungen etwas größer. Dies ist vermutlich durch die schnellere Konvergenz gegen die asymptotischen Werte verursacht.

Als drittes berechnen wir die finite Verteilung des finiten minimax Schätzers und vergleichen diese mit der Normalverteilung, die den kleinsten Kolmogorov Abstand d_κ hat; siehe Unterabschnitte 11.4.1.3 und 11.4.2.3. Unsere numerischen

Ergebnisse deuten darauf hin, dass die Konvergenzgeschwindigkeit im Fall von Totalvariationsumgebungen ($* = v$) nicht nur in der Null (entspricht dem finiten Risiko), sondern uniform über den ganzen Träger der finiten Verteilung des finiten minimax Schätzers schneller ist. Die Unterschiede treten aber in erster Linie für sehr kleine Stichprobenumfänge zu Tage und bereits für $n \geq 10$ erhalten wir in beiden Fällen $d_\kappa < 0.02$.

In Kapitel 12 spezialisieren wir die Resultate aus Kapitel 10 von beliebigen Fehlern mit endlicher Fisher-Information auf normalverteilte Fehler. Wir betrachten unbedingte ("Fehler in den Variablen") Kontaminiatons- ($* = c, t = 0$) und Totalvariationsumgebungen ($* = v, t = 0$) sowie bedingte ("fehlerfreie Variablen") Kontaminations- ($* = c, t = \varepsilon$) und Totalvariationsumgebungen ($* = v, t = \delta$); vergleiche Abschnitt 12.1 bzw. 12.2.

Ähnlich zur normalen Lokation besitzen der finite und der asymptotische minimax Schätzer dieselbe Gestalt und können anhand der entsprechenden optimalen Stutzhöhen ($* = c, v; t = 0$) bzw. Funktionen ($* = c, v; t = \varepsilon, \delta$) identifiziert werden. Daher vergleichen wir zuerst die optimalen Stutzhöhen bzw. Funktionen. Wie sich herausstellt, sind die Ergebnisse analog zur normalen Lokation. Das heißt, wir erhalten mit Hilfe von Taylorentwicklungen, dass die Konvergenzgeschwindigkeit im Fall von Kontaminationsumgebungen ($* = c; t = 0, \varepsilon$) von der Ordnung $n^{-1/2}$, im Fall von Totalvariationsumgebungen ($* = c; t = 0, \varepsilon$) hingegen von der Ordnung n^{-1} ist. In allen Fällen erfolgt die Konvergenz gegen die asymptotische Werte von unten; d.h., die (erste Ordnungs-) Asymptotik ist zu optimistisch. Darüber hinaus liefern diese Taylorentwicklungen $\mathrm{O}(n^{-1/2})$-korrigierte ($* = c; t = 0, \varepsilon$) und $\mathrm{O}(n^{-1})$-korrigierte ($* = v; t = 0, \delta$) asymptotisch optimale Stutzhöhen bzw. Funktionen.

Außerdem leiten wir erneut Algorithmen für die Berechnung des finiten Risikos her, welche auf FFT basieren und ähnlich zu den Algorithmen sind, die im Fall der eindimensionalen normalen Lokation verwendet wurden; siehe Unterabschnitte 12.1.3 und 12.2.3. Aufgrund der Abhängigkeit vom Regressor haben wir dieses Mal, im Gegensatz zur normalen Lokation, keine analytischen Ergebnisse für Stichprobenumfang $n = 2$ mit deren Hilfe wir die Genauigkeit unserer Algorithmen überprüfen könnten. Jedoch deuten verschiedene Gegenproben und Ergebnisse, die wir mittels numerischen Simulationen erhielten, darauf hin, dass unsere Algorithmen wieder sehr genau sind.

In Abschnitt 12.3 stellen wir einige numerischen Vergleiche an. Im Regressionskontext, im Gegensatz zur Lokation, ergibt sich das Problem, dass ideale Regressorverteilungen K ausgewählt werden müssen. Für die Zwecke dieses Abschnitts betrachten wir $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ und $K = \mathrm{Unif}\,([-1, 2])$.

Wir beginnen mit der Behandlung der unbedingten Umgebungen; siehe Unterabschnitte 12.3.1 und 12.3.2. Zuerst vergleichen wir die finit, die asymptotisch und die O-korrigierten asymptotisch optimalen Stutzhöhen; siehe Unterabschnitte 12.3.1.1 und 12.3.2.1. Wie im Fall der normalen Lokation gibt es deutliche Unterschiede zwischen den finit und den asymptotisch optimalen Stutzhöhen. Hingegen liegen die O-korrigierten asymptotisch optimalen Stutzhöhen bereits für Stichproben, die aus ca. 20 Beobachtungen bestehen, sehr nahe bei den finit opti-

malen.

Als nächstes behandeln wir die finiten Risiken. Mit Hilfe unserer FFT Algorithmen vergleichen wir die finiten Risiken der finiten und der asymptotischen minimax Schätzer sowie der Schätzer, die auf der O-korrigierten asymptotisch optimalen Stutzhöhe basieren; siehe Unterabschnitte 12.3.1.2 und 12.3.2.2. Wieder stellt sich heraus, dass in den meisten Fällen die Unterschiede (in absoluten Zahlen) zwischen den entsprechenden finiten Risiken nur klein sind.

Wie im Fall der normalen Lokation deuten die numerischen Untersuchungen der Konvergenzgeschwindigkeit der finiten Risiken gegen die asymptotischen Risiken darauf hin, dass diese im Fall unbedingter Kontaminationsumgebungen $(* = c,\, t = 0)$ von der Ordnung $n^{-1/2}$, hingegen im Fall unbedingter Totalvariationsumgebungen $(* = v,\, t = 0)$ von der Ordnung n^{-1} ist.

Anschließend betrachten wir relative Risiken. Wie zuvor sind die Ergebnisse ähnlich zu denen in der normalen Lokation und nur die numerischen Werte sind unterschiedlich. In den betrachteten Fällen bleibt der maximale Effizienzverlust des asymptotischen minimax Schätzers deutlich unter 10% für Stichprobenumfänge $n \geq 50$ $(* = c,\, t = 0)$ bzw. $n \geq 20$ $(* = v,\, t = 0)$. Im Fall der Schätzer, die auf der O-korrigierten asymptotisch optimalen Stutzhöhe basieren, können kleine Werte von n $(* = c,\, t = 0\colon n < 10,\ * = v,\, t = 0\colon n < 20)$ zu sehr großen Subeffizienzen führen. Dies wird dadurch verursacht, dass die O-Korrektur für kleine Stichprobenumfänge zu groß ist und sogar zu unzulässigen (negativen) Stutzhöhen führen kann. Um zulässige Werte zu erhalten, ersetzen wir diese Werte daher durch 0. Hingegen liegen die Effizienzverluste für Stichprobenumfänge $n \geq 10$ $(* = c,\, t = 0)$ bzw. $n \geq 20$ $(* = v,\, t = 0)$ deutlich unter 5% $(* = c,\, t = 0)$ bzw. 10% $(* = v,\, t = 0)$.

In vielen Anwendungen ist die Größe der Kontamination unbekannt bzw. unbekannt bis auf ein Intervall. Daher bestimmen wir wieder die ungünstigsten Radien und die entsprechenden Ineffizienzen, welche analog zu Abschnitt 2.2 definiert sind. In den betrachteten Fällen sind die Effizienzverluste etwas größer als im Fall der normalen Lokation, bleiben aber deutlich unter 60% $(\rho = 0)$, 30% $(\rho = 1/2)$ und 15% $(\rho = 1/3)$. Aufgrund der Symmetrie der Normalverteilung sind im asymptotischen Setup, in dem die größten Ineffizienzen auftreten, die Subeffizienzen für unbedingte Kontaminations- $(* = c,\, t = 0)$ und Totalvariationsumgebungen $(* = v,\, t = 0)$ identisch. Im Kontext finiter Stichproben sind die Resultate unterschiedlich und aufgrund der schnelleren Konvergenz ist es nicht überraschend, dass die Effizienzverluste im Fall der unbedingten Totalvariationsumgebungen etwas größer sind.

Abschließend berechnen wir die finite Verteilung des finiten minimax Schätzers und vergleichen diese mit der Normalverteilung, die den geringsten Kolmogorov-Abstand d_κ aufweist; siehe Unterabschnitte 12.3.1.3 und 12.3.2.3. Analog zur normalen Lokation deuten unsere numerischen Ergebnisse darauf hin, dass die Konvergenzgeschwindigkeit im Fall unbedingter Totalvariationsumgebungen $(* = v,\, t = 0)$ nicht nur in der Null (entspricht dem finiten Risiko), sondern uniform über den gesamten Träger der finiten Verteilung des finiten minimax Schätzers schneller ist. Wieder treten die Unterschiede besonders bei sehr kleinen Stichprobenumfängen auf und bereits für $n \geq 10$ erhalten wir in beiden Fällen $d_k < 0.02$.

Als zweites betrachten wir bedingte ("fehlerfreie Variablen") Umgebungen; siehe

Unterabschnitte 12.3.3 und 12.3.4. In den Unterabschnitten 12.3.3.1 und 12.3.4.1 bestimmen wir die Kontaminationskurven, für die der Hampel-Krasker Schätzer der finite bzw. asymptotische minimax Schätzer ist. Wie sich herausstellt, sind die entsprechenden Kontaminationskurven für kleine Werte des Regressors sehr ähnlich. Jedoch ist der Umfang der Kontamination, der vom asymptotischen Ansatz für große Werte vorgeschlagen wird, unrealistisch groß und kann im Fall finiter Stichproben nicht realisiert werden. Im Gegensatz dazu ist der Umfang der Kontamination im finiten Setup beschränkt. Dies deutet darauf hin, dass die (erste Ordnungs-) Asymptotik für große Regressoren nur ein schlechte Approximation liefert.

Man kann aber für geeignete Kontaminationskurven auch den Huber Schätzer als finiten bzw. asymptotischen minimax Schätzer erhalten. Die Berechnungen in den Unterabschnitten 12.3.3.2 und 12.3.4.2 zeigen, dass die entsprechenden Kontaminationskurven für kleine Regressorwerte wieder sehr ähnlich sind. Jedoch muss der Umfang der Kontamination für große Werte im finiten Setup sehr schnell gegen 0 tendieren wohingegen im asymptotischen Setup wieder ein unrealistisch großer Kontaminationsumfang benötigt wird. Das heißt, die (erste Ordnungs-) Asymptotik scheint erneut nur eine schlechte Approximation für große Regressoren zu liefern.

Für die verbleibenden numerischen Vergleiche wählen wir konstante Kontaminationskurven und verwenden $K = \frac{1}{3}\left(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\right)$ und $K = \mathrm{Unif}\left([-1, 2]\right)$ als ideale Regressorverteilungen.

Wie wir in den Unterabschnitten 12.3.3.3 und 12.3.4.3 sehen, ist die optimale Klippingfunktion im Fall konstanter Kontaminationskurven gleich Null für kleine Regressorwerte. Dies gilt sowohl für den finiten als auch für den asymptotischen Setup. Hingegen ist die Klippingfunktion im asymptotischen Setup für große Werte unbeschränkt während sie im finiten Setup gegen 0 tendiert. Außerdem sind die O-korrigierten asymptotisch optimalen Klippingfunktionen negativ für (sehr) kleine sowie für (sehr) große Regressorwerte. Für mittelgroße Werte funktionieren die Approximationen ganz gut bis hinunter zu Stichprobenumfängen von $n = 10$ $(* = c, t = \varepsilon)$ bzw. $n = 5$ $(* = v, t = \delta)$. Diese Ergebnisse zeigen einmal mehr, dass die (erste Ordnungs-) Asymptotik für große Regressoren wohl eine schlechte Approximation liefert.

In den Unterabschnitten 12.3.3.4 und 12.3.4.4 verwenden wir unseren FFT Algorithmus, um die finiten Risiken des finiten und des asymptotischen minimax Schätzers sowie des Schätzers, der auf der O-korrigierten asymptotisch optimalen Stutzhöhe basiert (negative Werte werden auf 0 gesetzt), zu vergleichen. Wie im Fall der normalen Lokation bzw. im Fall unbedingter Umgebungen sind die Unterschiede (in absoluten Zahlen) zwischen den entsprechenden finiten Risiken in allen betrachteten Fällen nur klein. Außerdem scheint, wie in den Fällen zuvor, die Konvergenzgeschwindigkeit gegen das asymptotische Risiko im Fall bedingter Kontaminationsumgebungen $(* = c, t = \varepsilon)$ von der Ordnung $n^{-1/2}$, hingegen im Fall bedingter Totalvariationsumgebungen $(* = v, t = \delta)$ von der Ordnung n^{-1} zu sein.

Wir beenden dieses Kapitel mit der Berechnung der finiten Verteilung des finiten minimax Schätzers. Wie zuvor vergleichen wir diese mit der Normalverteilung,

die den kleinsten Kolmogorov-Abstand d_κ aufweist; siehe Unterabschnitte 12.3.3.5 und 12.3.4.5. Dieses Mal sind die Unterschiede zwischen Kontaminations- und Totalvariationsumgebungen kleiner als in den vorangegangenen Situationen. Genauer gesagt erhalten wir in beiden Fällen $d_\kappa < 0.02$ bereits für $n \geq 5$.

Introduction

This thesis consists of five parts. Each part opens with a short description of the previous treatment in literature and a summary of our own results.
Along with this dissertation comes a CD which contains the .pdf and .ps versions of this document as well as the Windows® installer and the sources for `R 2.1.1 patched` (cf. R Development Core Team (2005)). In addition, one can find on this CD: Our `R` bundle `RobASt` (cf. Appendix D), which includes the `R` packages `distrEx`, `RandVar`, `ROptEst`, `RobLox`, `ROptRegTS` and `RobRex`, as well as the required `R` packages `distr` (cf. Ruckdeschel et al. (2005)), `setRNG` (cf. Gilbert (2004)) and `evd` (cf. Stephenson (2004)).

The Need for and Justification of Robust Statistics

For detailed answers to the question "Why Robust Procedures?", respectively "Why Robust Statistics?" we refer to Section 1.1 of Huber (1981) and Chapter 1 of Huber (1997), respectively Section 1.2 of Hampel et al. (1986).

In addition, Marazzi (1993), in his introduction, gives a nice motivation for robust methods which is based on linear regression and covariance matrices. We instead use the even simpler one-dimensional normal location model; i.e., $P_\theta = \mathcal{N}(\theta, \sigma)$ where $\sigma = 1$ is known. Although this is probably the best known model in robust statistics, some new aspects (finite-sample results, higher order asymptotics) and ideas will be presented.

In our approach; the setup of infinitesimal neighborhoods, the aim of robustness is to safeguard against deviations from the assumptions which are below or near the limits of detectability; confer also p 61 of Huber (1997). The purpose of this introduction is to demonstrate, in a quantitative manner, that such small deviations, may have nontrivial effects on statistical procedures, while they cannot be detected surely by goodness-of-fit tests; confer Remark 4.2.7 of Rieder (1994). On the other hand, robust procedures are very stable and lose only little efficiency in the ideal model.

Gross Error Model

As noted in Subsection 1.2c of Hampel et al. (1986), $1-10\%$ "wrong values" (gross errors, outliers) are typical in routine data. Such real data sets can be modeled by

the well-known gross error model (convex contamination)

$$Q = (1-\varepsilon)P_\theta + \varepsilon H$$

where H is some arbitrary probability measure and $\varepsilon \in [0,1]$ is the amount of gross errors (contamination); confer Tukey (1960).

Infinitesimal Neighborhoods

In our asymptotic setup, which is based on neighborhoods that are shrinking at a rate of $\sqrt{n}$, we have to identify ε with $r/\sqrt{n}$ where $r \in [0,\infty]$. A motivation for this shrinkage in terms of the outlier probability is given in Ruckdeschel (2005a). Moreover, in the finite-sample setup we use a modification of this model. That is, for sample size $n \in \mathbb{N}$ and random variables $U_1,\ldots,U_n \overset{\text{i.i.d.}}{\sim} \text{Binom}\,(1, r/\sqrt{n}\,)$ we instead work with the following conditional probabilities

$$Q_n(r) = \Big\{ \mathcal{L}\Big([(1-U_i)X_i + U_i Y_i]_{i=1,\ldots,n} \,\Big|\, \sum U_i < n/2 \Big) \Big\}$$

where $X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} P_\theta$, $(Y_1,\ldots,Y_n) \sim H_n \in \mathcal{M}_1(\mathbb{B}^n)$ and all random variables are stochastically independent. This modification is motivated by the observation that no meaningful estimator can draw useful information out of a sample where $\sum U_i \geq n/2$. This is a similar phenomenon as breakdown point 0.5. An application of Theorem 2 in Hoeffding (1963) shows that

$$P\Big(\sum U_i \geq n/2 \Big) \leq \exp\big\{ -2n(0.5 - r/\sqrt{n}\,)^2 \big\}$$

decays exponentially fast. Hence, the above modification is asymptotically negligible; i.e., all results on weak convergence over infinitesimal neighborhoods remain unchanged. For more details we refer to Sections 2.2–2.4 of Ruckdeschel (2004c).

Remark It is a result of Ruckdeschel (2004b) that, with this modification of the $1/\sqrt{n}$ neighborhoods, the maximum mean square error (MSE) of asymptotically linear estimators with bounded (!) influence curves converges even without clipping of the square loss function. As for the artificial clipping of unbounded loss functions confer Le Cam (1986), Rieder (1994), Bickel et al. (1998) or van der Vaart (1998).

////

Estimators

As estimators we choose mean, median and robust estimators with influence curves (cf. Definition 1.1.1) of Hampel-type form[5]

$$\eta(x) = A[-c \vee x \wedge c] \qquad \text{with } A = [2\Phi(c) - 1]^{-1}$$

where $c \in (0,\infty)$ is a suitable clipping bound and Φ is the cumulative distribution function of $\mathcal{N}(0,1)$.

[5] in allusion to the solution derived in Lemma 5 of Hampel (1968)

Optimality Properties In case of normal location, extending the list on p 285 of Huber (1981), estimators with influence curves of Hampel-type form are optimal in various aspects. They minimize:

(1) The maximum asymptotic variance for symmetric convex contaminations; confer Huber (1964).

(2) The maximum asymptotic variance subject to a bias bound for infinitesimal contamination and total variation neighborhoods; confer Lemma 5 of Hampel (1968), Section 2.5 of Hampel et al. (1986), Section 5.5 of Rieder (1994) and Subsection 1.3.3.

(3) The maximum finite-sample under-/overshoot probability for contamination/total variation neighborhoods; confer Huber (1968), Rieder (1989) and Part V.

(4) The maximum asymptotic under-/overshoot probability for infinitesimal contamination/total variation neighborhoods; confer Rieder (1980) and Part V.

(5) The maximum asymptotic mean square error (MSE) for infinitesimal contamination and total variation neighborhoods; confer Section 5.5 of Rieder (1994) and Subsection 1.3.4.

(6) More generally: The maximum asymptotic risk for infinitesimal contamination and total variation neighborhoods, where risk may be any convex and isotone function of asymptotic variance and bias; confer Ruckdeschel and Rieder (2004).

(7) The second order expansion of the maximum asymptotic MSE for infinitesimal contamination neighborhoods; confer Ruckdeschel (2004b). ////

For the purpose of this introduction, we put $n = 16$ and radius $r = 0.2$ (i.e., 5% gross errors) and choose quadratic loss (i.e., MSE). We consider the asymptotically optimal-robust estimator for $r = 0.2$ (i.e., $c = 1.492$) as well as the radius–minimax estimators for $r \in [0.1, 0.4]$ (i.e., radius known up to factor 2, $\varepsilon \in [0.025, 0.1]$), $r \in [0, 2.0]$ (i.e., $\varepsilon \in [0, 0.5]$) and $r \in [0, \infty]$. The corresponding asymptotic optimal clipping bounds are $c = 1.356$, $c = 0.824$ and $c = 0.718$, respectively. For the definition of the radius–minimax estimator we refer to Section 2.2.

Choice of Clipping Bounds

(1) We use the asymptotically optimal estimators since we want to demonstrate that these estimators work well down to small sample sizes. But, the clipping bound could also be chosen optimally with respect to the finite-sample maximum MSE, respectively the second or third order expansion of the asymptotic MSE. As numerical results in Ruckdeschel (2004b) show, the differences between these various choices are small.

(2) There are also only small efficiency losses when we use different asymptotic risks to determine the optimal clipping bound c; confer Section 7.2 of Ruckdeschel and Rieder (2004).

(3) Under an additional homogeneity condition on the loss function, which for instance holds for all L_q risks with $q \geq 1$, the radius–minimax estimator for $r \in [0, \infty]$ is independent from the chosen loss function; confer Sections 6 and 7.3 of Ruckdeschel and Rieder (2004). This is in fact the reason why we included this estimator. Actually, $\varepsilon \in [0, 1]$ entails $r \in [0, 4]$ at sample size $n = 16$; i.e., radii $r > 4$ are actually not admitted. ////

Finite-Sample Maximum MSE

Our finite-sample investigation proves and makes precise what has been asserted in robust statistics on asymptotic grounds all along: In the ideal situation (i.e., $r = 0$), suitable chosen asymptotically optimal-robust estimators have a slightly larger finite-sample maximum MSE than the mean. However, they do not lose much efficiency and perform clearly better than the median in the ideal model. Contrary, for $r > 0$ the finite-sample maximum MSE of the mean is unbounded, whereas robust estimators have a bounded finite-sample maximum MSE. That is, already small deviations from the ideal model may lead to very large errors in case of the mean. In particular, the asymptotically optimal-robust estimators again perform better than the median. These are common statements; confer for instance Sections 1.1 and 1.2 of Huber (1981) or Sections 1.1 and 1.2 of Hampel et al. (1986).

The (numerically) exact finite-sample distribution and corresponding finite-sample maximum risk for robust estimators with Hampel-type influence curves, which are constructed by means of the M principle, can be computed via algorithms developed in Subsection 11.3.2 and Ruckdeschel and Kohl (2005). These procedures use the fast Fourier transform (FFT) in crucial way. In Table 4 one can find the finite-sample maximum MSE for $n = 16$ and $r = 0, 0.2$. In these situations the median shows an efficiency loss larger than 22% ($r = 0$) and 16% ($r = 0.2$), respectively.

r	mean	$r = 0.2$	$r \in [0.1, 0.4]$	$r \in [0, 2.0]$	$r \in [0, \infty]$	median
0	1.000	1.035	1.049	1.145	1.176	1.446
0.2	∞	1.450	1.431	1.443	1.465	1.713

Table 4: Finite-sample maximum MSE for normal location and sample size $n = 16$.

Finite-Sample versus Asymptotic Optimal Clipping Bounds In comparison with the asymptotic optimal clipping bound, our investigation shows that the clipping bound, which is optimal in the finite-sample sense, is in general smaller, that is, more conservative. This follows by higher oder asymptotics and numerical evaluations; confer Part V and Ruckdeschel (2004b). This fact is also reflected by the results contained in Table 4 where the asymptotically optimal-robust estimator ($c = 1.492$) has a larger finite-sample maximum MSE than the asymptotic radius–minimax estimators for $r \in [0.1, 0.4]$ ($c = 1.356$) and $r \in [0, 2.0]$ ($c = 0.824$). In fact, the numerically determined finite-sample optimal clipping bound is $c = 1.130$ and leads to a finite-sample minimax MSE of 1.418. ////

Cniper Contamination

This notion means nice and pernicious and threatens the accuracy of estimators in an unexpected and dangerous way as a sniper does; confer Section 5 of Ruckdeschel

(2004a). Now, we do not admit arbitrary $H_n \in \mathcal{M}_1(\mathbb{B}^n)$ but only contaminations by Dirac measures at $a \in \mathbb{R}$; i.e.,

$$Q_n(r,a) = \left[(1 - r/\sqrt{n}\,)P_\theta + r/\sqrt{n}\;\mathrm{I}_{\{a\}}\right]^{\otimes n}$$

We determine the gross error point a minimal such that a given robust estimator under $Q_n(r,a)$ eliminates the classically optimal estimator (in this case: the mean $\bar{X}_n$); i.e.,

$$a = \sup\left\{z > 0 \,\middle|\, \mathrm{MSE}_{Q_n(r,z)}(\bar{X}_n) \le \mathrm{MSE}_{Q_n(r,z)}(S_n^c)\right\}$$

where S_n^c is a robust estimator with influence curve of Hampel-type form for some given clipping bound $c \in (0,\infty)$. As a consequence, the robust estimator S_n^c has a smaller MSE for any contaminating distribution H with support $[a,\infty)$ (or $(-\infty,-a]$); confer Proposition 5.1 of Ruckdeschel (2004a). Under $Q_n(r,a)$ he obtains

$$n\,\mathrm{MSE}_{Q_n(r,a)}(\bar{X}_n) = (1 - r/\sqrt{n}\,) + a^2(r^2 + r/\sqrt{n} - r^2/n)$$

confer Section 5.3 (ibid.). Hence, for $M_c := n\,\mathrm{maxMSE}\,(S_n^c)$ we get

$$a = \sqrt{\frac{M_c - (1 - r/\sqrt{n}\,)}{r^2 + r/\sqrt{n} - r^2/n}}$$

For our robust estimators given in Table 4 this leads to $a = 2.391$ $(r = 0.2)$, $a = 2.345$ $(r \in [0.1, 0.4])$, $a = 2.374$ $(r \in [0, 2.0])$ and $a = 2.427$ $(r \in [0, \infty])$, respectively. These small contaminations lie well within 2.5 standard deviations from zero. Note, that under cniper contamination we even encounter less outliers, if outliers under standard normal are defined as observations with absolute value larger than 2.5; more precisely, this identifies the largest 1.24% as outliers in the ideal model whereas under cniper contamination we obtain $(1 - r/\sqrt{n}\,)1.24\% = 1.18\%$. Thus, this situation, which destroys the superiority of the mean, is surely innocent.

A Small Simulation Study

Next, we present the results of a small simulation study in the submodel introduced above of the type II errors of goodness-of-fit tests and of the MSE of location estimators. We computed $M = 1\mathrm{e}05 = 10^5$ samples of size $n = 16$ with radius $r = 0.2$ (i.e., $\varepsilon = 0.05$ and $P(\sum U_i \ge 8) = 3.50\mathrm{e}-07$). In view of the above results, $a = 2.45$ $(\Phi(-2.45) \approx 0.71\%)$ should be sufficient such that our robust estimators outperform the mean. To avoid ties, we used $H = \mathrm{Unif}\,([2.45, 2.46])$ instead of $H = \mathrm{I}_{\{2.45\}}$.

First, we tried some diagnostics. That is, we computed well-known tests for normality using `R` package `fBasics`; confer Wuertz et al. (2005). In Table 5 one can find the empirical type II errors (the null hypothesis is not rejected when it is false) of the considered tests using a significance level of 5%. As we see, the results for the chosen tests are very similar and indicate that the power (ability to reject

the null hypothesis when it is actually false) of goodness-of-fit tests is very small in case of such innocent contaminations. Thus, estimators should also be evaluated and compared under such innocent deviations from the normal.

Test for Normality	Type II Error
Anderson-Darling	93.3%
Cramér-von Mises	93.7%
Kolmogorov-Smirnov (Lilliefors)	94.2%
Shapiro-Wilk	93.4%

Table 5: Empirical Type II error of tests for normality under cniper contamination.

Remark These empirical results suggest to change the null hypothesis from exact normality to approximate normality. This is in the spirit of Section 3 in Rieder (1981b) where he extends the null hypothesis of exact symmetry to approximate symmetry and derives a nonparametric asymptotic maximin test. The corresponding modification of goodness-of-fit tests seems open. ////

Second, we computed the empirical MSE based on 1e05 samples of size 16 and corresponding 95% confidence intervals (based on the central limit theorem) of mean, median and our robust estimators; confer Table 6. This study is similar to

Estimator		$n \times$ Emp. MSE	95% conf. interval
mean		1.480	[1.467, 1.493]
$r = 0.2$:	M principle	1.445	[1.431, 1.458]
	one-step construction	1.434	[1.420, 1.447]
$r \in [0.1, 0.4]$:	M principle	1.428	[1.414, 1.441]
	one-step construction	1.423	[1.410, 1.436]
$r \in [0, 2.0]$:	M principle	1.441	[1.428, 1.454]
	one-step construction	1.448	[1.435, 1.461]
$r \in [0, \infty]$:	M principle	1.462	[1.449, 1.476]
	one-step construction	1.468	[1.455, 1.481]
median		1.712	[1.696, 1.727]

Table 6: Empirical MSE for normal location, sample size $n = 16$ and radius $r = 0.2$ under cniper contamination.

the study presented in Section 5 of Ruckdeschel (2004b). It also is in the spirit of the Princeton robustness study; confer Andrews et al. (1972). However, we choose particular asymptotically optimal estimators, compare these estimators with respect to their finite-sample MSE and consider only cniper contamination. Under

the chosen cniper contamination the mean has a (numerically) exact finite-sample MSE of 1.477 which lies well within the given empirical confidence interval. In case of our robust estimators, the corresponding estimates are determined as M estimators, respectively as one-step estimators starting with the median. In view of the general construction problem we also included the corresponding one-step estimates. As we see, our robust estimators indeed outperform mean and median where the results for the M principle and the one-step method are very similar.

Remark

(1) M principle and one-step construction work equally well down to even smaller sample sizes. It follows by the work of Ruckdeschel (2004b) and Ruckdeschel (2005e) on higher order asymptotics of the MSE of robust estimators with Hampel-type influence curves, that in normal location the M estimators and the one-step estimators are asymptotically equivalent up to second order.

(2) The median has the property which is required for initial estimators ($\sqrt{n}$ consistency on full $1/\sqrt{n}$ Kolmogorov neighborhoods). This will be shown in Subsection 2.3.4. For more details on one-step constructions we refer to Section 6.4 of Rieder (1994) and Section 2.3.

(3) In his Theorem 3.4 (b) Ruckdeschel (2004b) shows that contamination to the right of $a_n := c(1 + A\sqrt{2\log(n)/n})$ with $A = [2\Phi(c) - 1]^{-1}$ is essentially sufficient such that a robust estimator with influence curve of Hampel-type for some given clipping bound $c \in (0, \infty)$ achieves his maximum asymptotic MSE up to third order. In case of our robust estimators this leads to $a_n = 2.508$ ($r = 0.2$), $a_n = 2.324$ ($r \in [0.1, 0.4]$), $a_n = 1.645$ ($r \in [0, 2.0]$) and $a_n = 1.520$ ($r \in [0, \infty]$), respectively. Hence, it is not surprising, that the previous empirical MSEs under cniper contamination (cf. Table 6) are already very close to the finite-sample maximum MSEs evaluated in Table 4. ////

Conclusions

Since $1 - 10\%$ gross errors are reported as typical in routine data, we draw the following conclusions:

(1) Under cniper contamination our asymptotically optimal-robust estimators supersede mean and median.

(2) Such small deviations cannot be detected surely by goodness-of-fit tests.

(3) Our asymptotically optimal-robust estimators perform well down to small sample sizes; in particular, the radius–minimax estimators for $r \in [0.1, 0.4]$ (i.e., $\varepsilon \in [0.025, 0.1]$) and $r \in [0, 2.0]$ (i.e., $\varepsilon \in [0, 0.5]$) seem to be good choices for routine data if the neighborhood radius is only roughly known.

(4) M principle and one-step construction work equally well.

Proposal

In case of routine data stemming from an ideal normal model the previous conclusions suggest to proceed as follows:

Step 1: Depending on the quality of the data, try to find a rough estimate for the amount $\varepsilon \in [0, 1]$ of gross errors such that $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$.

Step 2: Compute the influence curve of our asymptotically optimal radius–minimax estimator for $r \in [\sqrt{n}\,\underline{\varepsilon}, \sqrt{n}\,\overline{\varepsilon}]$ using `S4` generic function `radiusMinimaxIC` of package `ROptEst` which is part of our `R` bundle `RobASt`; confer Appendix D.

Step 3: Choose and evaluate an appropriate initial estimate. Possible implemented candidates are the median, the MAD or the Kolmogorov(–Smirnov) MD estimator (cf. `S4` generic function `ksEstimator` of package `ROptEst`).

Step 4: Estimate the parameter of interest by means of the one-step construction using `S4` generic function `oneStepEstimator` of package `ROptEst`. ////

In this thesis we will show that the proposal given above works not only in case of normal location but, in case of general smoothly parameterized ideal models like exponential families or linear regression models. In addition, we provide the implementation of these models and of the corresponding optimally robust estimators by means of our `R` bundle `RobASt`.

Genesis of this Thesis

The work on this thesis started with some numerical investigations about the "costs of not knowing the radius"; confer Rieder et al. (2001). We computed the optimally robust influence curves for several well-known robust models like location, scale, and linear regression. For this purpose, we for instance had to invent new numerical algorithms in case of conditional (error-free-variables) regression neighborhoods.

The results contained in Part II started as an extension of these radius investigations to some non-standard robust models. In particular, they led us to some new (theoretical) results like an extension of the classical Cramér-Rao bound for quadratic loss (cf. Section 2.1) or the convergence of robust models (cf. Section 2.4). Moreover, we were able to supply the mathematical results (cf. Section 2.2) which support and complement the purely numerical determination in Rieder et al. (2001). The models treated in Part II are covered by our `R` package `ROptEst`; confer Appendix D.3.

At that time Bednarski and Müller (2001) on robust linear regression with unknown scale was published which, in our view, did not look very convincing. Hence, we decided to research into this model.
The fewest reader seem to realize, that already for scale alone, in the symmetrically contaminated model about a centered normal, the minimax variance approach of Huber (1981) remains incomplete. In addition, he shows no quantitative, let alone optimal, robustness of his joint location and scale estimates.
Moreover, the robust linear regression model with unknown error scale is covered by the local and asymptotic, infinitesimal robustness theories of Hampel et al. (1986) and Rieder (1994), however, is not treated very explicitly. This is true already for the simpler model of joint location and scale. In particular, Rieder (1994) has not specialized his optimality results contained in Section 5.5 to concrete models.

From our work Part III arose which is supplemented by Appendix B. Again, we also focussed on the numerical evaluation of the derived optimally robust influence curves. For this purpose, we implemented our `R` packages `ROptRegTS`, `RobRex` and `RobLox`; confer Appendix D.4, Section 7.6 and Section 8.9.

If the distribution of the unknown error is symmetric, the linear regression model is adaptive with respect to the unknown error scale. That is, the estimation of the regression parameter with unknown error scale is asymptotically no harder than the estimation of the regression parameter with known error scale. A very detailed treatment of adaptivity in semiparametric models is given in Bickel et al. (1998). However, robustness properties of the introduced procedures are not considered; confer p 4 (ibid.). The lectures on semiparametrics and (robust) time series models of my supervisor Prof. Dr. Rieder in 2001 and 2002 led to our definition and investigation of adaptivity in the context of infinitesimal robust regression and time series models; confer Part IV and Appendix A.
We define robust adaptivity by means of the same value of two robust optimization problems; confer Section 9.1. As a consequence, adaptivity is no longer only a dichotomous criterion but, in contrast to previous literature, now has a quantitative meaning, too. That is, we were able to evaluate the amount of non-adaptivity numerically. We restricted our considerations to the optimally robust influence curves which can be computed via our `R` package `ROptRegTS` (cf. Appendix D.4) and did not treat the construction of the corresponding robust–adaptive estimators as in Stabla (2005).

The aim of Part V was, to check the asymptotics against exact finite-sample results. More precisely, we wanted to compare the finite-sample and asymptotic results given by Huber (1968), Rieder (1989) and Rieder (1980), respectively. While we were looking for a way to compute the finite-sample risk or at least an approximation of it, we developed a convolution algorithm based on the fast Fourier transform (cf. Appendix C and Kohl et al. (2005)) which is implemented in our `R` package `distr` (cf. Ruckdeschel et al. (2005)). This procedure enables us to compute a very accurate numerical approximation of the finite-sample distribution and the corresponding finite-sample risk of robust estimators which are constructed by means of the M principle; confer Part V and Ruckdeschel and Kohl (2005). The corresponding finite-sample and asymptotic minimax estimators are implemented in our `R` packages `ROptEst` (cf. Appendix D.3) and `ROptRegTS` (cf. Appendix D.4). Motivated by the work of P. Ruckdeschel (cf. Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c), Ruckdeschel (2005e)) we also integrated some results on higher order asymptotics.

Chapter 1 on the fundamentals of the asymptotic theory of robustness was included to make this thesis better readable. It is based on Chapters 4 and 5 of Rieder (1994). Actually, Chapter 2 of Part I was not intended. As already mentioned above, the contained results for the most part emerge from the work on the other parts and many of the included (theoretical) results were originated by our numerical evaluations. The results on one-step constructions (cf. Section 2.3) arose during the work on our `R` package `ROptEst` (cf. Appendix D.3).

The `R` bundle `RobASt` (cf. Appendix D), which consists of the `R` packages `distrEx`, `RandVar`, `ROptEst`, `RobLox`, `ROptRegTS` and `RobRex` can in principle be used to re-

compute all numerical results contained in this thesis. However, the main intention was to provide a way to determine optimally robust estimators for various smooth parametric models. Our implementation is based on `S4` classes and methods (cf. Chambers (1998)) and uses our `R` package `distr` (cf. Ruckdeschel et al. (2005)) in a crucial way. With this approach we, contrary to Marazzi (1993), were able to uncouple our algorithms from specific distributional assumptions and to implement them at one stroke for whole classes of models. For example, one can use our `R` package `ROptEst` to compute optimally robust estimators for any smooth parametric model which is based on a univariate distribution.

Acknowledgements

Many people helped and supported me during the work on this thesis.

First of all I thank my supervisor Prof. Dr. Rieder who taught me robust statistics, contributed many ideas to this work and was always willing to spend his valuable time for discussions.

Also many thanks to my permanent dialogue partner Dr. Peter Ruckdeschel. Without him many of the ideas and results contained in this thesis would not have been found. In particular, he helped me to design the `R` bundle `RobASt`.

Last but not least I thank my wife Katrin, my daughters Anna and Emma, and my family and family-in-law. I owe much more than I can express here to them. Without their love, support and patience this thesis never would have been finished.

Contents

List of Figures

List of Tables

Notation

Abreviations

a.e.	almost everywhere, almost surely
ibid.	ibidem, in the same place; confer the book, chapter, article, or page cited just before
i.i.d.	stochastically independent, identically distributed
s.t.	subject to
w.r.t.	with respect to, relative to
AL	asymptotically linear
DFT	discrete Fourier transform
IC	influence curve
FFT	fast Fourier transform
ksMD	Kolmogorov(–Smirnov) minimum distance (estimator)
MD	minimum distance
MLE	maximum likelihood estimator
MSE	mean square error
maxMSE	minimax asymptotic MSE
relMSE	MSE–inefficiency
RHS	right-hand side
////	QED

Sets and Functions

$\mathbb{N}$	the natural numbers $\{1, 2, \ldots\}$
$\mathbb{N}_0$	the natural numbers including 0 $\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	the integers $\{\ldots, -1, 0, 1, \ldots\}$
$\mathbb{R}$	the real numbers $(-\infty, \infty)$
$\bar{\mathbb{R}}$	the extended real numbers $[-\infty, \infty]$, homeomorphic to $[0, 1] \subset \mathbb{R}$ via the smooth isometry $z \mapsto \frac{e^z}{e^z+1}$
$\mathbb{C}$	the complex numbers
$\times$	Cartesian product of sets
I_A, $\mathrm{I}(A)$	indicator function of a set or statement A; i.e., for any set A, we may write $\mathrm{I}_A(x) = \mathrm{I}(x \in A)$
sign	$\mathrm{sign}(x) = -1, 0, 1$ for x negative/zero/positive
$f(x \pm 0)$	right/left-hand limit at x of a function f
Λ	L_2 derivative
$\mathcal{I}$	Fisher information

$\boldsymbol{\sigma}$ Algebras

$\mathbb{B}$, $\bar{\mathbb{B}}$	Borel σ algebras on $\mathbb{R}$ and $\bar{\mathbb{R}}$, respectively
$\bigotimes$	product of σ algebras

Measures

$\mathcal{M}_1(\mathcal{A})$	the probability measures (mass 1) on $\mathcal{A}$
support P	smallest closed subset A of Ω (separable, metric) such that $P(\Omega \setminus A) = 0$; cf. II Definition 2.1 of Parthasarathy (1967)
$\ll$	domination of measures
$\bigotimes$	product of measures
$*$	convolution of measures
$\xrightarrow{\text{w}}$	weak convergence of (bounded) measures

Random Variables and Expectation

$\sim$	distributed according to
$\mathcal{L}_P(X)$	law of X under P
$\mathrm{E}\, X$	expectation of X
$\mathrm{Var}\, X$	variance of X
$\mathrm{Cov}\, X$	covariance of X
$E_\bullet$	conditional expectation given the regressor
$\xrightarrow{P_n}$	stochastic convergence, convergence in P_n probability

Laws

I_a	(Dirac) one-point measure in a; $\mathrm{I}_a(A) = \mathrm{I}_A(a)$
$\mathrm{Binom}\,(m, p)$	binomial distribution with size $m \in \mathbb{N}$ and probability of success $p \in [0, 1]$
$\mathrm{Exp}\,(\lambda)$	exponential distribution with scale $\lambda \in (0, \infty)$
$\mathrm{Gamma}\,(\sigma, \alpha)$	Gamma distribution with scale $\sigma \in (0, \infty)$ and shape $\alpha \in (0, \infty)$
$\mathrm{Gumbel}\,(\mu, \sigma)$	Gumbel distribution with location parameter $\mu \in \mathbb{R}$ and scale $\sigma \in (0, \infty)$
$\mathcal{N}\,(\mu, \sigma^2)$	normal law on $(\mathbb{R}^m, \mathbb{B}^m)$ with mean $\mu \in \mathbb{R}^m$ and standard deviation σ
φ, Φ	standard normal density and distribution function on $\mathbb{R}$
$\mathrm{Pois}\,(\lambda)$	Poisson distribution with mean $\lambda \in (0, \infty)$
$\mathrm{Unif}\,(M)$	uniform distribution on $M \subset \mathbb{R}$

Mathematical Symbols

$\subset$, $\supset$	subset/supset
$\leq$	less or equal, coordinatewise on $\mathbb{R}^m$
$*$	convolution of periodic sequences of complex numbers
x_+, x_-	positive, negative parts
$\wedge$, min	minimum
$\vee$, max	maximum

$\inf$, $\sup$	pointwise infimum/supremum
$\inf_P$, $\sup_P$	P essential infimum, supremum
$\inf_\bullet$, $\sup_\bullet$	conditional essential extrema given the regressor
$\uparrow$, $\downarrow$	monotone convergence form below/above of numbers
o, O	the Landau symbols
$\operatorname{lin}(x_1, \ldots, x_k)$	linear space generated by $x_1, \ldots, x_k$

Matrices and Vectors

$\mathbb{I}_k$	the unit $k \times k$ matrix
$A \in \mathbb{R}^{p\times k}$	a real matrix with p rows and k columns
A^τ	transpose of a matrix A
$\operatorname{rk} A$	rank of A
$\operatorname{tr} A$	trace of A
$\operatorname{minev}(A)$	minimal eigenvalue of A
$\mathcal{C}(A)$	column space of A
$\mathcal{N}(A)$	null space of A
$A \succ B$	$A - B$ positive definite
$A \succeq B$	$A - B$ positive semidefinite
$A \otimes B$	Kronecker product of matrices A and B; cf. Definition B.1.1
$\operatorname{vec}(A)$	vec operator: transforms matrix $A \in \mathbb{R}^{p\times k}$ to a pk-dimensional column vector; cf. Definition B.1.2
$\operatorname{vech}(A)$	vech operator: transforms matrix $A \in \mathbb{R}^{k\times k}$ to a $k(k+1)/2$-dimensional column vector; cf. Definition B.2.1
$\operatorname{diag}(a_1, \ldots, a_k)$	diagonal matrix with diagonal $a_1, \ldots, a_k$
$\|\|A\|\|_{\mathrm{op}}$	operator norm of A; i.e., $\|\|A\|\|_{\mathrm{op}} = \sup_{\|x\|\le 1} \|Ax\|$
$\|x\|$	Euclidean norm of $x \in \mathbb{R}^m$

Function Spaces

$\mathcal{D}_1$	functions: $\bar{\mathbb{R}} \to \mathbb{R}$ which are infinitely differentiable on $\mathbb{R}$, continuous on $\bar{\mathbb{R}}$ and whose derivative φ' has compact support
$\mathcal{C}_c^1$	functions: $\mathbb{R} \to \mathbb{R}$ which are continuously differentiable functions and have compact support
$\mathcal{C}_c^\infty$	functions: $\mathbb{R} \to \mathbb{R}$ which are infinitely differentiable and have compact support
$L_2^k(P)$	the Hilbert space of (equivalence classes of) $\mathbb{R}^k$-valued functions f such that $\int \|f\|^2 dP < \infty$; $L_2(P) = L_2^1(P)$
$L_\infty^k(P)$	the space of (equivalence classes of) $\mathbb{R}^k$-valued functions f such that $\sup_P \|f\| < \infty$; $L_\infty(P) = L_\infty^1(P)$
$Z_\infty(\theta)$	$L_\infty^k(P_\theta) \cap \{\mathrm{E}_\theta = 0\}$; the tangents at P_θ
$\Psi_\alpha(\theta)$, $\Psi_\alpha^D(\theta)$	set of square integrable ($\alpha = 2$), and bounded ($\alpha = \infty$), influence curves at P_θ; respectively, partial influence curves at P_θ, with some matrix $D \in \mathbb{R}^{p\times k}$ such that $\operatorname{rk} D = p \le k$; cf. Definition 1.1.1

$\Psi_{\alpha\bullet}(\theta)$, $\Psi^D_{\alpha\bullet}(\theta)$	set of square integrable ($\alpha = 2$), and bounded ($\alpha = \infty$), conditionally centered (partial) influence curves, respectively; cf. Definition A.1.2
$\Psi_{\mathrm{M}}(\theta)$	square integrable influence curves for general M estimators; cf. (7.1.50)
$\Psi_{\mathrm{Mc}}(\theta)$	conditionally centered, square integrable influence curves for general M estimators; cf. (7.1.51)
$\Psi_{\mathrm{Ms}}(\theta)$	conditionally centered, square integrable influence curves for sectionwise M estimators; cf. (7.1.53)
$\Psi_{\mathrm{BM}}(\theta)$	conditionally centered, square integrable influence curves for BM estimators; cf. (7.1.52)

Neighborhoods and Bias Terms

$* = c, v, \kappa$	type of balls and metric: contamination, total variation, Kolmogorov
$t = 0, \varepsilon, \delta, \alpha$	type of neighborhoods: (general parameter) unconditional ($t = 0$), and conditional regression neighborhoods with fixed contamination curve ($* = \varepsilon, \delta$), respectively average (square) conditional regression balls of exponent $\alpha = 1, 2$; cf. Sections 10.1 and A.1
$\mathcal{U}_*(\theta)$, $\mathcal{U}_{*,t}(\theta)$	neighborhood system about P_θ
$U_*(\theta, r)$, $U_{*,t}(\theta, r)$	such a neighborhood about P_θ of radius $r \in (0, \infty)$; in the infinitesimal robust setup, usually $r = \mathrm{O}(1/\sqrt{n}\,)$
$\mathcal{G}_*(\theta)$, $\mathcal{G}_{*,t}(\theta)$	corresponding tangent classes
ω_*, $\omega_{*,t}$	standardized (infinitesimal) bias terms
ω_{M}	standardized (infinitesimal) bias terms for general M estimators
ω_{Mc}	standardized (infinitesimal) bias terms for general M estimators with conditionally centered ICs
ω_{Ms}	standardized (infinitesimal) bias terms for sectionwise M estimators
ω_{BM}	standardized (infinitesimal) bias terms for BM estimators

Part I

Asymptotic Theory of Robustness

CHAPTER 1 provides a description and summary of the asymptotic theory which forms the basis of this thesis. In the robustness context it is associated with the names of Bickel and Rieder; confer Bickel (1981) and Rieder (1994). Our presentation is based on Chapters 4 and 5 of Rieder (1994). It is restricted to the estimation of a finite-dimensional parameter in the one sample i.i.d. case. More precisely, we assume a smoothly parameterized family

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A})$$

of probability measures on some sample space $(\Omega, \mathcal{A})$, whose parameter space Θ is an open subset of some finite dimensional $\mathbb{R}^k$. At some fixed $\theta \in \Theta$ this family $\mathcal{P}$ is assumed to be L_2 differentiable,

$$\|\sqrt{dP_{\theta+t}} - \sqrt{dP_\theta}\,(1 + \tfrac{1}{2}t^\tau \Lambda_\theta)\| = o(|t|)$$

with L_2 derivative $\Lambda_\theta \in L_2^k(P_\theta)$ and Fisher information of full rank k,

$$\mathcal{I}_\theta = \mathrm{E}_\theta\, \Lambda_\theta \Lambda_\theta^\tau$$

For more details on L_2, respectively L_r $(r \geq 1)$ differentiability we refer to Section 2.3 of Rieder (1994), Section 1.8 of Witting (1985) and Rieder and Ruckdeschel (2001), respectively.

In Section 1.1 we introduce the notion of (partial) square integrable influence curves (involving a matrix $D \in \mathbb{R}^{p\times k}$ of full rank $p \leq k$) and show that a necessary and sufficient condition for their existence is

$$\exists A \in \mathbb{R}^{p\times k}\colon D = A\mathcal{I}_\theta$$

Next, we introduce asymptotically linear estimators and derive the Cramér-Rao bound for this class of estimators.

The infinitesimal robust setup which is based on neighborhoods about the ideal model $\mathcal{P}$ that are shrinking at a rate of $\sqrt{n}$ is presented in Section 1.2. Throughout this thesis we consider neighborhoods of contamination, total variation and occasionally of Kolmogorov type. That is, we omit Hellinger and Cramér von Mises neighborhoods as treated in Rieder (1994).

Subsequently, tangent classes for these neighborhoods are defined and then simple perturbations along such tangents instead of full neighborhoods are considered. As a consequence of Le Cam's third lemma, one gets the asymptotic normality of asymptotically linear estimators under such simple perturbations. Using quadratic loss this leads to the asymptotic mean square error (MSE) problem stated in Subsection 1.3.1. This convex optimization problem involves certain bias terms (depending on the type of neighborhood) which can be calculated more or less explicitly; confer Subsection 1.3.2.

The solution to this optimization problem has been derived in detail in Section 5.5 of Rieder (1994) using Lagrange multiplier theorems developed in Appendix B (ibid.). To obtain the corresponding MSE solution, beforehand the problem of minimizing the trace of the asymptotic covariance subject to a bound on

the various bias types is solved. Thus, we also give the solution (optimal influence curve) to this supplementary problem; confer Subsection 1.3.3. Additionally, the minimum asymptotic bias and the influence curve which achieves this minimum bias is specified. As for the original minimax MSE problem, the optimal influence curve is of the same form as in case of the minimum trace problem with a suitable bias bound. This bound is determined by an additional implicit equation; confer Subsection 1.3.4. The MSE solution is always of main case form; confer Theorem 1.3.9 (a).

CHAPTER 2 presents supplements to the asymptotic theory of robustness which have proved necessary for this thesis.

First, we show in Subsection 2.1.1 that the Lagrange multiplier A, which occurs in the optimal influence curves and is determined by an optimization problem using Lagrange multiplier arguments, has a statistical interpretation:

$$\text{minimaxMSE} = \operatorname{tr} A$$

This identity extends the classical Cramér-Rao bound for quadratic loss and is remarkable since, in addition to variance, bias is involved.

Next, we treat discrete models which have rarely been considered in robustness literature; confer Subsection 2.1.2. The models show peculiar aspects: Under an additional "gap" condition, the MSE solution (always of main case form) in fact coincides with the minimum bias solution. This happens for radii r greater than some finite radius $\bar{r} \in [0, \infty)$, the so called lower case radius. Another phenomenon which has not been studied in literature so far is non-uniqueness of the Lagrange multipliers as part of the (unique) optimal influence curves.

In the remaining part of Section 2.1, we derive technical properties of the Lagrange multipliers contained in the MSE solution: Boundedness (cf. Subsection 2.1.3), uniqueness (cf. Subsection 2.1.4) and continuity (cf. Subsection 2.1.5). These properties are important for the following purposes: Determination of unknown neighborhood radius according to a minimax criterion (cf. Section 2.2), estimator construction (cf. Section 2.3) and convergence of robust models (cf. Section 2.4).

In Section 2.2 we consider the notions of least favorable radius and radius–minimax estimator introduced in Rieder et al. (2001). This concept serves as a strategy how to proceed if the true neighborhood radius is unknown, respectively unknown except to belong to some radius interval. We supply the mathematical results on the least favorable radius which support and complement the purely numerical determination in Rieder et al. (2001).

Another important problem is the construction of optimally robust estimators. So far, the results concern the MSE optimal influence curve whose derivation is based only on simple perturbations. Given a family of influence curves $(\psi_\theta)_{\theta\in\Theta}$, we have to construct an asymptotic estimator S, without knowing the parameter $\theta \in \Theta$, such that S is asymptotically linear with influence curve ψ_θ at P_θ. Moreover, the risk of this estimator should not increase if one passes over from simple perturbations to full neighborhoods about P_θ. These goals can (under additional assumptions) be achieved by means of one-step constructions, and sufficient conditions are given in Subsection 2.3.1.

The general sufficient conditions are taken from Section 6.4 of Rieder (1994). We use them to derive sufficient conditions for the MSE optimal influence curve; confer Subsection 2.3.2. We verify these conditions for exponential families of full rank; confer Subsection 2.3.3. Thus, we can use the one-step method in several important models which are widely used in parametric statistics. In particular, these results apply to most of the models considered in this thesis.

The one-step construction requires a suitable initial estimator. By Theorem 6.3.7 of Rieder (1994) the Kolmogorov minimum distance estimator has the necessary properties if we consider $1/\sqrt{n}$ neighborhoods of Kolmogorov type. Consequently, this is also true if we consider smaller $1/\sqrt{n}$ neighborhoods like contamination or total variation neighborhoods. However, in robust literature most frequently the simpler median and median absolute deviation (MAD) are proposed as appropriate initial estimators. Since there seems to be no reference that these estimators also have the asserted properties, we prove their uniform $\sqrt{n}$ consistency on $1/\sqrt{n}$ Kolmogorov neighborhoods even if there is no location or scale structure; confer Subsection 2.3.4.

In the remaining part of the current chapter we derive some results which may be interpreted as convergence of robust models; confer Section 2.4. We do not aim at the abstract framework of Le Cam (1986) involving arbitrary decision rules. But, from the beginning, we base our concept solely on the optimally robust estimators. We prove that under certain weak assumptions and with appropriate standardizations the Lagrange multipliers of the MSE optimal influence curve of one model converge towards the Lagrange multipliers of the MSE optimal influence curve of another model. Hence, the minimax asymptotic MSE, the standardized asymptotic bias, and the asymptotic variance converge, too. Thus, if there is some infinitesimal robust model where the optimally robust influence curve is hard to determine, we can try to find another robust model which may serve as approximation and where the computation of the corresponding optimally robust influence curve is much easier. Using this influence curve we are able to construct approximations to the optimally robust influence curve for the model of interest which also is in the spirit of Le Cam (1986). Convincing examples are given in Chapters 3–5. The concept — convergence of robust models — may certainly be expanded more abstractly.

Chapter 1

Asymptotic Theory of Robustness – an Abridge

This initial chapter, for the most part, is based on Chapters 4 and 5 of Rieder (1994). In Section 1.1 we give the definition of partial influence curves (ICs) and specify a necessary and sufficient condition for their existence. We then introduce asymptotically linear estimators (AL estimators) and verify the Cramér-Rao bound in this smooth parametric i.i.d. case and restricted to the class of AL estimators. In the infinitesimal robust setup presented in Section 1.2, a parametric family $\mathcal{P}$ serves as ideal center model and at least under the null hypothesis $P_\theta \in \mathcal{P}$ the observations $y_1, \ldots, y_n$ at time $n \in \mathbb{N}$ are assumed to be i.i.d.. In Section 1.3 we give the solutions (i.e., the optimally robust ICs) to the optimization problems motivated in Subsection 1.3.1.
In the sequel, expectation and covariance unter P_θ are denoted by E_θ and Cov_θ, respectively.

1.1 Asymptotically Linear Estimators

Often influence curves (ICs) are introduced as Gâteaux derivatives of statistical functionals; confer Section 2.5 of Huber (1981) and Section 2.1 of Hampel et al. (1986), respectively. But, most proofs of asymptotic normality in the i.i.d. case head for an estimator expansion, in which ICs canonically occur as summands; confer M, L, R, S and MD (minimum distance) estimates. The following definition corresponds to Definition 4.2.10 of Rieder (1994).

Definition 1.1.1 *Suppose $\mathcal{P}$ is L_2 differentiable at θ, and assume some matrix $D \in \mathbb{R}^{p\times k}$ of full rank $p \le k$. Let $\alpha = 2, \infty$, respectively.*

(a) *Then, the set $\Psi_2(\theta)$ of all* square integrable *and the subset $\Psi_\infty(\theta)$ of all* bounded ICs at *P_θ, respectively, are*

$$\Psi_\alpha(\theta) = \left\{\psi_\theta \in L_2^k(P_\theta) \mid \mathrm{E}_\theta\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k\right\} \tag{1.1.1}$$

(b) *The set* $\Psi_2^D(\theta)$ *of all* square integrable *and the subset* $\Psi_\infty^D(\theta)$ *of all* bounded, partial ICs at P_θ, *respectively, are*

$$\Psi_\alpha^D(\theta) = \left\{\psi_\theta \in L_2^p(P_\theta) \,|\, \mathrm{E}_\theta\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = D\right\} \tag{1.1.2}$$

In this context we repeat Remark 4.2.11 of Rieder (1994) where we omit part (d) about L_1 differentiability, note part (e) without the proof and add *time series models* to part (f).

Remark 1.1.2 **(a)** The attribute *square integrable* will usually be omitted.

(b) The *classical scores* and the *classical partial scores*,

$$\psi_{h,\theta} = \mathcal{I}_\theta^{-1}\Lambda_\theta \in \Psi_2(\theta) \tag{1.1.3}$$

$$\eta_{h,\theta} = D\psi_{h,\theta} = D\mathcal{I}_\theta^{-1}\Lambda_\theta \in \Psi_2^D(\theta) \tag{1.1.4}$$

are always ICs, respectively, partial ICs, at P_θ.

(c) The definition of $\Psi_2(\theta)$ and $\Psi_\infty(\theta)$ requires $\mathcal{I}_\theta \succ 0$, and Λ_θ nondegenerate in the sense that, for all $t \in \mathbb{R}^k$,

$$t^\tau \Lambda_\theta = 0 \quad \text{a.e. } P_\theta \qquad \Longrightarrow \qquad t = 0 \tag{1.1.5}$$

(d) [...]

(e) [...] $\Psi_\alpha^D(\theta) = \left\{D\psi_\theta \,|\, \psi_\theta \in \Psi_\alpha(\theta)\right\}$ [...]

(f) Of course, $\Psi_\alpha^D(\theta) = \Psi_\alpha(\theta)$ for $D = \mathbb{I}_k$. Partial ICs with general D occur when there are nuisance components. In robust regression – *respectively, time series models* – moreover, conditionally centered (partial) ICs will occur.

(g) $\Psi_\alpha^D(\theta)$ are closed convex subsets of $L_\alpha^p(P_\theta)$; $\alpha = 2, \infty$. ////

The following lemma gives a necessary and sufficient condition for the existence of (square integrable) partial ICs. Its proof is based on arguments provided by H. Rieder.

Lemma 1.1.3 *It holds*

$$\Psi_2^D(\theta) \neq \emptyset \qquad \Longleftrightarrow \qquad \exists A \in \mathbb{R}^{p\times k}\colon D = A\mathcal{I}_\theta \tag{1.1.6}$$

Proof Let $A \in \mathbb{R}^{p\times k}$ be some matrix (of full rank p) such that $D = A\mathcal{I}_\theta$. Then,

$$D = A\mathcal{I}_\theta = A\,\mathrm{E}_\theta\, \Lambda_\theta\Lambda_\theta^\tau = \mathrm{E}_\theta\, A\Lambda_\theta\Lambda_\theta^\tau \tag{1.1.7}$$

That is, if we consider $\psi_\theta := A\Lambda_\theta$, we additionally obtain $\mathrm{E}_\theta\, \psi_\theta = A\,\mathrm{E}_\theta\, \Lambda_\theta = 0$ and

$$\mathrm{E}_\theta\, |\psi_\theta|^2 = \mathrm{tr}\,\mathrm{E}_\theta\, A\Lambda_\theta\Lambda_\theta^\tau A^\tau = \mathrm{tr}\, A\mathcal{I}_\theta A^\tau = \mathrm{tr}\, AD^\tau < \infty \tag{1.1.8}$$

Hence, $\psi_\theta \in \Psi_2^D(\theta)$.

Conversely, assume there exists some $\psi_\theta \in \Psi_2^D(\theta)$. We project ψ_θ onto

$$[\mathrm{lin}\,\{\Lambda_{\theta,1}, \ldots, \Lambda_{\theta,k}\}]^p = \{A\Lambda_\theta \,|\, A \in \mathbb{R}^{p\times k}\} \tag{1.1.9}$$

in $L_2^p(\theta)$. That is, $\psi_\theta = \chi_\theta + A\Lambda_\theta$ where $\mathrm{E}_\theta\, \chi_\theta\Lambda_\theta^\tau = 0$. Consequentially,

$$D = \mathrm{E}_\theta\, \psi_\theta\Lambda_\theta^\tau = \mathrm{E}_\theta(\chi_\theta + A\Lambda_\theta)\Lambda_\theta^\tau = A\mathcal{I}_\theta \tag{1.1.10}$$

////

Remark 1.1.4 **(a)** The previous lemma shows that we do not necessarily need $\mathcal{I}_\theta \succ 0$ for the existence of partial ICs. But, since $\operatorname{rank}(D) = p$, it has to hold $\operatorname{rank}(A\mathcal{I}_\theta) = p$ where

$$\operatorname{rank}(A\mathcal{I}_\theta) = \operatorname{rank}(\mathcal{I}_\theta) - \dim\left(\mathcal{C}(\mathcal{I}_\theta) \cap \mathcal{N}(A)\right) \tag{1.1.11}$$

with $\mathcal{C}(\mathcal{I}_\theta)$ the column space of $\mathcal{I}_\theta$ and $\mathcal{N}(A)$ the null space of A; confer Theorem 17.5.4 of Harville (1997). Consequentially, the Fisher information $\mathcal{I}_\theta$ at least has to have rank p.

(b) Another necessary condition for the existence of (square integrable) partial influence curves is $D\mathcal{I}_\theta D^\tau \succ 0$. However, this condition is not sufficient. Consider for example $k = 2$, $p = 1$, $D = (1, 1)$ and

$$\mathcal{I}_\theta = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{1.1.12}$$

Then, $D\mathcal{I}_\theta D^\tau = 1 > 0$ but there exists no $A \in \mathbb{R}^{1\times 2}$ such that $D = A\mathcal{I}_\theta$. ////

Next, we give the definition of asymptotically linear estimators (AL estimators); confer Definition 4.2.16 of Rieder (1994).

Definition 1.1.5 *An asymptotic estimator*

$$S = (S_n) \qquad S_n : (\Omega^n, \mathcal{A}^n) \to (\mathbb{R}^k, \mathbb{B}^k) \tag{1.1.13}$$

is called asymptotically linear at P_θ *if there is an IC* $\psi_\theta \in \Psi_2(\theta)$ *such that*

$$R_n = \sqrt{n}\,(S_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\theta(y_i) + o_{P_\theta^n}(n^0) \tag{1.1.14}$$

We call $R = (R_n)$ standardization, *and* ψ_θ *the* IC, *of* S *at* P_θ.

We now state Remark 4.2.17 of Rieder (1994) where we omit part (c) on L_1 differentiability and part (f) on the nonparametric convolution and asymptotic minimax theorems.

Remark 1.1.6 **(a)** The expansion (1.1.14) determines the IC ψ_θ uniquely, because $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta(y_i)$ with $\eta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta\, \eta = 0$, can tend to 0 in P_θ^n probability only if $\mathrm{E}_\theta\, |\eta|^2 = 0$; that is, $\eta = 0$ a.e. P_θ.

(b) If S is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$, then

$$\sqrt{n}\,(S_n - \theta)(P_\theta^n) \xrightarrow{\mathrm{w}} \mathcal{N}(0, \mathrm{Cov}_\theta(\psi_\theta)) \tag{1.1.15}$$

because of $\psi_\theta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta\, \psi_\theta = 0$, and the Lindeberg-Lévy CLT. The third condition $\mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k$, as already noted in the remarks of Rieder (1980) (p 108), is equivalent to the locally uniform extension of this asymptotic normality; see Lemma 4.2.18 of Rieder (1994).

(c) [...]

(d) Extending general M estimates, the class of AL estimators has in the case $k = 1$ been introduced by Rieder (1980). Bickel (1981) defined the related notion CULAN, employing however compact subsets of Θ instead of compacts in the local parameter space.

(e) The class of AL estimators contains the common asymptotically normal M, L, R and MD (minimum distance) estimates; confer Chapters 1 and 6 of Rieder (1994). In fact, most proofs of asymptotic normality in the i.i.d. case end up with an extension (1.1.14); the corresponding conditions need to be verified only under the ideal model.

(f) [...]

(g) The previous robustness theories of Huber (1964), Hampel (1974b), Rieder (1980) and Bickel (1981) have been formulated but for AL estimators or, even more specialized, for M estimates. ////

In this smooth parametric context one can now easily prove the Cramér-Rao bound for AL estimators.

Proposition 1.1.7 *Consider an estimator $S = (S_n)$ that is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$. Then,*

$$\mathrm{Cov}_\theta(\psi_\theta) \succeq \mathcal{I}_\theta^{-1} = \mathrm{Cov}_\theta(\psi_{h,\theta}) \tag{1.1.16}$$

in the positive definite sense, with equality iff $\psi_\theta = \psi_{h,\theta}$.

Proof An immediate consequence of

$$0 \preceq \mathrm{E}_\theta(\psi_\theta - \psi_{h,\theta})(\psi_\theta - \psi_{h,\theta})^\tau = \mathrm{Cov}_\theta(\psi_\theta) - \mathcal{I}^{-1} \tag{1.1.17}$$

////

Remark 1.1.8 This optimality holds not only for AL estimators but can be extended to estimators that are regular in some weaker sense than asymptotic linearity or even to arbitrary, measurable estimators. The corresponding optimality results are subject of the convolution and asymptotic minimax theorems. For more details we refer to Sections 3.2, 3.3 of Rieder (1994), Sections 8.5, 8.7 of van der Vaart (1998) or Section 2.3 of Bickel et al. (1998), respectively. ////

1.2 Infinitesimal Robust Setup

For a very detailed introduction and motivation of robust statistics we refer to Chapter 1 of Hampel et al. (1986). A quick introduction to robustness is also given by Huber (1997). In this section we present the infinitesimal robust setup which can be found in Subsection 4.2.1 of Rieder (1994). For more details on this we also refer to Bickel (1981). Let

$$\mathcal{U}(\theta) = \big\{U(\theta, r) \,\big|\, r \in [0, \infty)\big\} \tag{1.2.1}$$

be any system of neighborhoods $U(\theta, r)$ of radius $r \in [0, \infty)$ about P_θ such that

$$P_\theta \in U(\theta, r_1) \subset U(\theta, r_2) \subset \mathcal{M}_1(\mathcal{A}) \qquad 0 \le r_1 < r_2 < \infty \tag{1.2.2}$$

Within this thesis we work with (convex) contamination ($* = c$), total variation ($* = v$) and occasionally with Kolmogorov ($* = \kappa$) neighborhood systems $\mathcal{U}_*(\theta)$. Rieder (1994) in addition considers Hellinger ($* = h$), Cramér-von Mises ($* = \mu$), Prokhorov ($* = \pi$) and Lévy ($* = \lambda$) neighborhood systems. In the cases $* = c, v$ the systems $\mathcal{U}_*(\theta)$ consist of closed balls about P_θ that are defined for an arbitrary sample space $(\Omega, \mathcal{A})$,

$$U_*(\theta, r) = B_*(P_\theta, r) \qquad r \in [0, \infty) \qquad * = c, v \tag{1.2.3}$$

where

$$B_c(P_\theta, r) = \big\{(1-r)_+ P_\theta + (1 \wedge r)\, Q \,\big|\, Q \in \mathcal{M}_1(\mathcal{A})\big\} \tag{1.2.4}$$

$$B_v(P_\theta, r) = \big\{Q \in \mathcal{M}_1(\mathcal{A}) \,\big|\, d_v(Q, P_\theta) \le r\big\} \tag{1.2.5}$$

with metric

$$d_v(Q, P_\theta) = \frac{1}{2} \int |dQ - dP_\theta| = \sup_{A \in \mathcal{A}} |Q(A) - P_\theta(A)| \tag{1.2.6}$$

and it holds $B_c(P_\theta, r) \subset B_v(P_\theta, r)$. In case $* = \kappa$ a finite-dimensional Euclidean sample space $(\Omega, \mathcal{A}) = (\mathbb{R}^m, \mathbb{B}^m)$ is employed and

$$U_\kappa(\theta, r) = \big\{Q \in \mathcal{M}_1(\mathbb{B}^m) \,\big|\, d_\kappa(Q, P_\theta) \le r\big\} \tag{1.2.7}$$

with metric

$$d_\kappa(Q, P_\theta) = \sup_{x \in \mathbb{R}^m} \big|Q\big(\{y \in \mathbb{R}^m \,|\, y \le x\}\big) - P_\theta\big(\{y \in \mathbb{R}^m \,|\, y \le x\}\big)\big| \tag{1.2.8}$$

where $\le$ is coordinatewise. Obviously, $B_v(P_\theta, r) \subset B_\kappa(P_\theta, r)$ as the supremum in case $* = v$ is taken over more sets.

Remark 1.2.1 The observations $y_1, \ldots, y_n$, which are i.i.d. under the null hypothesis P_θ, may now be allowed to follow any law $Q \in U_*(\theta, r)$, while still the parameter θ has to be estimated. Since the equation

$$Q = P_\theta + (Q - P_\theta) \tag{1.2.9}$$

involving the nuisance component $Q - P_\theta$, has multiple solutions θ, the parameter θ is obviously no longer identifiable. This problem may also be considered using functionals that are Fisher consistent at the ideal model and extend the parametrization to the neighborhoods. As noted in the preface and in Subsection 4.3.3 of Rieder (1994), both approaches lead to the same optimally robust ICs and procedures if the choice of the functional is subject to robustness criteria. ////

We now fix $\theta \in \Theta$ and define the following classes of one-dimensional bounded tangents

$$\mathcal{G}_c(\theta) = \big\{ q \in Z_\infty(\theta) \,\big|\, \inf\nolimits_{P_\theta} q \geq -1 \big\} \tag{1.2.10}$$

and

$$\mathcal{G}_v(\theta) = \big\{ q \in Z_\infty(\theta) \,\big|\, \mathrm{E}_\theta\, |q| \leq 2 \big\} \tag{1.2.11}$$

where

$$Z_\infty(\theta) = \big\{ q \in L_\infty(P_\theta) \,|\, \mathrm{E}_\theta\, q = 0 \big\} \tag{1.2.12}$$

Then, simple perturbations along $q \in Z_\infty(\theta)$ are, for $\sqrt{n} \geq -r \inf_{P_\theta} q$, defined as

$$dQ_n(q,r) = \left(1 + \frac{r}{\sqrt{n}}\, q\right) dP_\theta \tag{1.2.13}$$

Remark 1.2.2 For a general definition of p-dimensional tangents and sequences of simple perturbations we refer to Definition 4.2.1 and pp 126 of Rieder (1994). ////

Lemma 1.2.3 *Given $q \in Z_\infty(\theta)$ and $r \in (0,\infty)$. Then, in the cases $* = c, v$, for every $n \in \mathbb{N}$ such that $\sqrt{n} \geq -r \inf_{P_\theta} q$,*

$$Q_n(q,r) \in B_*(P_\theta, r/\sqrt{n}\,) \quad \Longleftrightarrow \quad q \in \mathcal{G}_*(\theta) \tag{1.2.14}$$

Proof Rieder (1994), Lemma 5.3.1. ////

In this context we repeat Remark 4.2.7 of Rieder (1994) as part (a) of the following remark.

Remark 1.2.4 (a) With the $1/\sqrt{n}$ scaling, a neighborhood system is also called *infinitesimal*. For sample size $n \to \infty$, neighborhoods and simple perturbations are scaled down so, because, on the one hand, such deviations from the ideal model have nontrivial effects on statistical procedures, while, on the other hand, they cannot be detected surely by goodness-of-fit tests.

(b) Ruckdeschel (2005a) gives a motivation why these neighborhoods are shrinking at a rate of $\sqrt{n}$ by constructing a minimax test for the outlier probability using upper probabilities. This simplifies the idea of goodness-of-fit tests which is noted in part (a) and also in Bickel (1981), pp 36. ////

AL estimators are asymptotically normal distributed under simple perturbations.

Proposition 1.2.5 *Let S be an estimator that is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$ and given $q \in Z_\infty(\theta)$ and $r \in (0,\infty)$ consider the simple perturbations $Q_n(q,r)$. Then,*

$$\sqrt{n}\,(S_n - \theta)\big(Q_n^n(q,r)\big) \xrightarrow{\mathrm{w}} \mathcal{N}_k\big(r\,\mathrm{E}_\theta\, \psi_\theta q,\, \mathrm{Cov}_\theta(\psi_\theta)\big) \tag{1.2.15}$$

PROOF Consequence of Lemma 4.2.4 of Rieder (1994) in combination with Slutzky's lemma, the Cramér-Wold device and Le Cam's third lemma. ////

Remark 1.2.6 Assume transforms $\tau\colon \mathbb{R}^k \to \mathbb{R}^p$ $(p \le k)$ which are differentiable at θ with bounded derivative $D = d\tau(\theta)$ of full rank p,

$$\tau(\theta + t) = \tau(\theta) + Dt + o(|t|) \qquad rkD = p \tag{1.2.16}$$

Then, one gets by the finite-dimensional delta method, setting $\eta_\theta := D\psi_\theta$,

$$\sqrt{n}\,\big(\tau \circ S_n - \tau(\theta)\big)\big(Q_n^n(q,r)\big) \xrightarrow{w} \mathcal{N}_p\big(r\,\mathrm{E}_\theta\,\eta_\theta q,\ \mathrm{Cov}_\theta(\eta_\theta)\big) \tag{1.2.17}$$

////

1.3 Optimally Robust Influence Curves

1.3.1 Introduction

In view of Proposition 1.2.5 and Remark 1.2.6 we obtain the following result.

Proposition 1.3.1 *Let S be an estimator that is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$ and given $q \in Z_\infty(\theta)$ and $r \in (0,\infty)$ consider the simple perturbations $Q_n(q,r)$. Moreover, assume transforms $\tau\colon \mathbb{R}^k \to \mathbb{R}^p$ $(p \le k)$ of form (1.2.16) and let $\eta_\theta = D\psi_\theta$ and*

$$\rho_0(q) = \int \ell\, d\mathcal{N}_p\big(r\,\mathrm{E}_\theta\,\eta_\theta q,\ \mathrm{Cov}_\theta(\eta_\theta)\big) \tag{1.3.1}$$

(a) *If $\ell\colon \mathbb{R}^p \to [0,\infty]$ is lower semicontinuous then for all $r \in (0,\infty)$,*

$$\liminf_{n\to\infty} \int \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) \ge \rho_0(q) \tag{1.3.2}$$

(b) *If $\ell\colon \mathbb{R}^p \to [0,\infty]$ is continuous a.e. λ^p then for all $r \in (0,\infty)$,*

$$\lim_{M\to\infty} \lim_{n\to\infty} \int M \wedge \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) = \rho_0(q) \tag{1.3.3}$$

PROOF Consequence of Proposition 1.2.5 and Remark 1.2.6 in combination with
(a) the Lemma of Fatou in the version of Lemma A.2.1 of Rieder (1994).
(b) the continuous mapping theorem. ////

Remark 1.3.2 The clipping of the function ℓ by M that occurs in Proposition 1.3.1 (b) is only necessary for attaining the lower bound $\rho_0(q)$. ////

This suggests the following limiting risk for AL estimators, in the cases $* = c, v$,

$$\sup_{q\in\mathcal{G}_*(\theta)} \lim_{M\to\infty} \lim_{n\to\infty} \int M \wedge \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) = \sup_{q\in\mathcal{G}_*(\theta)} \rho_0(q) \tag{1.3.4}$$

Choosing quadratic loss $\ell(z) = |z|^2$, one obtains the subsequent asymptotic mean square error (MSE) problems,

$$\operatorname{maxMSE}_\theta(\eta_\theta, r) := E_\theta|\eta_\theta|^2 + r^2\omega_{*,\theta}(\eta_\theta)^2 = \min\,! \qquad \eta_\theta \in \Psi_2^D(\theta) \quad (1.3.5)$$

with

$$\omega_{*,\theta}(\eta_\theta) \;=\; \sup\big\{|\operatorname{E}_\theta \eta_\theta q| \,\big|\, q \in \mathcal{G}_*(\theta)\big\} \quad (1.3.6)$$

where the radius $r \in (0,\infty)$ of the simple perturbations (1.2.13) is fixed. The solution to this optimization problem is given in Subsection 1.3.4. The determination of the solution is based on Langrange multiplier theorems developed in Appendix B of Rieder (1994) and canonically leads to the following Hampel type problem[1], with bound $b \in (0,\infty)$ fixed,

$$E_\theta|\eta_\theta|^2 = \min\,! \qquad \eta_\theta \in \Psi_2^D(\theta),\ \omega_{*,\theta}(\eta_\theta) \le b \quad (1.3.7)$$

Thus, the solution to this Hampel type problem is given beforehand in Subsection 1.3.3. The standardized (infinitesimal) bias terms $\omega_{*,\theta}(\eta_\theta)$ that occur in the optimization problems are more or less explicitly calculated in Subsection 1.3.2.

Remark 1.3.3 **(a)** Actually, one is interested in the following limiting risk

$$\lim_{M\to\infty}\lim_{n\to\infty}\sup_{Q\in U_*(\theta,r/\sqrt{n})}\int M\wedge\ell\big(\sqrt{n}\,(\tau\circ S_n-\tau(\theta))\big)\,dQ^n \quad (1.3.8)$$

Thus, it has to be made sure that at least for the optimal ICs, the interchanging of $\lim_M \lim_n$ and $\sup_q$ and the passage from the neighborhood submodel to full neighborhoods does not increase the asymptotic risk (1.3.4). Under additional assumptions on the optimally robust ICs this goal can be achieved by suitable estimator constructions. In case of one-step estimators we verify that these extra conditions, which are spelled out in Subsection 2.3.1, are fulfilled for exponential families of full rank; confer Subsection 2.3.3. Thus, we can apply the one-step construction to most models considered in this thesis. Exceptions are regression models in combination with conditional neighborhoods and times series models where the construction problem is still unsolved. For more details on stable constructions including M and MD (minimum distance) estimates we refer to Chapter 6 of Rieder (1994).

(b) Since the normal distribution is fully specified by its first two moments, one might, analogously to pp 197 of Fraiman et al. (2001), think of the following general optimization problem

$$\sup_{q\in\mathcal{G}_*(\theta)} g\big(r\operatorname{E}_\theta\eta_\theta q, \operatorname{Cov}_\theta(\eta_\theta)\big) = \min\,! \qquad \eta_\theta \in \Psi_2^D(\theta) \quad (1.3.9)$$

for suitable functions g. By choosing $g(x_1,x_2) = |x_1|^2 + \operatorname{tr}(x_2)$ and $g(x_1,x_2) = \infty\,\mathrm{I}_{\{|x_1|>b\}}(x_1) + \operatorname{tr}(x_2)$, respectively, this problem also covers the MSE and the

[1] in allusion to the problem solved in Lemma 5 of Hampel (1968)

Hampel type problem stated above. Ruckdeschel and Rieder (2004) consider the similar problem

$$G\big(r\omega_*(\eta_\theta), \sqrt{\mathrm{E}_\theta\,|\eta_\theta|^2}\,\big) = \min\,! \qquad \eta_\theta \in \Psi_2^D(\theta) \tag{1.3.10}$$

where G is some positive and convex function which is strictly isotone in both arguments. They show that the solution to (1.3.10) also solves the corresponding Hampel type problem (1.3.7), respectively the corresponding MSE problem (1.3.5) where one only has to transform the bias weight according to the given risk; confer Section 8.1 (ibid.). Using this fact, they derive necessary and sufficient conditions for the optimally robust ICs including an additional equation for the determination of the optimal bias bound b; confer Theorem 3.1 (ibid.). ////

To lighten the notation, we drop the fixed parameter θ and write $\omega_* = \omega_{*,\theta}$ and $\eta = \eta_\theta$ as well as $\mathcal{G}_* = \mathcal{G}_*(\theta)$ and $\Psi_2^D = \Psi_2^D(\theta)$. Moreover, let $\mathrm{E} = \mathrm{E}_\theta$ denote expectation, $\mathrm{Cov} = \mathrm{Cov}_\theta$ covariance and $\inf_P$, $\sup_P$ the essential extrema under $P = P_\theta$.

1.3.2 Bias Terms

The standardized (infinitesimal) bias terms ω_* for $* = c, v$ have the following general properties.

Lemma 1.3.4 *Let $* = c, v$ and $\eta \in L_1^p(P)$. Then,*

$$\omega_*(\eta) = \omega_*(\eta - \mathrm{E}\,\eta) \tag{1.3.11}$$

$$\omega_*(\eta) = \sup\big\{\omega_*(e^\tau\eta)\,\big|\,e \in \mathbb{R}^p,\ |e| = 1\big\} \tag{1.3.12}$$

$$\omega_c(\eta) \le \omega_v(\eta) \le 2\omega_c(\eta) \tag{1.3.13}$$

The terms ω_* are positively homogeneous, subadditive, hence convex on $L_1^p(P)$, and weakly lower semicontinuous on $L_2^p(P)$.

Proof Rieder (1994), Lemma 5.3.2. ////

One gets the following explicit expressions for ω_*.

Proposition 1.3.5 *Let $\eta \in L_1(P)$ with $\mathrm{E}\,\eta = 0$. Then,*

$$\omega_c(\eta) = \sup_P |\eta| \tag{1.3.14}$$

$$\omega_v(\eta) = \sup\big\{\sup_P e^\tau\eta - \inf_P e^\tau\eta\,\big|\,e \in \mathbb{R}^p,\ |e| = 1\big\} \tag{1.3.15}$$

Proof Rieder (1994), Proposition 5.3.3 (a). ////

Remark 1.3.6 For $\eta \in L_1(P)$ such that η is bounded and $\mathrm{E}\,\eta = 0$, it turns out that the standardized bias terms evaluated over full contamination and total variation balls do not exceed ω_* by more than the increase of some P essential to pointwise extrema; confer Lemma 5.3.4 of Rieder (1994). ////

1.3.3 Minimum Trace Subject to Bias Bound

In this section we give the unique solutions to the Hampel type problem (1.3.7). For various aspects of this problem confer pp 196 of Rieder (1994). We first give the unique solution for $* = c$.

Theorem 1.3.7 (a) *In case* $\omega_c^{\min} < b \le \omega_c(\eta_h)$, *there exist some* $a \in \mathbb{R}^p$ *and* $A \in \mathbb{R}^{p\times k}$ *such that the solution is of the form*

$$\tilde\eta = (A\Lambda - a)w \qquad w = \min\left\{1, \frac{b}{|A\Lambda - a|}\right\} \tag{1.3.16}$$

Conversely, if some $\tilde\eta \in \Psi_2^D$ *is of form (1.3.16) for any* $b \in (0,\infty)$, $a \in \mathbb{R}^p$, *and* $A \in \mathbb{R}^{p\times k}$, *then* $\tilde\eta$ *is the solution, and the following representations hold,*

$$a = Az \qquad 0 = \mathrm{E}(\Lambda - z)w \qquad D = A\,\mathrm{E}(\Lambda - z)(\Lambda - z)^\tau w \tag{1.3.17}$$

where $AD^\tau = DA^\tau \succ 0$.

(b) *It holds that*

$$\omega_c^{\min} = \max\left\{\frac{\operatorname{tr} AD^\tau}{\mathrm{E}\,|A\Lambda - a|}\,\middle|\, a \in \mathbb{R}^p,\ A \in \mathbb{R}^{p\times k}\setminus\{0\}\right\} \tag{1.3.18}$$

There exist $a \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times k}\setminus\{0\}$ *and* $\bar\eta \in \Psi_2^D$ *achieving* $\omega_c^{\min} = b$, *respectively. And then necessarily*

$$\bar\eta = b\frac{A\Lambda - a}{|A\Lambda - a|} \qquad \text{on } \{A\Lambda \ne a\} \tag{1.3.19}$$

Moreover, $a = Az$ *for some* $z \in \mathbb{R}^k$, *and* $AD^\tau = DA^\tau \succeq 0$.

If $\bar\eta$ *in addition is constant on* $\{A\Lambda = a\}$, *then it is the solution.*

Proof Rieder (1994), Theorem 5.5.1. ////

Remark 1.3.8 (a) In dimension $k = 1$ we obtain

$$\omega_c^{\min} = \frac{|D|}{\mathrm{E}\,|\Lambda - m|} \tag{1.3.20}$$

with any $m = \mathrm{med}(\Lambda)$ and the solution $\bar\eta$ reads

$$\bar\eta = \omega_c^{\min}\operatorname{sign}(D)\Big[\mathrm{I}(\Lambda > m) - \mathrm{I}(\Lambda < m) + \beta\,\mathrm{I}(\Lambda = m)\Big] \tag{1.3.21}$$

with

$$\beta = \Big[P(\Lambda < m) - P(\Lambda > m)\Big]\Big/P(\Lambda = m) \tag{1.3.22}$$

where $|\beta| \le 1$.

(b) By Lemma 1.3.4 and Proposition 1.3.5 we have

$$\omega_c(\eta) = \sup\left\{\omega_c(e^\tau\eta)\,\middle|\, e \in \mathbb{R}^k,\ |e| = 1\right\} = \sup{}_P|\eta| \tag{1.3.23}$$

for $\eta \in \Psi_2$. Now, consider $\chi := e^\tau \eta$; i.e., $\chi \in \Psi_2^D$ with $D = e^\tau$. Then,

$$\omega_c(\chi) \geq \max \left\{ \frac{\operatorname{tr} Ae}{\mathrm{E}\,|A\Lambda - a|} \;\middle|\; a \in \mathbb{R},\ A \in \mathbb{R}^{1\times k} \setminus \{0\} \right\} \tag{1.3.24}$$

By choosing $A = e^\tau$ it follows,

$$\omega_c(\chi) \geq \max \left\{ \frac{1}{\mathrm{E}\,|e^\tau\Lambda - a|} \;\middle|\; a \in \mathbb{R} \right\} = \frac{1}{\mathrm{E}\,|e^\tau\Lambda - \operatorname{med}(e^\tau\Lambda)|} \tag{1.3.25}$$

Consequently, we obtain the following lower bound

$$\omega_c(\eta) \geq \sup_{e\in\mathbb{R}^k,\,|e|=1} \frac{1}{\mathrm{E}\,|e^\tau\Lambda - \operatorname{med}(e^\tau\Lambda)|} = \frac{1}{\inf\limits_{e\in\mathbb{R}^k,\,|e|=1} \mathrm{E}\,|e^\tau\Lambda - \operatorname{med}(e^\tau\Lambda)|} \tag{1.3.26}$$

By choosing $e^\tau = (0,\ldots,1,\ldots,0)$ with 1 at position $1 \leq i \leq k$ this entails

$$\omega_c(\eta) \geq \max_{1\leq i\leq k} \frac{1}{\mathrm{E}\,|\Lambda_i - \operatorname{med}(\Lambda_i)|} = \frac{1}{\min\limits_{1\leq i\leq k} \mathrm{E}\,|\Lambda_i - \operatorname{med}(\Lambda_i)|} \tag{1.3.27}$$

Moreover,

$$\inf_{e\in\mathbb{R}^k,\,|e|=1} \mathrm{E}\,|e^\tau\Lambda - \operatorname{med}(e^\tau\Lambda)| \leq \inf_{e\in\mathbb{R}^k,\,|e|=1} \mathrm{E}\,|e^\tau\Lambda| \tag{1.3.28}$$

$$\leq \inf_{e\in\mathbb{R}^k,\,|e|=1} \Big(\mathrm{E}(e^\tau\Lambda)^2\Big)^{1/2} \tag{1.3.29}$$

$$\leq \big(\operatorname{minev}\mathcal{I}\big)^{1/2} \tag{1.3.30}$$

where $\operatorname{minev}\mathcal{I}$ is the minimal eigenvalue of $\mathcal{I}$. That is,

$$\omega_c(\eta) \geq \frac{1}{\big(\operatorname{minev}\mathcal{I}\big)^{1/2}} \tag{1.3.31}$$

////

Since the exact bias terms for $* = v$ and $p > 1$ are difficult to handle (c.f. Rieder (1994), p 205), the Hampel type problem is only solved for $p = 1$.

Theorem 1.3.9 (a) *In case $\omega_v^{\min} < b \leq \omega_v(\eta_h)$, there exist some $c \in (-b, 0)$ and $A \in \mathbb{R}^{1\times k} \setminus \{0\}$ such that*

$$\tilde\eta = c \vee A\Lambda \wedge (c + b) \tag{1.3.32}$$

is the solution, and

$$\omega_v(\tilde\eta) = b \tag{1.3.33}$$

Conversely, if some $\tilde\eta \in \Psi_2^D$ is of form (1.3.32) for any $b \in (0,\infty)$, $c \in (-b, 0)$, and $A \in \mathbb{R}^{1\times k}$, then $\tilde\eta$ is the solution, and the following representations hold,

$$\mathrm{E}(c - A\Lambda)_+ = \mathrm{E}\big(A\Lambda - (c + b)\big)_+ \qquad D = \mathrm{E}\big[c \vee A\Lambda \wedge (c + b)\big]\Lambda^\tau \tag{1.3.34}$$

where $AD^\tau = DA^\tau > 0$.

(b) *It holds that*

$$\omega_v^{\min} = \max\left\{ \frac{AD^\tau}{\mathrm{E}(A\Lambda)_+} \;\middle|\; A \in \mathbb{R}^{1\times k} \setminus \{0\} \right\} \tag{1.3.35}$$

There exist $A \in \mathbb{R}^{1\times k} \setminus \{0\}$ *and* $\bar\eta \in \Psi_2^D$ *achieving* $\omega_v^{\min} = b$, *respectively. And then necessarily*

$$\bar\eta\,\mathrm{I}(A\Lambda \neq 0) = c\,\mathrm{I}(A\Lambda < 0) + (c+b)\,\mathrm{I}(A\Lambda > 0) \tag{1.3.36}$$

for some $c \in (-b, 0)$. *In the case* $k = 1$, *the solution is*

$$\bar\eta = b\,\mathrm{sign}(D)\left(\frac{P(\Lambda<0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda>0) - \frac{P(\Lambda>0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda<0) \right) \tag{1.3.37}$$

Proof Rieder (1994), Theorem 5.5.5. ////

Remark 1.3.10 **(a)** In dimension $k = 1$ we obtain

$$\omega_v^{\min} = \frac{|D|}{\mathrm{E}(\Lambda)_+} \tag{1.3.38}$$

(b) By Lemma 1.3.4 and Proposition 1.3.5 we have

$$\omega_v(\eta) = \sup\left\{ \omega_v(e^\tau\eta) \,\middle|\, e \in \mathbb{R}^k,\, |e| = 1 \right\} \tag{1.3.39}$$

for $\eta \in \Psi_2$. Now, consider $\chi := e^\tau\eta$; i.e., $\chi \in \Psi_2^D$ with $D = e^\tau$. Then,

$$\omega_v(\chi) \geq \max\left\{ \frac{\mathrm{tr}\,Ae}{\mathrm{E}(A\Lambda)_+} \;\middle|\; A \in \mathbb{R}^{1\times k} \setminus \{0\} \right\} \tag{1.3.40}$$

By choosing $A = e^\tau$ it follows,

$$\omega_c(\chi) \geq \frac{1}{\mathrm{E}(e^\tau\Lambda)_+} \tag{1.3.41}$$

Hence, we obtain the following lower bound

$$\omega_v(\eta) \geq \sup_{e\in\mathbb{R}^k,\,|e|=1} \frac{1}{\mathrm{E}(e^\tau\Lambda)_+} = \frac{1}{\inf\limits_{e\in\mathbb{R}^k,\,|e|=1} \mathrm{E}(e^\tau\Lambda)_+} \tag{1.3.42}$$

By choosing $e^\tau = (0,\ldots,1,\ldots,0)$ with 1 at position $1 \leq i \leq k$ this yields

$$\omega_v(\eta) \geq \max_{1\leq i\leq k} \frac{1}{\mathrm{E}(\Lambda_i)_+} = \frac{1}{\min\limits_{1\leq i\leq k} \mathrm{E}(\Lambda_i)_+} \tag{1.3.43}$$

Since, $\omega_v(\eta) \geq \omega_c(\eta)$ it also holds

$$\omega_v(\eta) \geq \frac{1}{\left(\mathrm{minev}\,\mathcal{I}\right)^{1/2}} \tag{1.3.44}$$

where $\mathrm{minev}\,\mathcal{I}$ is the minimal eigenvalue of $\mathcal{I}$. ////

1.3.4 Mean Square Error

In this section we give the solutions to the MSE problem (1.3.5).

Theorem 1.3.11 **(a)** *The solutions to the MSE problem (1.3.5) for* $* = c$ *and* $(* = v,\ p = 1)$*, respectively, are unique.*

(b) *The solution to problem (1.3.5) and* $* = c$ *coincides with the solution to problem (1.3.7) and* $* = c$*, with* $b \in (0, \infty)$ *and* $r \in (0, \infty)$ *related by*

$$r^2 b = \mathrm{E}\left(|A\Lambda - a| - b\right)_+ \tag{1.3.45}$$

(c) *The solution to problem (1.3.5) and* $(* = v,\ p = 1)$ *coincides with the solution to problem (1.3.7) and* $(* = v,\ p = 1)$*, with* $b \in (0, \infty)$ *and* $r \in (0, \infty)$ *related by*

$$r^2 b = \mathrm{E}\left(c - A\Lambda\right)_+ \tag{1.3.46}$$

Proof Rieder (1994), Theorem 5.5.7. ////

Chapter 2

Supplements to the Asymptotic Theory of Robustness

This chapter provides supplementary results which mainly concern the mean square error solutions specified in Subsection 1.3.4. In Subsection 2.1.1 we show that there is an analogy between the matrix A included in the optimally robust ICs and the inverse Fisher information. We then prove that the minimum bias solution (under additional assumptions) can be the solution to the corresponding asymptotic MSE problem for finite radius r; confer Subsection 2.1.2. The subsequent Subsections 2.1.3–2.1.5 contain results about the boundedness, uniqueness and continuity of the Lagrange multipliers contained in the optimal ICs. In Section 2.2 we introduce the notions of least favorable radius and radius–minimax estimator. The construction of optimally robust estimators by means of one-step constructions is considered in Section 2.3 where general sufficient conditions are given in Subsection 2.3.1. Based on these we derive sufficient conditions for the MSE optimal ICs (cf. Subsection 2.3.2) and verify them in case of exponential families of full rank (cf. Subsection 2.3.3). Finally, we state some results about the convergence of robust models; confer Section 2.4.

2.1 Mean Square Error Solution

2.1.1 Matrix A – an Analogue to the Inverse Fisher Information

The following result may, in terms of statistical risk, be interpreted as an extension of the classical Cramér-Rao bound under quadratic loss, in which case $\operatorname{tr} AD^\tau = \operatorname{tr} D\mathcal{I}^{-1}D^\tau$.

Proposition 2.1.1 *Given some radius* $r \in (0,\infty)$, *the solution* $\tilde{\eta}$ *to the MSE problem (1.3.5) for* $* = c$ *and* $* = v$, $p = 1$, *respectively, fulfills*

$$\operatorname{maxMSE}(\tilde{\eta}, r) = \operatorname{tr} AD^\tau \tag{2.1.1}$$

PROOF

$\boldsymbol{* = c}$**:** We define $Y = A\Lambda - a$. Then

$$\begin{aligned}
\operatorname{maxMSE}(\tilde{\eta}, r) &= \mathrm{E}\,|\tilde{\eta}|^2 + r^2\omega_c^2(\tilde{\eta}) \\
&= \mathrm{E}\,|Y|^2 w^2 + r^2 b^2 \\
&= \mathrm{E}\,Y^\tau \tilde{\eta} - \mathrm{E}\,Y^\tau Y w(1-w) + r^2 b^2 \\
&\overset{(1.3.17)}{=} \operatorname{tr} AD^\tau - \mathrm{E}\,Y^\tau Y w(1-w) + r^2 b^2 \qquad (2.1.2)
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{E}\,Y^\tau Y w(1-w) &= \mathrm{E}\,|Y|^2 \frac{b}{|Y|}\left(1 - \frac{b}{|Y|}\right)_+ \\
&= b\,\mathrm{E}\,\big(|Y| - b\big)_+ \qquad (2.1.3) \\
&\overset{(1.3.45)}{=} r^2 b^2 \qquad (2.1.4)
\end{aligned}$$

$\boldsymbol{* = v}$**,** $\boldsymbol{p = 1}$**:** We define $Y = A\Lambda$ and

$$w = 1 \wedge \max\left\{\frac{c}{Y}, \frac{c+b}{Y}\right\} \tag{2.1.5}$$

Then, $\tilde{\eta} = Yw$ and we get analogously to the case $* = c$

$$\operatorname{maxMSE}(\tilde{\eta}, r) \overset{(1.3.34)}{=} AD^\tau - \mathrm{E}\,Y^2 w(1-w) + r^2 b^2 \tag{2.1.6}$$

with

$$\begin{aligned}
\mathrm{E}\,Y^2 w(1-w) &= \mathrm{E}\,Y^2 \max\left\{\frac{c}{Y}, \frac{c+b}{Y}\right\}\left(1 - \max\left\{\frac{c}{Y}, \frac{c+b}{Y}\right\}\right)_+ \\
&= (c+b)\,\mathrm{E}\,\big(Y - (c+b)\big)_+ - c\,\mathrm{E}\,\big(c - Y\big)_+ \\
&\overset{(1.3.34)}{=} b\,\mathrm{E}\,\big(c - Y\big)_+ \qquad (2.1.7) \\
&\overset{(1.3.46)}{=} r^2 b^2 \qquad (2.1.8)
\end{aligned}$$

////

Remark 2.1.2 This correspondence for the minimax asymptotic MSE holds more generally and can be verified for the cases $* = c, v$, $t = 0, \varepsilon, \alpha$, $s = 0, e, 2$ considered in Rieder (1994). Exceptions are the cases $* = h, t = 0, s = 0, e$ and $* = h, t = \alpha = 2, s = e$, where the optimally robust ICs are identical to η_h and $\operatorname{maxMSE}(\eta_h, r) = \operatorname{tr} D\mathcal{I}^{-1}D^\tau + r^2 b^2$. ////

2.1.2 Discrete Models and the Gap Condition

Under the assumptions of the following proposition the lower case $\bar{\eta}$ is the solution to the MSE problem (1.3.5) for finite radii $r \in [0, \infty)$ and $* = c$.

Proposition 2.1.3 *Let $* = c$, dimension $k = 1$ and m any $\mathrm{med}(\Lambda)$. Moreover, define*

$$\gamma = \inf_P \left\{ |\Lambda - m| \,\middle|\, |\Lambda - m| > 0 \right\} \tag{2.1.9}$$

$$\gamma_1 = \inf_P \left\{ (\Lambda - m) \,\middle|\, \Lambda > m \right\} \tag{2.1.10}$$

and

$$\gamma_2 = \sup_P \left\{ (\Lambda - m) \,\middle|\, \Lambda < m \right\} \tag{2.1.11}$$

(a) *Assume $P(\Lambda = m) > 0$ and let $\bar{\eta}$ be the lower case solution given in Theorem 1.3.7 (b). Then, the following statements are equivalent:*

(i) $\bar{\eta}$ is the solution to the MSE problem (1.3.5) with radius r.

(ii) $\bar{\eta} = \rho$ a.e. P where ρ is of form (1.3.16) and solves the MSE equation (1.3.45) for radius r.

(iii) $\gamma > 0$ and

$$r \geq \left[M \,\mathrm{E}\,|\Lambda - m| - P(\Lambda \neq m) - \beta^2 P(\Lambda = m) \right]^{1/2} =: \bar{r} \tag{2.1.12}$$

where β is defined in (1.3.22) and

$$M = \begin{cases} \frac{1+|\beta|}{\gamma} & \text{if } P(\Lambda < m) = 0 \text{ or } P(\Lambda > m) = 0 \\ \max\left\{ \frac{1-\beta}{\gamma_1}, \frac{1+\beta}{-\gamma_2} \right\} & \text{else} \end{cases} \tag{2.1.13}$$

(b) *Assume $P(\Lambda = m) = 0$ and let $\bar{\eta}$ be the lower case solution given in Theorem 1.3.7 (b). Then, the following statements are equivalent:*

(i) $\bar{\eta}$ is the solution to the MSE problem (1.3.5) with radius r.

(ii) $\bar{\eta} = \rho$ a.e. P where ρ is of form (1.3.16) and solves the MSE equation (1.3.45) for radius r.

(iii) $\gamma_1 > 0$ or $\gamma_2 < 0$ and

$$r \geq \left[M \,\mathrm{E}\,|\Lambda - m| - 1 \right]^{1/2} =: \bar{r} \qquad M = \tfrac{2}{\gamma_1 - \gamma_2} \tag{2.1.14}$$

Proof

(a) Assume $P(\Lambda = m) > 0$ and let $\gamma > 0$ and $r \geq \bar{r}$. We consider

$$\rho(\Lambda) = A(\Lambda - z)w \qquad w = \min\left\{1, \frac{\omega_c^{\min}}{|A(\Lambda - z)|}\right\} \tag{2.1.15}$$

with

$$AD = \left(\omega_c^{\min}\right)^2 \left[r^2 + P(\Lambda \neq m) + \beta^2 P(\Lambda = m) \right] \tag{2.1.16}$$

and

$$z = m - A^{-1} \omega_c^{\min} \beta \,\mathrm{sign}(D) \tag{2.1.17}$$

Since $\operatorname{sign}(A) = \operatorname{sign}(D)$, we obtain $AD \geq \omega_c^{\min} M|D|$, respectively $|A| \geq \omega_c^{\min} M$ by the definition of $\bar{r}$. Thus, if $P(\Lambda > m) > 0$ and $P(\Lambda < m) > 0$, then a.e. P,

$$|A|(\Lambda - z) = |A|(\Lambda - m) + \omega_c^{\min}\beta \geq \omega_c^{\min}(M\gamma_1 + \beta) \geq \omega_c^{\min} \qquad \text{if } \Lambda > m \tag{2.1.18}$$

$$|A|(\Lambda - z) = |A|(\Lambda - m) + \omega_c^{\min}\beta \leq \omega_c^{\min}(M\gamma_2 + \beta) \leq -\omega_c^{\min} \qquad \text{if } \Lambda < m \tag{2.1.19}$$

which implies $\rho = \bar{\eta}$ a.e. P on $\{\Lambda \neq m\}$. If $P(\Lambda > m) = 0$ or $P(\Lambda < m) = 0$, respectively, we only need to take into consideration (2.1.18) or (2.1.19), respectively where we have $\beta < 0$, respectively $\beta > 0$; i.e., $\rho = \bar{\eta}$ a.e. P on $\{\Lambda \neq m\}$. In addition, $\rho(m) = \omega_c^{\min}\beta \operatorname{sign}(D) = \bar{\eta}(m)$ as $|\beta| \leq 1$; i.e., $\rho = \bar{\eta}$ a.e. P. Thus, $\rho \in \Psi_2^D$ and is of form (1.3.16). Moreover,

$$\begin{aligned}
\omega_c^{\min} \operatorname{E}(|A(\Lambda - z)| - \omega_c^{\min})_+ &\overset{(2.1.3)}{=} \operatorname{E} A^2|\Lambda - z|^2 w(1-w) \\
&= A \operatorname{E} \rho\Lambda - \operatorname{E}\rho^2 \\
&= AD - \big(\omega_c^{\min}\big)^2\big[P(\Lambda \neq m) + \beta^2 P(\Lambda = m)\big] \\
&\overset{(2.1.16)}{=} r^2\big(\omega_c^{\min}\big)^2
\end{aligned} \tag{2.1.20}$$

which is equivalent to the MSE equation (1.3.45). Consequentially, ρ, respectively $\bar{\eta}$ is the solution to the MSE problem (1.3.5) with radius $r \geq \bar{r}$.

Conversely, under the condition $P(\Lambda = m) > 0$, z has to be of form (2.1.17), otherwise the corresponding $\rho \in \Psi_2^D$ of form (1.3.16) can not fulfill $\rho = \bar{\eta}$ a.e. P. In case $\gamma = 0$, i.e., $\gamma_1 = 0$ or $\gamma_2 = 0$, respectively, then $P(\Lambda < m) < 0.5$ or $P(\Lambda > m) < 0.5$, respectively and therefore $\beta < 1$ or $\beta > -1$, respectively. But, then for any $A \in \mathbb{R} \setminus \{0\}$ there exists a subset $T_1 \subset \{\Lambda > m\}$, respectively $T_2 \subset \{\Lambda < m\}$ such that $P(T_1) > 0$, respectively $P(T_2) > 0$ and

$$|A|(\Lambda - m) + \omega_c^{\min}\beta < \omega_c^{\min} \qquad \text{on } T_1 \tag{2.1.21}$$

respectively

$$|A|(\Lambda - m) + \omega_c^{\min}\beta > -\omega_c^{\min} \qquad \text{on } T_2 \tag{2.1.22}$$

i.e., there is no $\rho \in \Psi_2^D$ of form (1.3.16) such that $\rho = \bar{\eta}$ a.e. P.
In case $\gamma > 0$, $\bar{\eta}$ can be rewritten as some ρ of form (1.3.16), only if $|A| \geq \omega_c^{\min} M$. Otherwise,

$$|A|\gamma_1 + \omega_c^{\min}\beta < \omega_c^{\min} \qquad \text{or} \qquad |A|\gamma_2 + \omega_c^{\min}\beta > -\omega_c^{\min} \tag{2.1.23}$$

respectively with positive probability; confer (2.1.18) and (2.1.19). But, if $|A| \geq \omega_c^{\min} M$, the corresponding ρ can not satisfy the MSE equation (1.3.45) in case $r < \bar{r}$; confer (2.1.20). That is, $\bar{\eta}$ cannot be the solution to the MSE problem (1.3.5) with radius r if $r < \bar{r}$.

(b) Under the condition $P(\Lambda = m) = 0$, we necessarily get $P(\Lambda > m) > 0$ and $P(\Lambda < m) > 0$.
If, $\gamma > 0$ and $r \geq \bar{r}$, we consider

$$\rho(\Lambda) = A(\Lambda - z)w \qquad w = \min\Big\{1, \frac{\omega_c^{\min}}{|A(\Lambda - z)|}\Big\} \tag{2.1.24}$$

with

$$AD = \big(\omega_c^{\min}\big)^2(r^2 + 1) \qquad \text{and} \qquad z = m + \tfrac{1}{2}(\gamma_1 + \gamma_2) \tag{2.1.25}$$

which implies $|A| \geq \omega_c^{\min} M$ by the definition of $\bar{r}$. Consequentially, a.e. P,

$$\begin{aligned} |A|(\Lambda - z) &= |A|\big(\Lambda - m - \tfrac{1}{2}(\gamma_1 + \gamma_2)\big) \\ &\geq \omega_c^{\min} M\big(\gamma_1 - \tfrac{1}{2}(\gamma_1 + \gamma_2)\big) = \omega_c^{\min} \qquad \text{if } \Lambda > m \qquad (2.1.26) \\ |A|(\Lambda - z) &= |A|\big(\Lambda - m - \tfrac{1}{2}(\gamma_1 + \gamma_2)\big) \\ &\leq \omega_c^{\min} M\big(\gamma_2 - \tfrac{1}{2}(\gamma_1 + \gamma_2)\big) = -\omega_c^{\min} \qquad \text{if } \Lambda < m \qquad (2.1.27) \end{aligned}$$

i.e., $\rho = \bar{\eta}$ a.e. P. In addition, we get analogously to (2.1.20)

$$\omega_c^{\min} \operatorname{E}(|A(\Lambda - z)| - \omega_c^{\min})_+ = AD - \big(\omega_c^{\min}\big)^2 \overset{(2.1.25)}{=} r^2\big(\omega_c^{\min}\big)^2 \qquad (2.1.28)$$

which is equivalent to the MSE equation (1.3.45). Thus, ρ, respectively $\bar{\eta}$ is the solution to the MSE problem (1.3.5) with radius $r \geq \bar{r}$.

Conversely, we have to choose $A \in \mathbb{R} \setminus \{0\}$ and $z \in \mathbb{R}$ such that a.e. P,

$$\begin{aligned} |A|(\Lambda - z) &\geq |A|(\gamma_1 + m - z) \geq \omega_c^{\min} \qquad \text{if } \Lambda > m \qquad (2.1.29) \\ |A|(\Lambda - z) &\leq |A|(\gamma_2 + m - z) \leq -\omega_c^{\min} \qquad \text{if } \Lambda < m \qquad (2.1.30) \end{aligned}$$

which is not possible if $\gamma_1 = 0$ and $\gamma_2 = 0$; i.e., in this case there is no $\rho \in \Psi_2^D$ of form (1.3.16) such that $\rho = \bar{\eta}$ a.e. P. Thus, $\bar{\eta}$ cannot be the solution to the MSE problem (1.3.5). However, if $\gamma_1 > 0$ or $\gamma_2 < 0$, this leads us to

$$|A| \geq \omega_c^{\min} \max\left\{\frac{1}{\gamma_1 + m - z}, \frac{1}{-\gamma_2 - m + z}\right\} \qquad \text{a.e. } P \qquad (2.1.31)$$

which is minimized by z as given in (2.1.25). For any other $z \in \mathbb{R}$ we obtain $|A| > \omega_c^{\min} M$; i.e., a larger maximum asymptotic MSE which corresponds to a larger radius r; confer (2.1.28). Consequentially, $\bar{\eta}$ can be rewritten as some ρ of form (1.3.16), only if $|A| \geq \omega_c^{\min} M$. But, if $|A| \geq \omega_c^{\min} M$, the corresponding ρ can not satisfy the MSE equation (1.3.45) in case $r < \bar{r}$; confer (2.1.28). Therefore, $\bar{\eta}$ cannot be the solution to the MSE problem (1.3.5) with radius r if $r < \bar{r}$. ////

Remark 2.1.4 **(a)** Since the lower case solution is also the solution to the MSE problem (1.3.5) in the setup of the preceding proposition, we call $\bar{r}$ *lower case radius*. The necessary condition $\gamma > 0$, respectively $\gamma_1 > 0$ or $\gamma_2 < 0$ is called *gap condition*.

(b) In case $\gamma_1 = -\gamma_2 = \gamma > 0$, the parts (a) and (b) of the preceding proposition coincide and we obtain $M = \frac{1+|\beta|}{\gamma}$ in all cases.

(c) In particular, Proposition 2.1.3 shows the centering constant a included in the solution (1.3.16) to the MSE problem (1.3.5) and $* = c$ is non-unique if $\operatorname{med}(\Lambda)$ is non-unique and $r \geq \bar{r}$. In case $r < \bar{r}$ and $\operatorname{med}(\Lambda)$ is non-unique, $\tilde{\eta}$ cannot attain only two points with probability 1 (otherwise $\tilde{\eta} = \bar{\eta}$). Thus, the uniqueness of A and b entails the uniqueness of a via $\operatorname{E}\tilde{\eta} = 0$ in this case. ////

The subsequent proposition is the analogue to Proposition 2.1.3 in case $* = v$.

Proposition 2.1.5 *Let $* = v$, dimension $k = 1$ and $\bar{\eta}$ be the lower case solution given in Theorem 1.3.9 (b). Then, the following statements are pairwise equivalent:*

(i) *$\bar{\eta}$ is the solution to the MSE problem (1.3.5) with radius r.*

(ii) *$\bar{\eta} = \rho$ a.e. P where ρ is of form (1.3.32) and solves the MSE equation (1.3.46) for radius r.*

(iii)

$$\inf_P \{|\Lambda| \,|\, |\Lambda| > 0\} =: \gamma > 0 \tag{2.1.32}$$

and

$$r \geq \left[M \, \mathrm{E}(\Lambda)_+ - \frac{P(\Lambda > 0)P(\Lambda < 0)}{P(\Lambda \neq 0)} \right]^{1/2} =: \bar{r} \tag{2.1.33}$$

where

$$M = \frac{1}{P(\Lambda \neq 0)} \max \left\{ \frac{P(\Lambda < 0)}{\gamma_1}, \frac{P(\Lambda > 0)}{-\gamma_2} \right\} \tag{2.1.34}$$

with

$$\gamma_1 = \inf_P \{\Lambda \,|\, \Lambda > 0\} \qquad \gamma_2 = \sup_P \{\Lambda \,|\, \Lambda < 0\} \tag{2.1.35}$$

Proof If $\gamma > 0$ and $r \geq \bar{r}$, we consider

$$\rho = \mathrm{sign}(D) \left[\left(-\omega_v^{\min} \frac{P(\Lambda > 0)}{P(\Lambda \neq 0)} \right) \vee |A|\Lambda \wedge \left(\omega_v^{\min} \frac{P(\Lambda < 0)}{P(\Lambda \neq 0)} \right) \right] \tag{2.1.36}$$

with

$$AD = \left(\omega_v^{\min}\right)^2 \left[r^2 + \frac{P(\Lambda > 0)P(\Lambda < 0)}{P(\Lambda \neq 0)} \right] \tag{2.1.37}$$

Since $\mathrm{sign}(A) = \mathrm{sign}(D)$, we obtain $AD \geq \omega_c^{\min} M |D|$, respectively $|A| \geq \omega_c^{\min} M$ by the definition of $\bar{r}$. Thus, a.e. P,

$$|A|\Lambda \geq \omega_v^{\min} M \gamma_1 \geq \omega_v^{\min} \frac{P(\Lambda < 0)}{P(\Lambda \neq 0)} \qquad \text{if } \Lambda > 0 \tag{2.1.38}$$

$$|A|\Lambda \leq -\omega_v^{\min} M \gamma_2 \leq -\omega_v^{\min} \frac{P(\Lambda > 0)}{P(\Lambda \neq 0)} \qquad \text{if } \Lambda < 0 \tag{2.1.39}$$

That is, $\rho = \bar{\eta}$ a.e. P as $\rho(0) = 0 = \bar{\eta}(0)$ no matter if $P(\Lambda = 0) > 0$. Thus, $\rho \in \Psi_2^D$ and is of form (1.3.32). Defining

$$c := -\omega_v^{\min} \begin{cases} \dfrac{P(\Lambda > 0)}{P(\Lambda \neq 0)} & \text{if } \mathrm{sign}(D) = 1 \\ \dfrac{P(\Lambda < 0)}{P(\Lambda \neq 0)} & \text{if } \mathrm{sign}(D) = -1 \end{cases} \tag{2.1.40}$$

we obtain

$$\begin{aligned} \mathrm{E}(c - A\Lambda)_+ &= \mathrm{E}(c - A\Lambda) I(A\Lambda \leq c) \\ &\overset{(2.1.38),(2.1.39)}{=} \mathrm{E}(c - A\Lambda) I(A\Lambda < 0) \\ &= cP(\mathrm{sign}(D)\Lambda < 0) + \mathrm{E}(A\Lambda)_- \\ &\overset{(2.1.37)}{=} r^2 \omega_v^{\min} \end{aligned} \tag{2.1.41}$$

since $\mathrm{E}(A\Lambda)_- = \mathrm{E}(A\Lambda)_+ = AD/\omega_v^{\min}$; i.e., the MSE equation (1.3.46) holds. Consequentially, ρ, respectively $\bar\eta$ is the solution to the MSE problem (1.3.5) with radius $r \geq \bar r$.

Conversely, if $\gamma = 0$ (i.e., $\gamma_1 = 0$ or/and $\gamma_2 = 0$), then for any $A \in \mathbb{R}\backslash\{0\}$ there exists a subset $T_1 \subset \{\Lambda > 0\}$, respectively $T_2 \subset \{\Lambda < 0\}$ such that $P(T_1) > 0$, respectively $P(T_2) > 0$ and

$$|A|\Lambda < \omega_v^{\min}\frac{P(\Lambda < 0)}{P(\Lambda \neq 0)} \qquad \text{on } T_1 \tag{2.1.42}$$

respectively

$$|A|\Lambda > -\omega_v^{\min}\frac{P(\Lambda > 0)}{P(\Lambda \neq 0)} \qquad \text{on } T_2 \tag{2.1.43}$$

i.e., there is no $\rho \in \Psi_2^D$ of form (1.3.16) such that $\rho = \bar\eta$ a.e. P. Moreover, $\bar\eta$ can be rewritten as some $\rho \in \Psi_2^D$ of form (1.3.32), only if $|A| \geq \omega_v^{\min} M$. Otherwise,

$$|A|\gamma_1 < \omega_v^{\min}\frac{P(\Lambda < 0)}{P(\Lambda \neq 0)} \qquad \text{or} \qquad |A|\gamma_2 > -\omega_v^{\min}\frac{P(\Lambda > 0)}{P(\Lambda \neq 0)} \tag{2.1.44}$$

respectively with positive probability; confer (2.1.38) and (2.1.39). However, in case $|A| \geq \omega_v^{\min} M$, the corresponding ρ can not satisfy the MSE equation (1.3.46) if $r < \bar r$; confer (2.1.41). That is, $\bar\eta$ cannot be the solution to the MSE problem (1.3.5) with radius r if $r < \bar r$. ////

Remark 2.1.6 **(a)** In case the gap condition is not fulfilled (i.e., $\gamma = 0$, respectively $\gamma_1 = 0$ and $\gamma_2 = 0$), we obtain $M = \infty$ and therefore also $\bar r = \infty$. That is, in any case the lower case radius $\bar r$ represents the minimal radius for which the lower case solution $\bar\eta$ also solves the corresponding MSE problem ($* = c, v$).

(b) Proposition 2.1.3 and Proposition 2.1.5 for instance apply to the binomial and Poisson models considered in Chapters 3 and 4. In particular, we obtain $\bar r = 0$ in case of $\mathrm{Binom}(1,\theta)$ ($\theta \in (0,1)$) and $\mathrm{Binom}(2, 0.5)$, respectively; i.e., the lower case solution $\bar\eta$ is the MSE optimal IC for all radii $r \in [0,\infty]$. However, in these cases there is only one IC. That is, the lower case solution coincides with the classical optimal IC η_h; confer Remark 3.1.3.

(c) If we consider the Hampel type problem (1.3.7), the proofs of Proposition 2.1.3 and Proposition 2.1.5 show that the Lagrange multipliers contained in the corresponding solutions are non-unique in case $b = \omega_c^{\min}$ and $\gamma > 0$, respectively $b = \omega_v^{\min}$ and $\gamma_1 > 0$ or $\gamma_2 < 0$. In these cases any $A \geq M$ leads to some ρ of the optimal form (1.3.16), respectively form (1.3.32) and therefore is the solution to the corresponding Hampel type problem (1.3.7).

(d) We conjecture that in some (discrete) models and under similar conditions the lower case can be the solution to the MSE problem (1.3.5) for finite radius also in dimension $k > 1$. As a possible starting point for the verification of this conjecture we see the multinomial model whose L_2 differentiability is spelled out in Example 3.4.12 of Rieder (1994). Since the multinomial model can be regarded as an exponential family of full rank (cf. Beispiel 1.159 of Witting (1985) or Example 5.3 of Lehmann and Casella (1998), respectively), it is also covered by Subsection 2.3.3. ////

2.1.3 Boundedness of Lagrange Multipliers

The subsequent proposition guarantees the boundedness of the Lagrange multipliers.

Proposition 2.1.7 *Let $D = \mathbb{I}_k$. Then, the Lagrange multipliers A_θ, a_θ, b_θ and c_θ contained in the MSE solutions $\tilde{\eta}_\theta$ for radius $r \in (0,\infty)$ given by Theorem 1.3.11 are bounded where*

$$|a_\theta| \le r^2 b_\theta \tag{2.1.45}$$

In addition,

$$m''_{\theta,i} - b_\theta \le a_{\theta,i} \le m'_{\theta,i} + b_\theta \qquad \text{for all} \quad i = 1,\dots,k \tag{2.1.46}$$

with

$$m'_{\theta,i} = \sup\Big\{ m \in \mathbb{R} \,\Big|\, P_\theta\big((A_\theta\Lambda_\theta)_i \le m\big) \le 1 - \big(2b_\theta^2\,\mathrm{E}_\theta\,\Lambda_{\theta,i}^2\big)^{-1} \Big\} \tag{2.1.47}$$

and

$$m''_{\theta,i} = \inf\Big\{ m \in \mathbb{R} \,\Big|\, P_\theta\big((A_\theta\Lambda_\theta)_i \le m\big) \ge \big(2b_\theta^2\,\mathrm{E}_\theta\,\Lambda_{\theta,i}^2\big)^{-1} \Big\} \tag{2.1.48}$$

In dimension $k = 1$ this can be strengthened to

$$m''_\theta - b_\theta \le a_\theta \le m'_\theta + b_\theta \tag{2.1.49}$$

with

$$m'_\theta = \sup\big\{ m \in \mathbb{R} \,\big|\, P_\theta\big(A_\theta\Lambda_\theta \le m\big) \le \tfrac{1}{2} \big\} \tag{2.1.50}$$

the maximal median of $A_\theta\Lambda_\theta$ and

$$m''_\theta = \inf\big\{ m \in \mathbb{R} \,\big|\, P_\theta\big(A_\theta\Lambda_\theta \le m\big) \ge \tfrac{1}{2} \big\} \tag{2.1.51}$$

the minimal median of $A_\theta\Lambda_\theta$.

Proof We fix $\theta \in \mathbb{R}$ and $r \in (0,\infty)$. By dominated convergence we have,

$$\lim_{M\to\infty} \mathrm{E}\,\Lambda\Lambda^\tau\,\mathrm{I}\big(|\mathcal{I}^{-1}\Lambda| \le M\big) = \mathcal{I} \succ 0 \tag{2.1.52}$$

Thus, the continuity of the determinant implies that there is some (sufficiently large) $M \in (0,\infty)$ such that

$$\mathrm{E}\,\Lambda\Lambda^\tau\,\mathrm{I}\big(|\mathcal{I}^{-1}\Lambda| \le M\big) \succ 0 \tag{2.1.53}$$

That is, we can apply the construction given on p 197 of Rieder (1994) which leads to the following bounded IC,

$$\chi = \big[\mathrm{E}\,\Lambda\Lambda^\tau\mathrm{J}\big]^{-1}\big[\Lambda\mathrm{J} - \mathrm{E}\,\Lambda\mathrm{J}\big] \qquad \mathrm{J} := \mathrm{I}\big(|\mathcal{I}^{-1}\Lambda| \le M\big) \tag{2.1.54}$$

Hence, the maximum asymptotic MSE of this IC,

$$\mathrm{maxMSE}(\chi, r) = \mathrm{E}\,|\chi|^2 + r^2 \sup{}_P |\chi|^2 \tag{2.1.55}$$

is bounded. Consequentially, the corresponding optimally robust IC $\tilde{\eta}$ must have a bounded maximum asymptotic MSE, too. That is, by Proposition 2.1.1, $\operatorname{tr} A$ is bounded. In case of $(* = v)$ we have $A \in \mathbb{R}$ which immediately implies A is bounded. In case of $(* = c)$ we get by the boundedness of $\operatorname{tr} A$ and the positive definiteness and symmetry of A that the operator norm of $A \in \mathbb{R}^{k\times k}$ is,

$$||A||_{\mathrm{op}} = \sup_{|x|\le 1} |Ax| = \sup_{|y|\le 1} |GAG^\tau y| = \max_{j=1,\dots,k} \lambda_j \le \operatorname{tr} A < \infty \tag{2.1.56}$$

where λ_j are the eigenvalues of A and G is an orthogonal matrix such that $GAG^\tau = \operatorname{diag}(\lambda_1, \dots, \lambda_k)$ (spectral decomposition). Thus, A is bounded in operator norm, hence in any other norm of $\mathbb{R}^{k\times k}$. The boundedness of maxMSE also immediately implies that b and $c \in (-b, 0)$ are bounded. Finally, we can derive bounds for a. The first bound (2.1.45) is a consequence of

$$-a = \mathrm{E}\, A\Lambda - a = \mathrm{E}(A\Lambda - a)\left[1 - \min\left\{1, \frac{b}{|A\Lambda - a|}\right\}\right] \tag{2.1.57}$$

$$= \mathrm{E}\, \frac{A\Lambda - a}{|A\Lambda - a|}\big(|A\Lambda - a| - b\big)_+ \tag{2.1.58}$$

and the MSE equation (1.3.45). To obtain the second bound (2.1.46), we use

$$0 = \mathrm{E}(A\Lambda - a) \min\left\{1, \frac{b}{|A\Lambda - a|}\right\} \tag{2.1.59}$$

which is equivalent to

$$0 = \mathrm{E}\big[-\gamma_i b \vee (A\Lambda - a)_i \wedge \gamma_i b\big] \qquad \text{for all} \quad i = 1, \dots, k \tag{2.1.60}$$

with

$$\gamma_i := |(A\Lambda - a)_i| / |A\Lambda - a| \qquad 0 \le \gamma_i \le 1 \tag{2.1.61}$$

Adding $b\,\mathrm{E}\,\gamma_i$ on both sides of (2.1.60), we get

$$b\,\mathrm{E}\,\gamma_i = \mathrm{E}\big[0 \vee (A\Lambda - a)_i + \gamma_i b \wedge 2\gamma_i b\big] \qquad \text{for all} \quad i = 1, \dots, k \tag{2.1.62}$$

Thus,

$$b\,\mathrm{E}\,\gamma_i \le 2b\,P\big((A\Lambda - a)_i + b\gamma_i > 0\big) \le 2b\,P\big((A\Lambda - a)_i + b > 0\big) \tag{2.1.63}$$

$$= 2b\,P\big((A\Lambda)_i > a_i - b\big) \qquad \text{for all} \quad i = 1, \dots, k \tag{2.1.64}$$

which is equivalent to

$$P\big((A\Lambda)_i \le a_i - b\big) \le 1 - \tfrac{1}{2}\,\mathrm{E}\,\gamma_i \qquad \text{for all} \quad i = 1, \dots, k \tag{2.1.65}$$

Furthermore, using Fisher consistency and Cauchy-Schwarz we obtain,

$$1^2 = \Big\{\mathrm{E}\,\Lambda_i\big[-b\gamma_i \vee (A\Lambda - a)_i \wedge b\gamma_i\big]\Big\}^2 \le \mathrm{E}\,\Lambda_i^2\,\mathrm{E}\,\big[(A\Lambda - a)_i^2 \wedge b^2\gamma_i^2\big] \tag{2.1.66}$$

$$\le b^2\,\mathrm{E}\,\Lambda_i^2\,\mathrm{E}\,\gamma_i^2 \qquad \text{for all} \quad i = 1, \dots, k \tag{2.1.67}$$

which implies

$$\mathrm{E}\,\gamma_i^2 \geq \left(b^2\,\mathrm{E}\,\Lambda_i^2\right)^{-1} > 0 \qquad \text{for all} \quad i = 1, \ldots, k \tag{2.1.68}$$

Since $\gamma_i \in [0,1]$, we additionally have $\mathrm{E}\,\gamma_i^2 \leq \mathrm{E}\,\gamma_i$, respectively

$$\tfrac{1}{2}\,\mathrm{E}\,\gamma_i \geq \tfrac{1}{2}\,\mathrm{E}\,\gamma_i^2 \geq \left(2b^2\,\mathrm{E}\,\Lambda_i^2\right)^{-1} > 0 \tag{2.1.69}$$

Therefore, we get with m_i' the maximal $1 - \left(2b^2\,\mathrm{E}\,\Lambda_i^2\right)^{-1}$ quantile of $(A\Lambda)_i$ as defined in (2.1.47),

$$a_i - b \leq m_i' \qquad \Longrightarrow \qquad a_i \leq m_i' + b \tag{2.1.70}$$

for all $i = 1, \ldots, k$. Analogously, we get

$$a_i \geq m_i'' - b \qquad \text{for all} \quad i = 1, \ldots, k \tag{2.1.71}$$

by subtracting $b\,\mathrm{E}\,\gamma_i$ on both sides of (2.1.60) where m_i'' is the minimal $\left(2b^2\,\mathrm{E}\,\Lambda_i^2\right)^{-1}$ quantile as defined in (2.1.48). In dimension $k = 1$, equation (2.1.60) reads

$$0 = \mathrm{E}\left[\,-b_\theta \vee A_\theta\Lambda_\theta - a_\theta \wedge b_\theta\right] \tag{2.1.72}$$

Adding b_θ, respectively $-b_\theta$ on both sides, we obtain

$$b_\theta = \mathrm{E}\left[0 \vee A_\theta\Lambda_\theta - a_\theta + b_\theta \wedge 2b_\theta\right] \tag{2.1.73}$$

respectively

$$-b_\theta = \mathrm{E}\left[\,-2b_\theta \vee A_\theta\Lambda_\theta - a_\theta - b_\theta \wedge 0\right] \tag{2.1.74}$$

Hence,

$$b_\theta \leq 2b_\theta P(A_\theta\Lambda_\theta > a_\theta - b_\theta) \tag{2.1.75}$$

respectively

$$-b_\theta \geq -2b_\theta P(A_\theta\Lambda_\theta < a_\theta + b_\theta) \tag{2.1.76}$$

which leads us to (2.1.49). ////

Remark 2.1.8 **(a)** The proof of the boundedness of a_θ is based on arguments provided by P. Ruckdeschel. Furthermore, he was able to generalize the previous proposition to arbitrary $D \in \mathbb{R}^{p\times k}$ with $\mathrm{rk}\,D = p$; confer Ruckdeschel (2005b).

(b) Another bound for a_θ may be read off from Lemma 2 of Shen (1995). He proves,

$$|a_\theta| \leq 3\,\mathrm{E}_\theta\,|A_\theta\Lambda_\theta| + 2b_\theta \tag{2.1.77}$$

Shen (1995) derives the corresponding solution by differentiation (not by Lagrange arguments) where he interchanges differentiation and integration. Therefore, he in our view needs the additional assumption $P_\theta(A_\theta\Lambda_\theta = a_\theta) = 0$.

(c) Obviously, in case $r < 1$ bound (2.1.45) is stricter than bound (2.1.46) and (2.1.49), respectively. Since $\mathrm{E}\,|A\Lambda| > 0$, bound (2.1.45) is stricter than bound (2.1.77) if $r \leq \sqrt{2}$. Moreover, if $k = 1$, the Markov inequality yields

$$\tfrac{1}{2} \leq P_\theta\big(\Lambda_\theta \geq \mathrm{med}(\Lambda_\theta)\big) \leq \frac{\mathrm{E}_\theta(\Lambda_\theta)_+}{\mathrm{med}(\Lambda_\theta)} \qquad \text{if } \mathrm{med}(\Lambda_\theta) > 0 \tag{2.1.78}$$

$$\tfrac{1}{2} \leq P_\theta\big(-\Lambda_\theta \geq -\mathrm{med}(\Lambda_\theta)\big) \leq \frac{\mathrm{E}_\theta(-\Lambda_\theta)_+}{-\mathrm{med}(\Lambda_\theta)} \qquad \text{if } \mathrm{med}(\Lambda_\theta) < 0 \tag{2.1.79}$$

As $\mathrm{E}_\theta(\Lambda_\theta)_+ = \mathrm{E}_\theta(\Lambda_\theta)_-$ and $\mathrm{E}_\theta(-\Lambda_\theta)_+ = \mathrm{E}_\theta(\Lambda_\theta)_-$, we get

$$|\,\mathrm{med}(\Lambda_\theta)| \leq \mathrm{E}_\theta\,|\Lambda_\theta| \tag{2.1.80}$$

Thus, in dimension $k = 1$ bound (2.1.49) is clearly stricter than bound (2.1.77) which in addition includes the term $2b_\theta$ instead of b_θ.

(d) The bounds (2.1.46), (2.1.49) and (2.1.77) are also valid in case of the Hampel type problem (1.3.7) and $* = c$. Whereas, bound (2.1.46) was derived using the MSE equation (1.3.45). Hence, it is only available if we consider the corresponding MSE problem (1.3.5) and $* = c$. ////

2.1.4 Uniqueness of Lagrange Multipliers

As noted in Remark B.2.10 (a) of Rieder (1994) the Lagrange multipliers (like the separating hyperplanes) need not be unique.
However, the clipping bounds $b \in (0, \infty)$ and $c \in (-b, 0)$ in the solutions $\tilde{\eta}_\theta$ given by Theorems 1.3.7–1.3.11 are uniquely determined since these bounds have been derived in terms of the solution $\tilde{\eta}_\theta$ as

$$b = \omega_*(\tilde{\eta}_\theta) \qquad c = \inf{}_{P_\theta}\tilde{\eta}_\theta \qquad * = c, v \tag{2.1.81}$$

confer Remark 5.5.8 of Rieder (1994). By Proposition 2.1.1 $\mathrm{tr}\, AD^\tau$ is unique, too. In addition, if $k = 1$ and $\mathrm{med}(\Lambda_\theta)$ is unique, then a_θ is unique by Lemma C.2.4 of Rieder (1994). Moreover, if $k = 1$, $\mathrm{med}(\Lambda_\theta)$ is non-unique and $r < \bar{r}$, then a is unique by Proposition 2.1.3.
In case $* = c$, $p > 1$ and

$$\mathrm{support}\,\Lambda_\theta(P_\theta) = \mathbb{R}^k \tag{2.1.82}$$

$A \in \mathbb{R}^{p\times k}$ and $a = Az$ with $z \in \mathbb{R}^k$ are unique; confer Remark 5.5.8 of Rieder (1994). In case $* = c$ and $p > 1$ the uniqueness of the Lagrange multipliers is also implied by

$$P_\theta(\Lambda_\theta \in H) < P_\theta(|\tilde{\eta}_\theta| < b) \tag{2.1.83}$$

for all $(k-1)$-dimensional hyperplanes H of $\mathbb{R}^k$; confer Rieder (1994), proof of Theorem 5.5.9. This condition is certainly satisfied if

$$e \in \mathbb{R}^k,\ \alpha \in \mathbb{R},\ P_\theta(e^\tau\Lambda_\theta = \alpha) > 0 \qquad \Longrightarrow \qquad e = 0 \tag{2.1.84}$$

that is, $P_\theta(\Lambda \in H) = 0$ for all $(k-1)$-dimensional hyperplanes H of $\mathbb{R}^k$. In turn, conditions (2.1.82) and (2.1.84) entail $\mathcal{I}_\theta \succ 0$.

2.1.5 Continuity Properties of Lagrange Multipliers

The Lagrange multipliers and hence the asymptotic MSE, the trace of the asymptotic variance and the standardized asymptotic bias of the MSE solution $\tilde{\eta}_r = \tilde{\eta}$ are continuous in $r \in (0, \infty)$.

Proposition 2.1.9 *Let $\tilde{\eta}_r$ be the MSE solution for $D = \mathbb{I}_k$ and radius $r \in (0, \infty)$ derived in Theorem 1.3.11 and $(r_n)_{n\in\mathbb{N}} \subset (0, \infty)$ with $r_n \to r$ as $n \to \infty$. Then,*

$$\lim_{n\to\infty} \operatorname{tr} A_{r_n} = \operatorname{tr} A_r \qquad \lim_{n\to\infty} b_{r_n} = b_r \qquad \lim_{n\to\infty} c_{r_n} = c_r \tag{2.1.85}$$

In case A_r and a_r are unique, then also

$$\lim_{n\to\infty} A_{r_n} = A_r \qquad \text{and} \qquad \lim_{n\to\infty} a_{r_n} = a_r \tag{2.1.86}$$

Proof To simplify the notation, we drop r as an index. Let $\tilde{\eta}$ be the MSE solution for $D = \mathbb{I}_k$ and radius $r \in (0, \infty)$ provided by Theorem 1.3.11 and $(r_n)_{n\in\mathbb{N}} \subset (0, \infty)$ with $r_n \to r$ as $n \to \infty$. By the construction given at the beginning of the proof of Proposition 2.1.7, we obtain

$$\operatorname{maxMSE}(\tilde{\eta}_n, r_n) = \mathrm{E}\,|\tilde{\eta}_n|^2 + r_n^2 \omega_*(\tilde{\eta}_n)^2 \tag{2.1.87}$$

is uniformly bounded in $n \in \mathbb{N}$. That is, by Proposition 2.1.1, $\operatorname{tr} A_n$ is uniformly bounded in $n \in \mathbb{N}$. In case $(* = v)$ we have $A_n \in \mathbb{R}$ which immediately implies A_n is uniformly bounded. In case $(* = c)$ we get by the boundedness of $\operatorname{tr} A_n$ and the positive definiteness and symmetry of A_n that the operator norm of A_n is,

$$||A_n||_{\mathrm{op}} = \sup_{|x|\le 1} |A_n x| = \sup_{|y|\le 1} |G_n A_n G_n^\tau y| = \max_{j=1,\ldots,k} \lambda_{n,j} \le \operatorname{tr} A_n < \infty \tag{2.1.88}$$

where $\lambda_{n,j}$ are the eigenvalues of A_n and G_n are orthogonal matrices such that $G_n A_n G_n^\tau = \operatorname{diag}(\lambda_{n,1}, \ldots, \lambda_{n,k})$ (spectral decomposition). Thus, A_n is bounded uniformly in n. In addition, the uniform boundedness of maxMSE entails b_n and $c_n \in (-b_n, 0)$ are bounded uniformly in n. Finally, a_n is uniformly bounded in n by bound (2.1.45). Consequentially, there is a subsequence $(m) \subset (n)$ such that $A_m \to A'$, $b_m \to b'$, $c_m \to c'$ and $a_m \to a'$. We now define

$$\chi = \chi(\Lambda) = (A'\Lambda - a') \min\left\{1, \frac{b'}{|A'\Lambda - a'|}\right\} \qquad (* = c) \tag{2.1.89}$$

respectively

$$\chi = \chi(\Lambda) = c' \vee A'\Lambda \wedge (c' + b') \qquad (* = v) \tag{2.1.90}$$

Since $\tilde{\eta}_m \to \chi$ and $|\tilde{\eta}_m|$ is uniformly bounded in m, we obtain by dominated convergence

$$0 = \lim_{m\to\infty} \mathrm{E}\,\tilde{\eta}_m = \mathrm{E}\,\chi \tag{2.1.91}$$

Furthermore,

$$\big(|A_m\Lambda - a_m| - b_m\big)_+ \le |A_m\Lambda - a_m| \le ||A_m||_{\mathrm{op}}|\Lambda| + |a_m| \tag{2.1.92}$$

where

$$\begin{aligned} \mathrm{E}\left[||A_m||_{\mathrm{op}}|\Lambda| + |a_m|\right] &\le ||A_m||_{\mathrm{op}}\left[\mathrm{E}\,|\Lambda|^2\right]^{1/2} + |a_m| \\ &= ||A_m||_{\mathrm{op}}\sqrt{\operatorname{tr}\mathcal{I}} + |a_m| \end{aligned} \tag{2.1.93}$$

is uniformly bounded in m, respectively

$$\big(c_m - A_m\Lambda\big)_+ \le A_m|\Lambda| + c_m \tag{2.1.94}$$

where

$$\mathrm{E}\,\big[A_m|\Lambda| + b_m\big] \le A_m\big[\,\mathrm{E}_m\,|\Lambda|^2\big]^{1/2} + c_m = A_m\sqrt{\mathrm{tr}\,\mathcal{I}} + c_m \tag{2.1.95}$$

is uniformly bounded in m. Hence, an application of dominated convergence yields

$$0 = \lim_{m\to\infty} \mathrm{E}\,\big(|A_m\Lambda - a_m| - b_m\big)_+ - r_m^2 b_m = \mathrm{E}\,\big(|A'\Lambda - a'| - b'\big)_+ - r^2 b' \tag{2.1.96}$$

respectively

$$0 = \lim_{m\to\infty} \mathrm{E}\,\big(c_m - A_m\Lambda\big)_+ - r_m^2 b_m = \mathrm{E}\,\big(c' - A'\Lambda\big)_+ - r^2 b' \tag{2.1.97}$$

Moreover,

$$|\tilde\eta_m\Lambda^\tau| \le |(A_m\Lambda - a_m)\Lambda^\tau| \le \|A_m\|_{\mathrm{op}}|\Lambda|^2 + |a_m|\,|\Lambda| \tag{2.1.98}$$

where

$$\mathrm{E}\,\big[\|A_m\|_{\mathrm{op}}|\Lambda|^2 + |a_m|\,|\Lambda|\big] = \|A_m\|_{\mathrm{op}}\,\mathrm{tr}\,\mathcal{I} + |a_m|\sqrt{\mathrm{tr}\,\mathcal{I}} \tag{2.1.99}$$

is uniformly bounded in m. That is, we can once again apply dominated convergence and get

$$\mathbb{I}_k = \lim_{m\to\infty} \mathrm{E}\,\tilde\eta_m\Lambda^\tau = \mathrm{E}\,\chi\Lambda^\tau \tag{2.1.100}$$

Thus, χ is the unique solution to the MSE problem (1.3.5) for radius r; i.e., $\chi = \eta$ in $L_2(P)$. In particular, we obtain $b' = b$, $c' = c$ and $\mathrm{tr}\,A' = \mathrm{tr}\,A$ by the uniqueness of b, c and $\mathrm{tr}\,A$. Hence, b, c and $\mathrm{tr}\,A$ are the unique accumulation points of the sequences b_n, c_n and $\mathrm{tr}\,A_n$ and the sequences converge. If, in addition A and a are unique, then also $A' = A$ and $a' = a$ where the accumulation points are unique; i.e., (2.1.86) holds. ////

Remark 2.1.10 **(a)** The proof of the previous proposition is similar to the proof of the subsequent Theorem 2.1.11, but easier. We can argue with dominated convergence, since $\mathcal{L}(\Lambda)$ is independent of n, whereas in Theorem 2.1.11 we have to invoke uniform integrability.

(b) As a direct consequence of Proposition 2.1.9, we obtain

$$\mathrm{E}\,|\tilde\eta_{r_n}|^2 = \mathrm{tr}\,A_{r_n} - r_n^2 b_{r_n}^2 \longrightarrow \mathrm{tr}\,A_r - r^2 b_r^2 = \mathrm{E}\,|\tilde\eta_r|^2 \tag{2.1.101}$$

as $n\to\infty$; i.e., the trace of the asymptotic covariance is continuous in $r\in(0,\infty)$.

(c) Rieder (1994) proves the uniqueness and continuity of a_r on $(0,\infty)$ in case $k = 1$ and $\mathrm{med}(\Lambda_\theta)$ unique; confer Lemma C.2.4 (ibid.).

(d) A generalization of Proposition 2.1.9 to arbitrary $D\in\mathbb{R}^{p\times k}$ with $\mathrm{rk}\,D = p$ is given in Ruckdeschel (2005b). ////

Under some additional assumptions the Lagrange multipliers and hence the asymptotic MSE, the trace of the asymptotic variance and the standardized asymptotic bias of the MSE solution $\tilde\eta_\theta = \tilde\eta$ are continuous in the parameter $\theta\in\Theta$. Moreover, the minimum bias $\omega_*^{\min}$ $(* = c, v)$ is continuous in $\theta\in\Theta$.

Theorem 2.1.11 *Let $D = \mathbb{I}_k$ and A_θ, a_θ, b_θ and c_θ be the Lagrange multipliers contained in the solution $\tilde{\eta}_\theta$ to the MSE problem (1.3.5) for radius $r \in (0,\infty)$ given in Theorem 1.3.11. Further assume*

$$\operatorname{tr} \mathcal{I}_{\theta_n} \longrightarrow \operatorname{tr} \mathcal{I}_\theta \qquad \text{as } n \to \infty \tag{2.1.102}$$

and

$$\mathcal{L}_{P_{\theta_n}}(\Lambda_{\theta_n}) \xrightarrow{\text{w}} \mathcal{L}_{P_\theta}(\Lambda_\theta) \qquad \text{as } n \to \infty \tag{2.1.103}$$

where $(\theta_n)_{n\in\mathbb{N}} \subset \Theta$ is some sequence such that $\theta_n \to \theta$ as $n \to \infty$.

(a) *Then,*

$$\lim_{n\to\infty} \operatorname{tr} A_{\theta_n} = \operatorname{tr} A_\theta \qquad \lim_{n\to\infty} b_{\theta_n} = b_\theta \qquad \lim_{n\to\infty} c_{\theta_n} = c_\theta \tag{2.1.104}$$

In case A_θ and a_θ are unique, then also

$$\lim_{n\to\infty} A_{\theta_n} = A_\theta \qquad \text{and} \qquad \lim_{n\to\infty} a_{\theta_n} = a_\theta \tag{2.1.105}$$

(b) *It holds for the minimum bias $\omega^{\min}_{*,\theta}$ given in Theorems 1.3.7 (b) and 1.3.9 (b), respectively*

$$\lim_{n\to\infty} \omega^{\min}_{*,\theta_n} = \omega^{\min}_{*,\theta} \qquad (* = c, v) \tag{2.1.106}$$

Proof To simplify the notation, we omit θ as an index and identify θ_n by n.

(a) Let $r \in (0,\infty)$ be fixed. We have

$$\mathrm{E}\,|\Lambda_{n,i}\Lambda_{n,j}| \le \big[\mathrm{E}_n\,|\Lambda_{n,i}|^2\big]^{1/2}\big[\mathrm{E}_n\,|\Lambda_{n,j}|^2\big]^{1/2} \le \mathrm{E}_n\,|\Lambda_n|^2 = \operatorname{tr}\mathcal{I}_n \to \operatorname{tr}\mathcal{I} < \infty \tag{2.1.107}$$

for all $i,j = 1,\dots,k$. Moreover, if we fix some $\varepsilon > 0$, there exists some $\delta(\varepsilon) > 0$ such that for any $A_n \in \mathbb{B}^k$: $P_n(A_n) < \delta(\varepsilon)$ implies

$$\int\limits_{A_n} |\Lambda_{n,i}\Lambda_{n,j}|\,dP_n \le \Big[\int\limits_{A_n} |\Lambda_{n,i}|^2\,dP_n\Big]^{1/2}\Big[\int\limits_{A_n} |\Lambda_{n,j}|^2\,dP_n\Big]^{1/2} \le \int\limits_{A_n} |\Lambda_n|^2\,dP_n \le \varepsilon \tag{2.1.108}$$

for all $i,j = 1,\dots,k$. That is, $(\Lambda_n\Lambda_n^\tau)$ is uniformly integrable by Theorem 4.5.3 of Chung (2000). Moreover, $\mathcal{L}_n(\Lambda_n\Lambda_n^\tau) \xrightarrow{\text{w}} \mathcal{L}(\Lambda\Lambda^\tau)$ by the continuous mapping theorem and we therefore can apply Vitali's theorem (cf. Corollary A.2.3 of Rieder (1994)) which yields $\mathcal{I}_n \to \mathcal{I}$, where $\mathcal{I} \succ 0$. Thus, by the continuity of the determinant, there is some $N \in \mathbb{N}$ such that for all $n \ge N$, $\mathcal{I}_n \succ 0$. In the sequel, we argue similarly to the proof of Proposition 2.1.7; i.e., we want to apply the construction given on p 197 of Rieder (1994). Therefore, we first verify that there is some $M \in (0,\infty)$ such that

$$\mathrm{E}_n\,\Lambda_n\Lambda_n^\tau\,\mathrm{I}\big(|\mathcal{I}_n^{-1}\Lambda_n| \le M\big) \tag{2.1.109}$$

is regular for sufficiently large $n \in \mathbb{N}$. Since $\mathcal{L}_n(\Lambda_n) \xrightarrow{\text{w}} \mathcal{L}(\Lambda)$, we get by the continuous mapping theorem for all $M \in (0,\infty)$ satisfying $P(|\mathcal{I}^{-1}\Lambda| = M) = 0$ that

$$\mathrm{E}_n\,\Lambda_n\Lambda_n^\tau\,\mathrm{I}\big(|\mathcal{I}_n^{-1}\Lambda_n| \le M\big) \to \mathrm{E}\,\Lambda\Lambda^\tau\,\mathrm{I}\big(|\mathcal{I}^{-1}\Lambda| \le M\big) \tag{2.1.110}$$

Therefore, by $\mathcal{I}_n \to \mathcal{I}$ we also have

$$\mathrm{E}_n\, \Lambda_n \Lambda_n^\tau\, \mathrm{I}\left(|\mathcal{I}_n^{-1}\Lambda_n| > M\right) \to \mathrm{E}\, \Lambda\Lambda^\tau\, \mathrm{I}\left(|\mathcal{I}^{-1}\Lambda| > M\right) \tag{2.1.111}$$

We can now choose $M \in (0,\infty)$ subject to $P(|\mathcal{I}^{-1}\Lambda| = M) = 0$ so large that the right hand side of (2.1.111) becomes arbitrarily small (e.g., in operator norm). Thus, by the continuity of the determinant there is some (sufficiently large) M such that

$$\mathrm{E}_n\, \Lambda_n \Lambda_n^\tau\, \mathrm{I}\left(|\mathcal{I}_n^{-1}\Lambda_n| \le M\right) \succ 0 \tag{2.1.112}$$

for all $n \ge N_1$ $(\ge N)$. That is, for $n \ge N_1$ we can define the following ICs,

$$\chi_n = \left[\mathrm{E}_n\, \Lambda_n \Lambda_n^\tau \mathrm{J}_n\right]^{-1} \left[\Lambda_n \mathrm{J}_n - \mathrm{E}_n\, \Lambda_n \mathrm{J}_n\right] \qquad \mathrm{J}_n := \mathrm{I}\left(|\mathcal{I}_n^{-1}\Lambda_n| \le M\right) \tag{2.1.113}$$

which are bounded uniformly in n. Hence, the maximum asymptotic MSE

$$\mathrm{maxMSE}(\chi_n, r) = \mathrm{E}_n\, |\chi_n|^2 + r^2 \sup\nolimits_{P_n} |\chi_n|^2 \tag{2.1.114}$$

is bounded uniformly in n. Thus, for $n \ge N_1$ the corresponding optimally robust ICs $\tilde{\eta}_n$ must have a uniformly bounded maximum asymptotic MSE; i.e., by Proposition 2.1.1, $\mathrm{tr}\, A_n$ is bounded for $n \ge N_1$. In case $(* = v)$ we have $A_n \in \mathbb{R}$ which immediately implies A_n is bounded. In case $(* = c)$ we get by the boundedness of $\mathrm{tr}\, A_n$ and the positive definiteness and symmetry of A_n that the operator norm of A_n is,

$$||A_n||_{\mathrm{op}} = \sup_{|x|\le 1} |A_n x| = \sup_{|y|\le 1} |G_n A_n G_n^\tau y| = \max_{j=1,\ldots,k} \lambda_{n,j} \le \mathrm{tr}\, A_n < \infty \tag{2.1.115}$$

where $\lambda_{n,j}$ are the eigenvalues of A_n and G_n are orthogonal matrices such that $G_n A_n G_n^\tau = \mathrm{diag}\,(\lambda_{n,1}, \ldots, \lambda_{n,k})$ (spectral decomposition). Thus, A_n is bounded uniformly in n. By the uniform boundedness of maxMSE, this immediately implies that b_n and $c_n \in (-b_n, 0)$ are bounded uniformly in n. Finally, a_n is uniformly bounded in n by bound (2.1.45). Consequentially, there is a subsequence $(m) \subset (n)$ such that $A_m \to A'$, $b_m \to b'$, $c_m \to c'$ and $a_m \to a'$. We now define

$$\chi = \chi(\Lambda) = (A'\Lambda - a') \min\left\{1, \frac{b'}{|A'\Lambda - a'|}\right\} \qquad (* = c) \tag{2.1.116}$$

respectively

$$\chi = \chi(\Lambda) = c' \vee A'\Lambda \wedge (c' + b') \qquad (* = v) \tag{2.1.117}$$

By assumption (2.1.103) and as $\tilde{\eta}_m(u_m) \to \chi(u)$ for $u_m \to u$, Theorem 5.5 of Billingsley (1968) yields $\mathcal{L}_m(\tilde{\eta}_m) \xrightarrow{\mathrm{w}} \mathcal{L}(\chi)$ and therefore we obtain by the uniform boundedness of $\tilde{\eta}_m$ and χ

$$0 = \lim_{m\to\infty} \mathrm{E}_m\, \tilde{\eta}_m = \mathrm{E}\, \chi \tag{2.1.118}$$

Moreover,

$$\begin{aligned}
\mathrm{E}_m \left(|A_m\Lambda_m - a_m| - b_m\right)_+ &\le \mathrm{E}_m\, |A_m\Lambda_m - a_m| \le ||A_m||_{\mathrm{op}}\, \mathrm{E}_m\, |\Lambda_m| + |a_m| \\
&\le ||A_m||_{\mathrm{op}} \left[\mathrm{E}_m\, |\Lambda_m|^2\right]^{1/2} + |a_m| \\
&= ||A_m||_{\mathrm{op}} \sqrt{\mathrm{tr}\, \mathcal{I}_m} + |a_m| \\
&\to ||A'||_{\mathrm{op}} \sqrt{\mathrm{tr}\, \mathcal{I}} + |a'| < \infty
\end{aligned} \tag{2.1.119}$$

respectively

$$\begin{aligned} \mathrm{E}_m\,(c_m - A_m\Lambda_m)_+ &\le A_m\,\mathrm{E}_m\,|\Lambda_m| + c_m \le A_m\big[\mathrm{E}_m\,|\Lambda_m|^2\big]^{1/2} + c_m \\ &= A_m\sqrt{\operatorname{tr}\mathcal{I}_m} + c_m \to A'\sqrt{\operatorname{tr}\mathcal{I}} + c' < \infty \end{aligned} \tag{2.1.120}$$

i.e., $\big((|A_m\Lambda_m - a_m| - b_m)_+\big)$, respectively $\big((c_m - A_m\Lambda_m)_+\big)$ is uniformly integrable. In addition, also $(\tilde{\eta}_m\Lambda_m^\tau)$ is uniformly integrable by Theorem 4.5.3 of Chung (2000) which may be shown analogously to the uniform integrability of $(\Lambda_n\Lambda_n^\tau)$ at the beginning of this proof. Furthermore, $\mathcal{L}_m(\tilde{\eta}_m\Lambda_m^\tau) \xrightarrow{\mathrm{w}} \mathcal{L}(\chi\Lambda^\tau)$ and

$$\mathcal{L}_m\left((|A_m\Lambda_m - a_m| - b_m)_+\right) \xrightarrow{\mathrm{w}} \mathcal{L}\left((|A'\Lambda - a'| - b')_+\right) \tag{2.1.121}$$

respectively

$$\mathcal{L}_m\left((c_m - A_m\Lambda_m)_+\right) \xrightarrow{\mathrm{w}} \mathcal{L}\left((c' - A'\Lambda)_+\right) \tag{2.1.122}$$

by a combination of the Cramér-Wold device, Slutzky's lemma and the continuous mapping theorem. That is, we may apply Vitali's theorem (cf. Corollary A.2.3 of Rieder (1994)) and obtain

$$\mathbb{I}_k = \lim_{m\to\infty} \mathrm{E}_m\,\tilde{\eta}_m\Lambda_m^\tau = \mathrm{E}\,\chi\Lambda^\tau \tag{2.1.123}$$

$$0 = \lim_{m\to\infty} \mathrm{E}_m\left(|A_m\Lambda_m - a_m| - b_m\right)_+ - r^2 b_m = \mathrm{E}\left(|A'\Lambda - a'| - b'\right)_+ - r^2 b' \tag{2.1.124}$$

respectively

$$0 = \lim_{m\to\infty} \mathrm{E}_m\left(c_m - A_m\Lambda_m\right)_+ - r^2 b_m = \mathrm{E}\left(c' - A'\Lambda\right)_+ - r^2 b' \tag{2.1.125}$$

Thus, χ is the unique solution to the MSE problem (1.3.5); i.e., $\chi = \eta$ in $L_2(P)$. In particular, we obtain $b' = b$, $c' = c$ and $\operatorname{tr} A' = \operatorname{tr} A$ by the uniqueness of b, c and $\operatorname{tr} A$. Hence, b, c and $\operatorname{tr} A$ are the unique accumulation points of the sequences b_n, c_n and $\operatorname{tr} A_n$ and the sequences converge. If, in addition A and a are unique, then also $A' = A$ and $a' = a$ where the accumulation points are unique; i.e., (2.1.105) holds.

(b) $*$ = c: Without restriction we can rewrite the minimum bias given in Theorem 1.3.7 (b) as

$$\omega_c^{\min} = \max\left\{\frac{\operatorname{tr} A}{\mathrm{E}\,|A\Lambda - a|}\;\middle|\; a \in \mathbb{R}^k,\ A \in \mathbb{R}^{k\times k}\setminus\{0\},\ \|A\|_{\mathrm{op}} = 1\right\} \tag{2.1.126}$$

Let $a \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k\times k}$ with $\|A\|_{\mathrm{op}} = 1$ be fixed such that

$$\frac{\operatorname{tr} A}{\mathrm{E}\,|A\Lambda - a|} = \omega_c^{\min} \tag{2.1.127}$$

Then,

$$\omega_{c,n}^{\min} \ge \frac{\operatorname{tr} A}{\mathrm{E}_n\,|A\Lambda_n - a|} \tag{2.1.128}$$

Since

$$\limsup_{n\to\infty} \mathrm{E}_n\,|A\Lambda_n - a| \le \|A\|_{\mathrm{op}} \limsup_{n\to\infty} \mathrm{E}_n\,|\Lambda_n| + |a| \tag{2.1.129}$$

$$\le \limsup_{n\to\infty} \sqrt{\operatorname{tr} \mathcal{I}_n} + |a| = \sqrt{\operatorname{tr} \mathcal{I}} + |a| < \infty \tag{2.1.130}$$

and $\mathcal{L}_n(\Lambda_n) \xrightarrow{\mathrm{w}} \mathcal{L}(\Lambda)$ by assumption (2.1.103), we can apply Vitali's theorem and obtain

$$\lim_{n\to\infty} \mathrm{E}_n\,|A\Lambda_n - a| = \mathrm{E}\,|A\Lambda - a| \tag{2.1.131}$$

Consequentially,

$$\liminf_{n\to\infty} \omega_{c,n}^{\min} \ge \lim_{n\to\infty} \frac{\operatorname{tr} A}{\mathrm{E}_n\,|A\Lambda_n - a|} = \frac{\operatorname{tr} A}{\mathrm{E}\,|A\Lambda - a|} = \omega_c^{\min} \tag{2.1.132}$$

Now, let $a_n \in \mathbb{R}^k$ and $A_n \in \mathbb{R}^{k\times k}$ with $\|A_n\|_{\mathrm{op}} = 1$ such that

$$\frac{\operatorname{tr} A_n}{\mathrm{E}_n\,|A_n\Lambda_n - a_n|} = \omega_{c,n}^{\min} \tag{2.1.133}$$

for $n \in \mathbb{N}$. We have,

$$\mathrm{E}_n\,|A_n\Lambda_n| \le \|A_n\|_{\mathrm{op}}\,\mathrm{E}\,|\Lambda_n| \le \sqrt{\operatorname{tr} \mathcal{I}_n} < \infty \tag{2.1.134}$$

Moreover,

$$\mathrm{E}_n\,|A_n\Lambda_n - a_n| \le \mathrm{E}_n\,|A_n\Lambda_n| \tag{2.1.135}$$

as a_n is a solution to

$$\frac{\operatorname{tr} A_n}{\mathrm{E}_n\,|A_n\Lambda_n - a_n|} = \max! \tag{2.1.136}$$

Thus, using the triangular inequality,

$$|a_n| \le |A_n\Lambda_n - a_n| + |A_n\Lambda_n| \tag{2.1.137}$$

and taking expectations, we obtain

$$|a_n| \le \mathrm{E}_n\,|A_n\Lambda_n - a_n| + \mathrm{E}_n\,|A_n\Lambda_n| \le 2\,\mathrm{E}_n\,|A_n\Lambda_n| \le 2\sqrt{\operatorname{tr} \mathcal{I}_n} < \infty \tag{2.1.138}$$

That is, a_n and A_n are bounded uniformly in n. Hence, there is a subsequence $(m) \subset (n)$ such that $a_m \to a' \in \mathbb{R}^k$ and $A_m \to A' \in \mathbb{R}^{k\times k}$. Since

$$\limsup_{m\to\infty} \mathrm{E}_m\,|A_m\Lambda_n - a_m| \le \|A_m\|_{\mathrm{op}} \limsup_{m\to\infty} \left[\mathrm{E}_m\,|\Lambda_m| + |a_m|\right] \tag{2.1.139}$$

$$\le \limsup_{m\to\infty} \left[\sqrt{\operatorname{tr} \mathcal{I}_m} + |a_m|\right] \tag{2.1.140}$$

$$= \sqrt{\operatorname{tr} \mathcal{I}} + |a'| < \infty \tag{2.1.141}$$

we obtain by Vitali's theorem

$$\lim_{m\to\infty} \frac{\operatorname{tr} A_m}{\mathrm{E}_m\,|A_m\Lambda_m - a_m|} = \frac{\operatorname{tr} A'}{\mathrm{E}\,|A'\Lambda - a'|} \le \omega_c^{\min} \tag{2.1.142}$$

Consequentially,

$$\limsup_{n\to\infty} \omega_{c,n}^{\min} = \limsup_{n\to\infty} \frac{\operatorname{tr} A_n}{\mathrm{E}_n\,|A_n\Lambda_n - a_n|} \le \omega_c^{\min} \tag{2.1.143}$$

$\boldsymbol{* = v,\ k = 1:}$ The minimum bias given in Theorem 1.3.9 (b) reads

$$\omega_{v,n}^{\min} = \big(\mathrm{E}_n(\Lambda_n)_+\big)^{-1} \tag{2.1.144}$$

where

$$\limsup_{n\to\infty} \mathrm{E}_n(\Lambda_n)_+ \le \limsup_{n\to\infty} \mathrm{E}_n\,|\Lambda_n| \le \limsup_{n\to\infty} \sqrt{\mathrm{E}_n(\Lambda_n)^2} \tag{2.1.145}$$

$$= \sqrt{\mathcal{I}} < \infty \tag{2.1.146}$$

Since $\mathcal{L}_n(\Lambda_n) \xrightarrow{\text{w}} \mathcal{L}(\Lambda)$ by assumption (2.1.103), an application of Vitali's theorem yields

$$\lim_{n\to\infty} \mathrm{E}_n(\Lambda_n)_+ = \mathrm{E}(\Lambda)_+ \tag{2.1.147}$$

i.e., (2.1.106) holds. ////

Remark 2.1.12 (a) As shown at the beginning of the proof of Theorem 2.1.11, conditions (2.1.102) and (2.1.103) imply the convergence of the corresponding Fisher information

$$\mathcal{I}_{\theta_n} = \mathrm{E}_{\theta_n}\,\Lambda_{\theta_n}\Lambda_{\theta_n}^\tau \to \mathrm{E}_\theta\,\Lambda_\theta\Lambda_\theta^\tau = \mathcal{I}_\theta \tag{2.1.148}$$

(b) In view of the approximation used in the proof of Lemma 6.4.4 of Rieder (1994) it might even be possible to abandon assumption (2.1.102).

(c) By the convergences stated in the previous theorem we at once get

$$\mathrm{E}\,|\tilde\eta_{\theta_n}|^2 = \operatorname{tr} A_{\theta_n} - r_n^2 b_{\theta_n}^2 \longrightarrow \operatorname{tr} A_\theta - r^2 b_\theta^2 = \mathrm{E}\,|\tilde\eta_\theta|^2 \tag{2.1.149}$$

as $n \to \infty$. That is, the trace of the asymptotic covariance is continuous in $\theta \in \Theta$.

(d) In view of Ruckdeschel (2005b) a generalization of Proposition 2.1.11 to arbitrary $D \in \mathbb{R}^{p\times k}$ with $\operatorname{rk} D = p$ seems to be in reach. ////

2.2 Least Favorable Radius

Given a neighborhood radius, we can now determine the optimally robust ICs via the implicit equations stated in Subsections 1.3.3 and 1.3.4. But, in most applications the neighborhood radius is unknown or unknown except to belong to some radius interval, respectively. That is, the radius appears as a one-dimensional nuisance parameter of the infinitesimal neighborhood models introduced in Section 1.2. In Rieder et al. (2001) we numerically solve the implicit equations for location, scale and linear regression models and calculate the increase of the maximum asymptotic risk over the minimax asymptotic risk in case that the optimally robust estimator for the false radius is used. More precisely, we determine the *inefficiency* - the limit

of the ratio of sample sizes such as to achieve the same accuracy asymptotically; i.e., in case of the MSE, we consider

$$\mathrm{relMSE}(\tilde{\eta}_{r_0}, r) = \frac{\mathrm{maxMSE}(\tilde{\eta}_{r_0}, r)}{\mathrm{maxMSE}(\tilde{\eta}_r, r)} \tag{2.2.1}$$

where

$$\mathrm{maxMSE}(\tilde{\eta}_{r_0}, r) = \mathrm{E}\,|\tilde{\eta}_{r_0}|^2 + r^2 \omega_*(\tilde{\eta}_{r_0})^2 \qquad * = c, v \tag{2.2.2}$$

That is, the maximum asymptotic MSE of the AL estimator with IC $\tilde{\eta}_{r_0}$ which is optimal for an infinitesimal neighborhood of radius $r_0 \in [0, \infty]$, is evaluated over an infinitesimal neighborhood of another radius $r \in [0, \infty]$, and is related to the minimax asymptotic MSE for that radius r. Often, we also use the term *subefficiency* which means inefficiency minus 1.

The trace of the asymptotic covariance of the MSE solution $\tilde{\eta}_r$ is increasing and the standardized asymptotic bias is decreasing in the radius $r \in (0, \infty)$. The proof of the following lemma is based on arguments provided by P. Ruckdeschel.

Lemma 2.2.1 *Let $\tilde{\eta}_r$ be the solution for radius $r \in (0, \infty)$ provided by Theorem 1.3.11. Then, $\mathrm{E}\,|\tilde{\eta}_r|^2$ is increasing and $\omega_*(\tilde{\eta}_r)$ is decreasing in r.*

Proof Given some $b \in (0, \infty)$ with $b \geq \omega_*^{\min}$, the solutions $\tilde{\eta}_b$ to the corresponding Hampel type problems (1.3.7) are given by Theorems 1.3.7 and 1.3.9. Obviously, $g(b) = \omega_*(\tilde{\eta}_b)^2 = b^2$ is non-negative, convex and strictly increasing in b; confer also Lemma 1.3.4. Moreover, given $t \in [0, 1]$ and $b_0, b_1 \in [\omega_*^{\min}, \infty)$, define $b_t = (1-t)b_0 + tb_1$. Then, $\eta_t = (1-t)\tilde{\eta}_{b_0} + t\tilde{\eta}_{b_1} \in \Psi_2^D$ with $|\eta_t| \leq b_t$ and

$$\mathrm{E}\,|\tilde{\eta}_{b_t}|^2 \leq \mathrm{E}\,|\eta_t|^2 \leq (1-t)\,\mathrm{E}\,|\tilde{\eta}_{b_0}|^2 + t\,\mathrm{E}\,|\tilde{\eta}_{b_1}|^2 \tag{2.2.3}$$

by the convexity of $|\cdot|^2$; i.e., $f(b) = \mathrm{E}\,|\tilde{\eta}_b|^2$ is convex in b. In addition, $f(b)$ is non-negative and strictly decreasing in b. Now, consider the corresponding MSE problem which reads

$$h(r, b) = f(b) + r^2 g(b) \tag{2.2.4}$$

and let b_r be the optimal clipping bound for radius r; i.e.,

$$h(r, b_r) = \min\{h(r, b) \,|\, b \in [\omega_*^{\min}, \infty)\} \tag{2.2.5}$$

By the convexity of f, g and h in b, one-sided derivatives in $b \in (\omega_*^{\min}, \infty)$ exist and we denote them by f', g' and $\partial_2 h$, respectively. Then, by convexity of h

$$\partial_2 h(r, b_r - 0) \leq 0 \leq \partial_2 h(r, b_r + 0) \tag{2.2.6}$$

for $b_r \in (\omega_*^{\min}, \infty)$ which implies

$$-\frac{f'(b_r - 0)}{g'(b_r - 0)} \geq r^2 \geq -\frac{f'(b_r + 0)}{g'(b_r + 0)} \tag{2.2.7}$$

where the convexity of f and g entails that $-f'$ and $1/g'$ are non-negative and decreasing. Now, assume $r_1 < r_2$ and $\omega_*^{\min} \leq b_{r_1} < b_{r_2}$. Then,

$$r_2^2 \leq -\frac{f'(b_{r_2} - 0)}{g'(b_{r_2} - 0)} \leq -\frac{f'(b_{r_1} + 0)}{g'(b_{r_1} + 0)} \leq r_1^2 \tag{2.2.8}$$

which is a contradiction. Hence, $b_r = \omega_*(\tilde{\eta}_r)$ is decreasing and consequentially $f(b_r) = \mathrm{E}\,|\tilde{\eta}_r|^2$ is increasing in $r \in (0,\infty)$. ////

Remark 2.2.2 As the notation in the proof suggests, the preceding lemma holds more generally, namely for all loss functions of the form

$$h(\alpha, x) = f(x) + \alpha g(x) \tag{2.2.9}$$

where $\alpha \in (0,\infty)$ and f and g are non-negative, convex and decreasing, respectively increasing in $x \in \mathbb{R}$. Hence, it for instance applies to the setup of Ruckdeschel and Rieder (2004); confer also Remark 1.3.3. ////

The following lemma shows that the MSE–inefficiency curves attain two relative maxima at the boundaries if we fix some interval for the neighborhood radius r.

Lemma 2.2.3 *Let $r \in [r_l, r_u]$ with $0 < r_l < r_u < \infty$.*

(a) *Then,*

$$\sup_{r\in[r_l,r_u]} \mathrm{relMSE}(\tilde{\eta}_s, r) \le \max\Big\{\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{E}\,|\tilde{\eta}_{r_l}|^2}, \frac{\omega_*(\tilde{\eta}_s)^2}{\omega_*(\tilde{\eta}_{r_u})^2}\Big\} \qquad \forall s \in [r_l, r_u] \tag{2.2.10}$$

and there exists some $r_0 \in [r_l, r_u]$ such that

$$\frac{\mathrm{E}\,|\tilde{\eta}_{r_0}|^2}{\mathrm{E}\,|\tilde{\eta}_{r_l}|^2} = \frac{\omega_*(\tilde{\eta}_{r_0})^2}{\omega_*(\tilde{\eta}_{r_u})^2} \tag{2.2.11}$$

Moreover, in case $r_l = 0$ and $r_u = \infty$,

$$\inf_{s\in[0,\infty]} \sup_{r\in[0,\infty]} \mathrm{relMSE}(\tilde{\eta}_s, r) = \frac{\mathrm{E}\,|\tilde{\eta}_{r_0}|^2}{\operatorname{tr} D\mathcal{I}^{-1}D^\tau} = \frac{\omega_*(\tilde{\eta}_{r_0})^2}{(\omega_*^{\min})^2} \tag{2.2.12}$$

(b) *It holds,*

$$\sup_{r\in[r_l,r_u]} \mathrm{relMSE}(\tilde{\eta}_s, r) = \mathrm{relMSE}(\tilde{\eta}_s, r_l) \vee \mathrm{relMSE}(\tilde{\eta}_s, r_u) \tag{2.2.13}$$

for all $s \in [r_l, r_u]$. Moreover, there exists some $r_0 \in [r_l, r_u]$ such that

$$\mathrm{relMSE}(\tilde{\eta}_{r_0}, r_l) = \mathrm{relMSE}(\tilde{\eta}_{r_0}, r_u) \tag{2.2.14}$$

and

$$\inf_{s\in[r_l,r_u]} \sup_{r\in[r_l,r_u]} \mathrm{relMSE}(\tilde{\eta}_s, r) = \mathrm{relMSE}(\tilde{\eta}_{r_0}, r_l) = \mathrm{relMSE}(\tilde{\eta}_{r_0}, r_u) \tag{2.2.15}$$

Proof

(a) Let $s, r \in [r_l, r_u]$ with $0 < r_l < r_u < \infty$. If $\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{E}\,|\tilde{\eta}_r|^2} \le \frac{\omega_*(\tilde{\eta}_s)^2}{\omega_*(\tilde{\eta}_r)^2}$,

$$\mathrm{E}\,|\tilde{\eta}_s|^2 + r^2\omega_*(\tilde{\eta}_s)^2 \le \frac{\omega_*(\tilde{\eta}_s)^2}{\omega_*(\tilde{\eta}_r)^2}\big(\mathrm{E}\,|\tilde{\eta}_r|^2 + r^2\omega_*(\tilde{\eta}_r)^2\big) \tag{2.2.16}$$

Otherwise,

$$\mathrm{E}\,|\tilde{\eta}_s|^2 + r^2\omega_*(\tilde{\eta}_s)^2 \le \frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{E}\,|\tilde{\eta}_r|^2}\big(\mathrm{E}\,|\tilde{\eta}_r|^2 + r^2\omega_*(\tilde{\eta}_r)^2\big) \tag{2.2.17}$$

Hence,

$$\mathrm{relMSE}(\tilde{\eta}_s, r) \le \max\left\{\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{E}\,|\tilde{\eta}_r|^2},\ \frac{\omega_*(\tilde{\eta}_s)^2}{\omega_*(\tilde{\eta}_r)^2}\right\} \tag{2.2.18}$$

Since $\mathrm{E}\,|\tilde{\eta}_r|^2$ and $\omega_*(\tilde{\eta}_r)$ are monotone increasing, respectively decreasing in $r \in [r_l, r_u]$ (cf. Lemma 2.2.1), we obtain (2.2.10). Moreover, $\mathrm{E}\,|\tilde{\eta}_s|^2/\,\mathrm{E}\,|\tilde{\eta}_{r_l}|^2$ is increasing in s and $\omega_*(\tilde{\eta}_s)^2/\omega_*(\tilde{\eta}_{r_u})^2$ is decreasing in s where

$$\lim_{s\downarrow r_l}\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{E}\,|\tilde{\eta}_{r_l}|^2} = 1 \qquad \text{resp.} \qquad \lim_{s\uparrow r_u}\frac{\omega_*(\tilde{\eta}_s)^2}{\omega_*(\tilde{\eta}_{r_u})^2} = 1 \tag{2.2.19}$$

by the continuity of $\mathrm{E}\,|\tilde{\eta}_r|^2$ and $\omega_*(\tilde{\eta}_r)$ in $r \in [r_l, r_u]$; confer Proposition 2.1.9 and Ruckdeschel (2005b). Hence, the intermediate value theorem yields some $r_0 \in [r_l, r_u]$ such that equality holds in (2.2.11). In case $r_l = 0$ and $r_u = \infty$, we obtain

$$\mathrm{relMSE}(\tilde{\eta}_s, 0) = \frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{tr}\,D\mathcal{I}^{-1}D^\tau} \qquad \text{and} \qquad \lim_{r\to\infty}\mathrm{relMSE}(\tilde{\eta}_s, r) = \frac{\omega_*(\tilde{\eta}_s)^2}{(\omega_*^{\min})^2} \tag{2.2.20}$$

Hence,

$$\inf_{s\in[0,\infty]}\sup_{r\in[0,\infty]}\mathrm{relMSE}(\tilde{\eta}_s, r) = \inf_{s\in(0,\infty)}\left[\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{\mathrm{tr}\,D\mathcal{I}^{-1}D^\tau} \vee \frac{\omega_*(\tilde{\eta}_s)^2}{(\omega_*^{\min})^2}\right] \tag{2.2.21}$$

and this infimum is attained by $s = r_0$.

(b) We first show that $r \mapsto \mathrm{relMSE}(\tilde{\eta}_s, r)$ is increasing for $r > s$. To simplify the notation, we identify $\omega_*(\tilde{\eta}_r)$ with b_r for all $r \in (0,\infty)$. Let $0 < s < r_1 < r_2 < \infty$, then, $\mathrm{E}\,|\tilde{\eta}_s|^2 \le \mathrm{E}\,|\tilde{\eta}_{r_1}|^2 \le \mathrm{E}\,|\tilde{\eta}_{r_2}|^2$ and $b_s \ge b_{r_1} \ge b_{r_2}$ (cf. Lemma 2.2.1); i.e.,

$$\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{b_s^2} \le \frac{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2}{b_{r_1}^2} \le \frac{\mathrm{E}\,|\tilde{\eta}_{r_2}|^2}{b_{r_2}^2} \tag{2.2.22}$$

Therefore, we obtain

$$\frac{1}{\frac{\mathrm{E}\,|\tilde{\eta}_s|^2}{b_s^2} + r_1^2} \ge \frac{1}{\frac{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2}{b_{r_1}^2} + r_1^2} \tag{2.2.23}$$

We multiply by $r_2^2 - r_1^2$ and rewrite this as

$$\frac{(r_2^2 - r_1^2)b_s^2}{\mathrm{E}\,|\tilde{\eta}_s|^2 + r_1^2 b_s^2} \ge \frac{(r_2^2 - r_1^2)b_{r_1}^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2 + r_1^2 b_{r_1}^2} \tag{2.2.24}$$

Now, we use $\mathrm{E}\,|\tilde{\eta}_{r_2}|^2 + r_2^2 b_{r_2}^2 \le \mathrm{E}\,|\tilde{\eta}_{r_1}|^2 + r_2^2 b_{r_1}^2$ and get for the right hand side of the previous inequality

$$\frac{(r_2^2 - r_1^2)b_{r_1}^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2 + r_1^2 b_{r_1}^2} \ge \frac{(r_2^2 - r_1^2)b_{r_1}^2 + \mathrm{E}\,|\tilde{\eta}_{r_2}|^2 + r_2^2 b_{r_2}^2 - \mathrm{E}\,|\tilde{\eta}_{r_1}|^2 - r_2^2 b_{r_1}^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2 + r_1^2 b_{r_1}^2} \tag{2.2.25}$$

$$= \frac{\mathrm{E}\,|\tilde{\eta}_{r_2}|^2 + r_2^2 b_{r_2}^2 - \mathrm{E}\,|\tilde{\eta}_{r_1}|^2 - r_1^2 b_{r_1}^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2 + r_1^2 b_{r_1}^2} \tag{2.2.26}$$

This leads us to

$$\left[\frac{(r_2^2-r_1^2)b_s^2}{\mathrm{E}\,|\tilde{\eta}_s|^2+r_1^2b_s^2}-\frac{\mathrm{E}\,|\tilde{\eta}_{r_2}|^2+r_2^2b_{r_2}^2-\mathrm{E}\,|\tilde{\eta}_{r_1}|^2-r_1^2b_{r_1}^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2+r_1^2b_{r_1}^2}\right]\frac{\mathrm{E}\,|\tilde{\eta}_s|^2+r_1^2b_s^2}{\mathrm{E}\,|\tilde{\eta}_{r_2}|^2+r_2^2b_{r_2}^2}\geq 0 \quad (2.2.27)$$

which can be simplified to

$$\frac{\mathrm{E}\,|\tilde{\eta}_s|^2+r_2^2b_s^2}{\mathrm{E}\,|\tilde{\eta}_{r_2}|^2+r_2^2b_{r_2}^2}-\frac{\mathrm{E}\,|\tilde{\eta}_s|^2+r_1^2b_s^2}{\mathrm{E}\,|\tilde{\eta}_{r_1}|^2+r_1^2b_{r_1}^2}\geq 0 \quad (2.2.28)$$

i.e., $\mathrm{relMSE}(\tilde{\eta}_s,r_2)\geq\mathrm{relMSE}(\tilde{\eta}_s,r_1)$. Analogously, we may verify $\mathrm{relMSE}(\tilde{\eta}_s,r)$ is decreasing for $r<s$. Hence, (2.2.13) holds. Moreover, we have

$$\begin{aligned}&\mathrm{E}\,|\tilde{\eta}_{r_2}|^2+s^2b_{r_2}^2-\mathrm{E}\,|\tilde{\eta}_{r_1}|^2-s^2b_{r_1}^2\\&\quad=(s^2-r_1^2)(b_{r_2}^2-b_{r_1}^2)+r_1^2(b_{r_2}^2-b_{r_1}^2)+\mathrm{E}\,|\tilde{\eta}_{r_2}|^2-\mathrm{E}\,|\tilde{\eta}_{r_1}|^2 \quad (2.2.29)\\&\quad=(s^2-r_1^2)(b_{r_2}^2-b_{r_1}^2)+\mathrm{E}\,|\tilde{\eta}_{r_2}|^2+r_1^2b_{r_2}^2-\mathrm{E}\,|\tilde{\eta}_{r_1}|^2-r_1^2b_{r_1}^2\geq 0 \quad (2.2.30)\end{aligned}$$

That is,

$$\mathrm{relMSE}(\tilde{\eta}_{r_1},s)\leq\mathrm{relMSE}(\tilde{\eta}_{r_2},s) \quad (2.2.31)$$

for $s<r_1<r_2$. Analogously, we may show

$$\mathrm{relMSE}(\tilde{\eta}_{r_1},s)\leq\mathrm{relMSE}(\tilde{\eta}_{r_2},s) \quad (2.2.32)$$

for $s>r_1>r_2$. Thus, using the fact $\mathrm{relMSE}(\tilde{\eta}_{r_l},r_l)=1=\mathrm{relMSE}(\tilde{\eta}_{r_u},r_u)$ and the continuity of $\mathrm{relMSE}(\tilde{\eta}_r,s)$ in r, which is entailed by the continuity of $\mathrm{E}\,|\tilde{\eta}_r|^2$ and b_r in r (cf. Proposition 2.1.9 and Ruckdeschel (2005b)), the shown monotonicity (2.2.31), respectively (2.2.32) together with the intermediate value theorem implies (2.2.14) and (2.2.15). ////

Remark 2.2.4 **(a)** The proof of the preceding lemma is based on arguments provided by P. Ruckdeschel.

(b) If $\bar{r}=0$ (cf. Remark 3.1.3 (b)), respectively $r_l\geq\bar{r}$, there is only one solution $\tilde{\eta}$ on $[0,\infty]$, respectively $[r_l,r_u]$; confer Propositions 2.1.3 and 2.1.5. Hence, we have to consider bounded intervals in the previous lemma.

(c) The calculation of some radius r_0 such that both boundary values are equal leads to an AL estimator that is *radius–minimax*; i.e., minimizes the maximum inefficiency over the given radius range. We call this radius r_0 *least favorable*. In case the true radius is completely unknown, respectively unknown except to belong to some radius interval, we recommend to use this optimally robust estimator.

(d) Ruckdeschel and Rieder (2004) prove that Lemma 2.2.3 (a) holds more generally for a large class of optimization problems of form (1.3.10) where G in addition has to fulfill a certain homogeneity condition; confer Theorem 6.1 (a) (idid.). In particular, by parameterizing $\tilde{\eta}$ not by the radius r but by the optimal clipping bound b, they show, that the radius–minimax IC for completely unknown radius r is the same for all L_q-risks ($q\in[1,\infty)$); confer Theorem 6.1 (b) (idid.).

(e) In Rieder et al. (2001) and also in this thesis we consider the cases that the radius r is completely unknown, respectively unknown up to a factor of 1/3 or 1/2;

i.e., any r_3 or r_2 such that the true radius r certainly would stay within $[\frac{1}{3}r_3, 3r_3]$ or $[\frac{1}{2}r_2, 2r_2]$, respectively. In a second step, we then determine least favorable values of r_3 and r_2 by maximizing the minimax subefficiencies over $[\frac{1}{3}r_3, 3r_3]$ and $[\frac{1}{2}r_2, 2r_2]$, respectively. ////

In the following remark we state the conclusions of Rieder et al. (2001), Section 1.6.

Remark 2.2.5 (a) The minimax subefficiency is small. Small in comparison with the most robust estimators, and small for practical purposes. Consistent estimation of the radius from the data hence seems neither necessary nor worthwhile – however, under the provision that the radius–minimax robust estimator is employed.

(b) The least favorable radii are small. This surprising fact seems to confirm Huber (1997), p 61, who distinguishes robustness from diagnostics by its purpose to safeguard against – as opposed to find and identify – deviations from the assumptions; in particular, to safeguard against deviations below or near the limits of detectability. Like Huber (1997), the small least favorable radii we obtain might question the breakdown literature, which is concerned only with (stability under) large contamination and, at most, (efficiency under) zero contamination. ////

2.3 One-Step Construction

2.3.1 Motivation and Setup

Having calculated the optimally robust IC, we finally need to construct some asymptotic estimator S, without knowing the parameter θ, such that S is asymptotically linear at P_θ with IC ψ_θ. Moreover, the risk of this estimator S should not increase if we interchange $\lim_M \lim_n$ and $\sup_q$ and pass over from simple perturbations to full neighborhoods; confer Remark 1.3.3 (a). In case of the one-step estimators we need initial estimators σ that are strict (take values only in Θ), $\sqrt{n}$ consistent on the full neighborhood system $\mathcal{U}(\theta)$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup \left\{ Q_n^{(n)}(\sqrt{n}|\sigma_n - \theta| > M) \,\middle|\, Q_{n,i} \in U(\theta, r/\sqrt{n}) \right\} = 0 \tag{2.3.1}$$

with $Q_n^{(n)} = Q_{n,1} \otimes \cdots \otimes Q_{n,n}$ and for technical reasons suitably discretized; confer Subsection 6.4.2 of Rieder (1994). In addition, the following conditions are necessary which are formulated about θ and understood to hold for all bounded sequences $\sqrt{n}\,(\theta_n - \theta)$ in $\mathbb{R}^k$, as $n \to \infty$ (cf. Rieder (1994), pp 247, 256):

$$\lim_{n\to\infty} \int |\psi_{n,\theta_n} - \psi_\theta|^2 \, dP_\theta = 0 \tag{2.3.2}$$

$$\sup_\Omega |\psi_{n,\theta_n}| = o(\sqrt[4]{n}) \tag{2.3.3}$$

$$\int \psi_{n,\theta_n} \, dP_{\theta_n} = o\left(\frac{1}{\sqrt{n}}\right) \tag{2.3.4}$$

$$\lim_{n\to\infty} \sup_\Omega |\psi_{n,\theta_n} - \psi_{n,\theta}| = 0 \tag{2.3.5}$$

where $\psi_{n,\theta} \in L_2^k(P_\theta)$, $\theta \in \Theta$, $n \in \mathbb{N}$, is an approximating sequence of the IC $\psi_\theta \in \Psi_2(\theta)$ by a sequence of families of suitable smooth and bounded functions. Then, the following asymptotic normality on full total variation balls and therefore also on full contamination balls holds for the one-step estimator S defined by

$$S_n = \sigma_n^* + \frac{1}{n}\sum_{i=1}^{n} \psi_{n,\sigma_n^*(y_1,\ldots,y_n)}(y_i) \tag{2.3.6}$$

with σ^* the discretized version of σ.

Theorem 2.3.1 *Assume (2.3.2)–(2.3.5). Then, for all $r \in (0,\infty)$ and all arrays $Q_{n,i} \in B_v(P_\theta, r/\sqrt{n})$,*

$$\sqrt{n}\left(S_n - \theta - \frac{1}{n}\sum_{i=1}^{n}\int \psi_{n,\theta}\, dQ_{n,i}\right)\left(Q_n^{(n)}\right) \xrightarrow{w} \mathcal{N}\left(0, \mathrm{Cov}_\theta(\psi_\theta)\right) \tag{2.3.7}$$

Proof Rieder (1994), Theorem 6.4.8 (b) ////

Remark 2.3.2 (a) In this thesis we mainly consider one-step constructions. Exceptions are Part V in which the finite-sample minimax estimators are M estimators via construction and Section 9.3 in which we treat time series models. As already noted in Remark 1.3.3 (a), the construction problem is still unsolved for time series models. Moreover, the construction problem in case of conditional regression neighborhoods, which are considered in Parts III and V, is a separate issue that we do not treat in this thesis.

(b) The use of one-step estimators has a clear advantage: Given some strict and $\sqrt{n}$ consistent starting estimator σ the one-step estimator S is very fast to compute and additionally unique. Estimates derived from M equations, however, besides being more difficult to determine, need not be unique; confer for example Reeds (1985). Moreover, the higher order asymptotics of Ruckdeschel (2004b) and Ruckdeschel (2005e) for the MSE show that at least in the location model and under symmetry the M estimators and the one-step estimators are asymptotically equivalent even up to second order (without symmetry this is true for the two-step estimator).

(c) The Kolmogorov(–Smirnov) and the Cramér-von Mises minimum distance estimator are strict and $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta)$ and $\mathcal{U}_\mu(\theta)$, respectively; confer Theorems 6.3.7 and 6.3.8 of Rieder (1994). Moreover, by Theorem 2.3.9 and Proposition 2.3.12 the median and the median absolute deviation are $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta)$. ////

2.3.2 Sufficient Conditions for Mean Square Error Solution

We now consider the optimally robust ICs $\tilde{\eta}_\theta$ specified in Section 1.3 for the full parameter; i.e., $D = \mathbb{I}_k$. We can verify (2.3.2)–(2.3.5) by choosing $\psi_{n,\theta} \equiv \tilde{\eta}_\theta$ for all $\theta \in \Theta$, $n \in \mathbb{N}$ and $\theta_n = \theta + h/\sqrt{n}$ for some $h \in \mathbb{R}^k$. Hence, (2.3.4) is fulfilled. Moreover, (2.3.2) is implied by (2.3.5). Condition (2.3.3) under the additional

conditions (2.1.102) and (2.1.103) follows by Theorem 2.1.11 and the boundedness of the Lagrange multipliers. The conditions (2.3.2) and (2.3.5) for $\psi_\theta = \tilde{\eta}_\theta$ are, under additional assumptions, verified in the following theorem.

Theorem 2.3.3 *Let $D = \mathbb{I}_k$ and $\psi_\theta = \tilde{\eta}_\theta$ in (2.3.2)–(2.3.5), where $\tilde{\eta}_\theta$ is the solution to the MSE problem (1.3.5) for radius $r \in (0,\infty)$ given in Theorem 1.3.11. Further assume*

$$\operatorname{tr} \mathcal{I}_{\theta_n} \longrightarrow \operatorname{tr} \mathcal{I}_\theta \qquad \text{as } n \to \infty \tag{2.1.102}$$

where $\theta_n \to \theta$ as $n \to \infty$.

(a) *If*

$$\Lambda_{\theta_n} \xrightarrow{P_\theta} \Lambda_\theta \qquad \text{as } n \to \infty \tag{2.3.8}$$

then condition (2.3.2) holds.

(b) *Assume the Lagrange multipliers contained in the solution $\tilde{\eta}_\theta$ are unique and (2.1.103) holds; i.e., $\mathcal{L}_{P_{\theta_n}}(\Lambda_{\theta_n}) \xrightarrow{w} \mathcal{L}_{P_\theta}(\Lambda_\theta)$ as $n \to \infty$.*

$* = c$: *Let*

$$\mathcal{D} = \left\{x \in \Omega \,\middle|\, |A_{\theta_n}\Lambda_{\theta_n}(x) - a_{\theta_n}| \le b_{\theta_n}\right\} \cup \left\{x \in \Omega \,\middle|\, |A_\theta\Lambda_\theta(x) - a_\theta| \le b_\theta\right\} \tag{2.3.9}$$

If

$$\sup_{\mathcal{D}} \left\{|\Lambda_{\theta_n} - \Lambda_\theta|\right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.10}$$

and

$$\sup_{{}^c\mathcal{D}} \left\{\frac{|\Lambda_{\theta_n} - \Lambda_\theta|}{|A_\theta\Lambda_\theta - a_\theta|}\right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.11}$$

condition (2.3.5) holds.

$* = v$, $k = 1$: *Let*

$$\mathcal{D} = \left\{x \in \Omega \,\middle|\, c_{\theta_n} \le A_{\theta_n}\Lambda_{\theta_n}(x) \le c_{\theta_n} + b_{\theta_n}\right\} \cup \left\{x \in \Omega \,\middle|\, c_\theta \le A_\theta\Lambda_\theta(x) \le c_\theta + b_\theta\right\} \tag{2.3.12}$$

If

$$\sup_{\mathcal{D}} \left\{|\Lambda_{\theta_n} - \Lambda_\theta|\right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.13}$$

and

$$\sup_{{}^c\mathcal{D}} \left\{\frac{|\Lambda_{\theta_n} - \Lambda_\theta|}{|\Lambda_\theta|}\right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.14}$$

condition (2.3.5) holds.

Proof We fix some radius $r \in (0,\infty)$. To simplify the notation, we omit θ as an index and identify θ_n by n.

(a) As a consequence of Theorem 2.1.11, we get $\mathrm{E}_n\,|\tilde{\eta}_n|^2 \to \mathrm{E}\,|\tilde{\eta}|^2$ as $n \to \infty$; confer Remark 2.1.12 (b). Moreover, under condition (2.3.8) every subsequence $(n') \subset (n)$ contains a further subsequence $(m) \subset (n')$ such that $\Lambda_m \to \Lambda$ a.e. P. Since $\tilde{\eta}_n(u_n) \to \tilde{\eta}(u)$ for $u_n \to u$ as shown in the proof of Theorem 2.1.11, also

$\tilde{\eta}_m(\Lambda_m) \to \tilde{\eta}(\Lambda)$ a.e. P which implies $\tilde{\eta}_n \xrightarrow{P} \tilde{\eta}$. Therefore, we can apply Vitali's Theorem (cf. Corollary A.2.3 of Rieder (1994)) which yields condition (2.3.2).

(b) We are in the situation of Theorem 2.1.11; i.e., (2.1.104) and (2.1.105) hold. $\boldsymbol{* = c:}$ We define $Y_n := A_n\Lambda_n - a_n$ and $Y := A\Lambda - a$ and partition Ω in the following four parts

$\cap$	$\{\lvert Y_n\rvert \le b_n\}$	$\{\lvert Y_n\rvert > b_n\}$
$\{\lvert Y\rvert \le b\}$	Ω_1	Ω_3
$\{\lvert Y\rvert > b\}$	Ω_2	Ω_4

On Ω_1 we obtain,

$$\sup_{\Omega_1} \left|\tilde{\eta}_n - \tilde{\eta}\right| = \sup_{\Omega_1} \left|A_n\Lambda_n - a_n - A\Lambda + a\right| \tag{2.3.15}$$

$$\le \sup_{\Omega_1} \left|A_n\Lambda_n - A\Lambda_n + A\Lambda_n - A\Lambda\right| + \left|a_n - a\right| \tag{2.3.16}$$

where the second summand on the RHS tends to 0 as $n \to \infty$ by (2.1.105). Furthermore,

$$\sup_{\Omega_1} \left|A_n\Lambda_n - A\Lambda_n + A\Lambda_n - A\Lambda\right|$$

$$\le \|A_n - A\|_{\mathrm{op}} \sup_{\Omega_1} |\Lambda_n| + \|A\|_{\mathrm{op}} \sup_{\Omega_1} |\Lambda_n - \Lambda| \tag{2.3.17}$$

where the first summand on the RHS tends to 0 as $n \to \infty$ by (2.1.105) and the uniform boundedness of Λ_n on Ω_1 which is a consequence of the uniform boundedness of A_n, a_n and b_n (cf. proof of Theorem 2.1.11). Since $\|A\|_{\mathrm{op}}$ is bounded, a sufficient condition for the convergence (2.3.5) on Ω_1 is

$$\sup_{\Omega_1} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.18}$$

On Ω_2 we get,

$$\sup_{\Omega_2} \left|\tilde{\eta}_n - \tilde{\eta}\right| = \sup_{\Omega_2} \left|Y_n - b\frac{Y}{|Y|}\right| = \sup_{\Omega_2} \left|Y_n - Y + Y - b\frac{Y}{|Y|}\right| \tag{2.3.19}$$

$$\le \sup_{\Omega_2} \left|Y_n - Y\right| + \sup_{\Omega_2} \left||Y| - b\right| \tag{2.3.20}$$

By analogous arguments as in case of Ω_1, a sufficient condition for the convergence of the first summand on the RHS is,

$$\sup_{\Omega_2} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.21}$$

For the second summand on the RHS we obtain,

$$\sup_{\Omega_2} \left||Y| - b\right| = \sup_{\Omega_2} |Y| - b = \sup_{\Omega_2} |Y| - b_n + b_n - b \tag{2.3.22}$$

$$\le \sup_{\Omega_2} |Y| - \sup_{\Omega_2} |Y_n| + b_n - b \tag{2.3.23}$$

$$\le \sup_{\Omega_2} |Y - Y_n| + b_n - b \tag{2.3.24}$$

which tends to 0 as $n \to \infty$ by (2.3.21) and (2.1.104). On Ω_3 we can argue analogously to the case Ω_2. We only need to interchange the role of Y_n and Y, respectively b_n and b. This leads to the sufficient condition

$$\sup_{\Omega_3} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.25}$$

for the convergence (2.3.5) on Ω_3. On Ω_4 we get,

$$\sup_{\Omega_4} |\tilde{\eta}_n - \tilde{\eta}| = \sup_{\Omega_4} \Big| b_n \frac{Y_n}{|Y_n|} - b \frac{Y}{|Y|} \Big| \tag{2.3.26}$$

$$= \sup_{\Omega_4} \Big| b_n \frac{Y_n}{|Y_n|} - b \frac{Y_n}{|Y_n|} + b \frac{Y_n}{|Y_n|} - b \frac{Y}{|Y|} \Big| \tag{2.3.27}$$

$$\leq |b_n - b| + b \sup_{\Omega_4} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| \tag{2.3.28}$$

where the first summand on the RHS tends to 0 as $n \to \infty$ by (2.1.104). Since b is bounded a first sufficient condition for the convergence (2.3.5) on Ω_4 is

$$\sup_{\Omega_4} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.29}$$

Furthermore, we get

$$\sup_{\Omega_4} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| = \sup_{\Omega_4} \Big| \frac{Y_n |Y|}{|Y_n||Y|} - \frac{Y_n |Y_n|}{|Y||Y_n|} + \frac{Y_n |Y_n|}{|Y||Y_n|} - \frac{Y}{|Y|} \Big| \tag{2.3.30}$$

$$\leq \sup_{\Omega_4} \Big| \frac{|Y| - |Y_n|}{|Y|} \Big| + \sup_{\Omega_4} \frac{|Y_n - Y|}{|Y|} \tag{2.3.31}$$

$$\leq 2 \sup_{\Omega_4} \frac{|Y_n - Y|}{|Y|} \tag{2.3.32}$$

Hence, another sufficient condition for the convergence (2.3.5) on Ω_4 is

$$\sup_{\Omega_4} \frac{|Y_n - Y|}{|Y|} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.33}$$

Moreover, we obtain

$$\sup_{\Omega_4} \frac{|Y_n - Y|}{|Y|} = \sup_{\Omega_4} \frac{|A_n(\Lambda_n - z_n) - A_n(\Lambda - z) + A_n(\Lambda - z) - A(\Lambda - z)|}{|Y|} \tag{2.3.34}$$

$$\leq \|A_n\|_{\text{op}} \sup_{\Omega_4} \frac{|\Lambda_n - z_n - \Lambda + z|}{|Y|} + \|A_n - A\|_{\text{op}} \sup_{\Omega_4} \frac{|\Lambda - z|}{|Y|} \tag{2.3.35}$$

where the second summand on the RHS tends to 0 as $n \to \infty$ by (2.1.105) and

$$\sup_{\Omega_4} \frac{|\Lambda - z|}{|Y|} = \sup_{\Omega_4} \frac{|\Lambda - z|}{|A(\Lambda - z)|} \leq \sup_{u \in \mathbb{R}^k, u \neq 0} \frac{|u|}{|Au|} \leq \|A^{-1}\|_{\text{op}} \tag{2.3.36}$$

as $\Lambda \neq 0$ and therefore $\Lambda - z \neq 0$ on Ω_4 since $A \succ 0$. In addition, we use the regularity of A; i.e., $u = A^{-1}Au$, hence $|Au| \geq \|A^{-1}\|_{\text{op}}^{-1}|u|$. Since $\|A_n\|_{\text{op}}$ is bounded uniformly in $n \in \mathbb{N}$, a further sufficient condition for the convergence (2.3.5) on Ω_4 can be obtained as follows,

$$\sup_{\Omega_4} \frac{|\Lambda_n - z_n - \Lambda + z|}{|Y|} \leq \sup_{\Omega_4} \frac{|\Lambda_n - \Lambda|}{|Y|} + |z_n - z| \sup_{\Omega_4} \frac{1}{|Y|} \tag{2.3.37}$$

where the second summand on the RHS tends to 0 as $n \to \infty$ since $|Y| > b > 0$ on Ω_4; i.e., (2.3.11) is sufficient on Ω_4. So, putting the results for the four parts $\Omega_1, \ldots, \Omega_4$ of Ω together, leads to the sufficient conditions (2.3.10) and (2.3.11).

$* = v, k = 1$: We define $Y_n := A_n\Lambda_n$ and $Y := A\Lambda$ and partition Ω in the following nine parts

$\cap$	$\{c_n \leq Y_n \leq c_n + b_n\}$	$\{Y_n < c_n\}$	$\{Y_n > c_n + b_n\}$
$\{c < Y < c + b\}$	Ω_1	Ω_3	Ω_5
$\{Y < c\}$	Ω_2	Ω_6	Ω_8
$\{Y > c + b\}$	Ω_4	Ω_7	Ω_9

A sufficient condition for the convergence (2.3.5) on Ω_1 analogously to (2.3.17) is

$$\sup_{\Omega_1} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.38}$$

On Ω_2 we obtain,

$$\sup_{\Omega_2} |\tilde{\eta}_n - \tilde{\eta}| = \sup_{\Omega_2} |A_n\Lambda_n - c| \leq \sup_{\Omega_2} |A_n\Lambda_n - A\Lambda| + \sup_{\Omega_2} |A\Lambda - c| \tag{2.3.39}$$

Thus, a sufficient condition for the convergence of the first summand on the RHS is,

$$\sup_{\Omega_2} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.40}$$

For the second summand on the RHS we get,

$$\sup_{\Omega_2} |Y - c| = c - \sup_{\Omega_2} Y = c - \sup_{\Omega_2} Y_n + \sup_{\Omega_2} Y_n - \sup_{\Omega_2} Y \tag{2.3.41}$$

$$\leq c - c_n + \sup_{\Omega_2} |Y_n - Y| \tag{2.3.42}$$

which tends to 0 as $n \to \infty$ by (2.1.104) and (2.3.40). On Ω_3 we can argue analogously to the case Ω_2; i.e., this leads to the sufficient condition

$$\sup_{\Omega_3} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.43}$$

for the convergence (2.3.5) on Ω_3. On Ω_4 we get,

$$\sup_{\Omega_4} |\tilde{\eta}_n - \tilde{\eta}| = \sup_{\Omega_4} |A_n\Lambda_n - (c+b)| \leq \sup_{\Omega_4} |A_n\Lambda_n - A\Lambda| + \sup_{\Omega_4} |A\Lambda - (c+b)| \tag{2.3.44}$$

Hence, a sufficient condition for the convergence of the first summand on the RHS is,

$$\sup_{\Omega_4} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.45}$$

where it holds for the second summand on the RHS,

$$\begin{aligned}
\sup_{\Omega_4} |Y - (c+b)| &= \sup_{\Omega_4} Y - (c+b) && (2.3.46)\\
&= \sup_{\Omega_4} Y - (c_n + b_n) + (c_n + b_n) - (c+b) && (2.3.47)\\
&\le \sup_{\Omega_4} Y - \sup_{\Omega_4} Y_n + c_n - c + b_n - b && (2.3.48)\\
&\le \sup_{\Omega_4} |Y_n - Y| + c_n - c + b_n - b && (2.3.49)
\end{aligned}$$

This tends to 0 as $n \to \infty$ by (2.1.104) and (2.3.45). On Ω_5 we can argue analogously to the case Ω_4. That is, we get the sufficient condition

$$\sup_{\Omega_5} |\Lambda_n - \Lambda| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.50}$$

for the convergence (2.3.5) on Ω_5. On Ω_6 and Ω_9 we have,

$$\sup_{\Omega_6} |\tilde{\eta}_n - \tilde{\eta}| = |c_n - c| \qquad \text{and} \qquad \sup_{\Omega_9} |\tilde{\eta}_n - \tilde{\eta}| = |c_n + b_n - c - b| \tag{2.3.51}$$

which tends to 0 as $n \to \infty$ by (2.1.104). On Ω_7 we get,

$$\sup_{\Omega_7} |\tilde{\eta}_n - \tilde{\eta}| = (c+b) - c_n \tag{2.3.52}$$

In case $-c \ge (c+b)$ we obtain,

$$\begin{aligned}
(c+b) - c_n &\le -c - c_n && (2.3.53)\\
&= -c + c_n - c_n \sup_{\Omega_7} \frac{Y}{|Y|} + c_n \sup_{\Omega_7} \frac{Y_n}{|Y_n|} && (2.3.54)\\
&\le c_n - c - c_n \sup_{\Omega_7} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| && (2.3.55)
\end{aligned}$$

In case $-c < (c+b)$ we get,

$$\begin{aligned}
(c+b) - c_n &= -(c+b) \sup_{\Omega_7} \frac{Y_n}{|Y_n|} - c_n \le c \sup_{\Omega_7} \frac{Y_n}{|Y_n|} - c_n && (2.3.56)\\
&= c \sup_{\Omega_7} \frac{Y_n}{|Y_n|} - c \sup_{\Omega_7} \frac{Y}{|Y|} + c - c_n && (2.3.57)\\
&\le c \sup_{\Omega_7} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| + c - c_n && (2.3.58)
\end{aligned}$$

By (2.1.104), a (first) sufficient condition for the convergence (2.3.5) on Ω_7 is

$$\sup_{\Omega_7} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.59}$$

Analogously, to the case $* = c$ and Ω_4 this leads to the sufficient condition

$$\sup_{\Omega_7} \frac{|Y_n - Y|}{|Y|} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.60}$$

where

$$\sup_{\Omega_7} \frac{|Y_n - Y|}{|Y|} = \sup_{\Omega_7} \frac{|A_n\Lambda_n - A_n\Lambda + A_n\Lambda - A\Lambda|}{|Y|} \tag{2.3.61}$$

$$\leq \frac{A_n}{A} \sup_{\Omega_7} \frac{|\Lambda_n - \Lambda|}{|\Lambda|} + \frac{|A_n - A|}{A} \tag{2.3.62}$$

Since $A > 0$ and (2.1.104), we get the sufficient condition

$$\sup_{\Omega_7} \frac{|\Lambda_n - \Lambda|}{|\Lambda|} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.63}$$

On Ω_8 it holds,

$$\sup_{\Omega_8} |\tilde{\eta}_n - \tilde{\eta}| = (c_n + b_n) - c \tag{2.3.64}$$

In case $-c \leq (c + b)$ we obtain,

$$(c_n + b_n) - c \leq (c_n + b_n) + (c + b) \tag{2.3.65}$$

$$= (c_n + b_n) - (c + b) - (c + b) \sup_{\Omega_8} \frac{Y}{|Y|} + (c + b) \sup_{\Omega_8} \frac{Y_n}{|Y_n|} \tag{2.3.66}$$

$$\leq c_n - c + b_n - b - (c + b) \sup_{\Omega_8} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| \tag{2.3.67}$$

In case $-c > (c + b)$ we get,

$$(c_n + b_n) - c = (c_n + b_n) + c \sup_{\Omega_8} \frac{Y}{|Y|} \leq (c_n + b_n) - (c + b) \sup_{\Omega_8} \frac{Y}{|Y|} \tag{2.3.68}$$

$$= (c_n + b_n) - (c + b) + (c + b) \sup_{\Omega_8} \frac{Y_n}{|Y_n|} - (c + b) \sup_{\Omega_8} \frac{Y}{|Y|} \tag{2.3.69}$$

$$\leq c_n - c + b_n - b + (c + b) \sup_{\Omega_8} \Big| \frac{Y_n}{|Y_n|} - \frac{Y}{|Y|} \Big| \tag{2.3.70}$$

Analogously to the case Ω_7 this leads to the sufficient condition

$$\sup_{\Omega_8} \frac{|\Lambda_n - \Lambda|}{|\Lambda|} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{2.3.71}$$

Now, putting the results for the nine parts $\Omega_1, \ldots, \Omega_9$ of Ω together, we obtain the sufficient conditions (2.3.13) and (2.3.14). ////

Remark 2.3.4 (a) If the considered parametric model has a certain equivariance structure like the regression, respectively regression and scale model, conditions (2.1.102) and (2.1.103) are automatically fulfilled. In case of the regression and scale model we have

$$\mathcal{L}_\theta(\Lambda_\theta) = \frac{1}{\sigma}\,\mathcal{L}_{\theta_0}(\Lambda_{\theta_0}) \qquad \text{and} \qquad \mathcal{I}_\theta = \frac{1}{\sigma^2}\mathcal{I}_{\theta_0} \tag{2.3.72}$$

with $\theta = (\beta, \sigma)^\tau$ and $\theta_0 = (0,1)^\tau$; confer Subsection 7.1.1 and Section 7.4.

(b) As shown in the proof of part (b) of the preceding theorem, conditions (2.3.10) and (2.3.29), respectively (2.3.10) and (2.3.33) are also sufficient for (2.3.5) in case $* = c$. Moreover, similar sufficient conditions can be read off from the proof of part (b) in case $* = v, k = 1$.

(c) Under the assumptions of part (b) of Theorem 2.3.3, Theorem 2.3.1 is in force where

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{Q_n \in B_*(P_\theta, r/\sqrt{n})} \sqrt{n} \left| \int \tilde{\eta}_{\theta_n}\, dQ_n \right| \\
&\qquad = \lim_{n\to\infty} \sup_{Q_n \in B_*(P_\theta, r/\sqrt{n})} \sqrt{n} \left| \int \tilde{\eta}_{\theta_n}\, (dQ_n - dP_{\theta_n}) \right| && (2.3.73)\\
&\qquad \le r \lim_{n\to\infty} b_{\theta_n} \overset{\text{Thm. } 2.1.11}{=} r b_\theta && (2.3.74)
\end{aligned}$$

Hence, the asymptotic minimax MSE on full contamination, respectively full total variation balls equals the minimax asymptotic MSE over simple perturbations which was derived in Section 1.3. ////

2.3.3 Sufficient Conditions in case of Exponential Families

In the subsequent lemma we show that Theorem 2.3.3 holds for a special class of exponential families. We assume a parametric family $\mathcal{P}$ of probability measures on some measurable space $(\Omega, \mathcal{A})$, whose parameter space Θ is an open subset of some finite dimensional $\mathbb{R}^k$. More precisely, $\mathcal{P}$ is of the form

$$\mathcal{P} = \{P_\theta \,|\, dP_\theta = p_\theta\, d\mu,\ \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}) \tag{2.3.75}$$

with some σ-finite measure μ on $(\Omega, \mathcal{A})$ and

$$p_\theta(x) = \exp\left\{\zeta(\theta)^\tau T(x) - \beta(\theta)\right\} h(x) \tag{2.3.76}$$

where $\alpha\colon \Theta \to \mathbb{R}^k$, $h\colon (\Omega, \mathcal{A}) \to ([0,\infty), \mathbb{B}\cap[0,\infty))$, $\beta(\theta) = \log \int \exp\{\zeta(\theta)^\tau T\}\, h\, d\mu$ and $T\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^k, \mathbb{B}^k)$ with $\mathrm{Cov}_\theta\, T \succ 0$.

Remark 2.3.5 (a) The family $\mathcal{P}$ forms a k-dimensional exponential family of full rank as introduced in Section 1.5 of Lehmann and Casella (1998) or in Section 1.7 of Witting (1985), respectively.

(b) The set $\zeta \in \mathbb{R}^k$ for which

$$0 < \int \exp\left\{\zeta^\tau T(x)\right\} h(x)\, \mu(dx) < \infty \tag{2.3.77}$$

is called the *natural parameter space* Z_* of an exponential family $\mathcal{P}$; confer pp 149 of Witting (1985) or p 24 of Lehmann and Casella (1998), respectively. ////

Lemma 2.3.6 *Consider the exponential family $\mathcal{P}$ of full rank given by (2.3.75) and (2.3.76). Let ζ be continuously differentiable in $\theta \in \Theta$ with regular Jacobian matrix $\mathcal{J}_\zeta$ and assume ζ maps Θ into the interior of the natural parameter set Z_* of $\mathcal{P}$.*

(a) *Then, $\mathcal{P}$ is L_2-differentiable with L_2 derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by*

$$\Lambda_\theta(x) = \mathcal{J}_\zeta^\tau \big(T(x) - \mathrm{E}_\theta\, T\big) \qquad \mathcal{I}_\theta = \mathcal{J}_\zeta^\tau\, \mathrm{Cov}_\theta(T) \mathcal{J}_\zeta \tag{2.3.78}$$

Moreover, conditions (2.1.102), (2.1.103) and (2.3.8) hold.

(b) *Assume the Lagrange multipliers A_θ and a_θ contained in the solution to the MSE problem (1.3.5) for radius $r \in (0, \infty)$ are unique. Then, conditions (2.3.10) and (2.3.11), respectively conditions (2.3.13) and (2.3.14) hold.*

Proof

(a) We have $\beta = \kappa \circ \zeta$ where κ is infinitely differentiable in ζ for all ζ in the interior of the natural parameter space Z_*; confer Witting (1985), Satz 1.164 (a). Hence, an application of the chain rule yields, β is continuously differentiable in $\theta \in \Theta$. That is,

$$\Lambda_\theta(x) := \nabla_\theta \log p_\theta(x) = \mathcal{J}_\zeta^\tau T(x) - \nabla_\theta \beta(\theta) \tag{2.3.79}$$

exists and is continuous in $\theta \in \Theta$ for all $x \in \Omega$. Since $\mathrm{E}_\theta\, \Lambda_\theta = 0$, we obtain $\nabla_\theta \beta(\theta) = \mathcal{J}_\zeta^\tau\, \mathrm{E}_\theta\, T$; i.e.,

$$\Lambda_\theta(x) = \mathcal{J}_\zeta^\tau \big(T(x) - \mathrm{E}_\theta\, T\big) \tag{2.3.80}$$

Consequentially,

$$\mathcal{I}_\theta = \mathrm{E}_\theta\, \Lambda_\theta \Lambda_\theta^\tau = \mathcal{J}_\zeta^\tau\, \mathrm{Cov}_\theta(T) \mathcal{J}_\zeta \succ 0 \tag{2.3.81}$$

as $\mathcal{J}_\zeta$ is regular for all $\theta \in \Theta$ and $\mathrm{Cov}_\theta(T) \succ 0$. Moreover, $\mathcal{I}_\theta$ is well defined and continuous in $\theta \in \Theta$ since all moments of T exist and $\mathrm{Cov}_\theta(T) = \tau \circ \zeta$ where τ is infinitely differentiable for all ζ in the interior of the natural parameter space Z_*; confer Witting (1985), Satz 1.164 (a); i.e., $\mathrm{Cov}_\theta(T)$ is even continuously differentiable in θ. Thus, $\mathcal{P}$ is L_2-differentiable by Lemma A.3 of Hajek (1972); confer also Satz 1.194 of Witting (1985) and Lemma 7.6 of van der Vaart (1998), respectively. In particular, the continuity of Λ_θ and $\mathcal{I}_\theta$ in θ implies conditions (2.1.102), (2.1.103) and (2.3.8).

(b) Let $r \in (0, \infty)$ and $\theta \in \Theta$ be fixed and $(\theta_n)_{n \in \mathbb{N}} \subset \Theta$ be some sequence such that $\theta_n \to \theta$ as $n \to \infty$.

$\boldsymbol{* = c}$**:** It holds,

$$\begin{aligned}
|A_\theta \Lambda_\theta - a_\theta| &\geq |A_\theta \mathcal{J}_\zeta^\tau (T - \mathrm{E}_\theta\, T)| - |a_\theta| && (2.3.82)\\
&\geq |A_\theta \mathcal{J}_\zeta^\tau T| - |A_\theta \mathcal{J}_\zeta^\tau\, \mathrm{E}_\theta\, T| - |a_\theta| && (2.3.83)\\
&\geq |A_\theta \mathcal{J}_\zeta^\tau T| - \|A_\theta\|_{\mathrm{op}} \|\mathcal{J}_\zeta\|_{\mathrm{op}} |\mathrm{E}_\theta\, T| - |a_\theta| && (2.3.84)\\
&\geq \|(A_\theta \mathcal{J}_\zeta^\tau)^{-1}\|_{\mathrm{op}}^{-1} |T| - \|A_\theta\|_{\mathrm{op}} \|\mathcal{J}_\zeta\|_{\mathrm{op}} |\mathrm{E}_\theta\, T| - |a_\theta| && (2.3.85)
\end{aligned}$$

That is,

$$|T| \leq \|(A_\theta \mathcal{J}_\zeta^\tau)^{-1}\|_{\mathrm{op}} \big[b_\theta + \|A_\theta\|_{\mathrm{op}} \|\mathcal{J}_\zeta\|_{\mathrm{op}} |\operatorname{E}_\theta T| + |a_\theta|\big] < \infty \tag{2.3.86}$$

on $\{|A_\theta \Lambda_\theta - a_\theta| \leq b_\theta\}$. Since A_θ, a_θ, b_θ, $\mathcal{J}_\zeta$ and $\operatorname{E}_\theta T$ are continuous in $\theta \in \Theta$, we also have

$$|T| \leq \|(A_{\theta_n} \mathcal{J}_{\zeta_n}^\tau)^{-1}\|_{\mathrm{op}} \big[b_{\theta_n} + \|A_{\theta_n}\|_{\mathrm{op}} \|\mathcal{J}_{\zeta_n}\|_{\mathrm{op}} |\operatorname{E}_{\theta_n} T| + |a_{\theta_n}|\big] < \infty \tag{2.3.87}$$

is bounded uniformly in $n \in \mathbb{N}$ on $\{|A_{\theta_n} \Lambda_{\theta_n} - a_{\theta_n}| \leq b_{\theta_n}\}$. Consequentially, T is bounded uniformly in $n \in \mathbb{N}$ on $\mathcal{D}$ $(= \mathcal{D}_n)$. Therefore,

$$\sup_{\mathcal{D}} |\Lambda_{\theta_n} - \Lambda_\theta| \;\leq\; \|\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau\|_{\mathrm{op}} \sup_{\mathcal{D}} |T| + |\mathcal{J}_{\zeta_n}^\tau \operatorname{E}_{\theta_n} T - \mathcal{J}_\zeta^\tau \operatorname{E}_\theta T| \tag{2.3.88}$$

$$\longrightarrow 0 \tag{2.3.89}$$

as $n \to \infty$ by the continuous differentiability of ζ and β; i.e., condition (2.3.10) holds. Moreover, we obtain

$$\begin{aligned} \sup_{{}^c\mathcal{D}} \frac{|\Lambda_{\theta_n} - \Lambda_\theta|}{|A_\theta \Lambda_\theta - a_\theta|} &\leq \sup_{{}^c\mathcal{D}} \frac{|(\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau)(T - \operatorname{E}_\theta T) + \mathcal{J}_{\zeta_n}^\tau \operatorname{E}_\theta T - \mathcal{J}_{\zeta_n}^\tau \operatorname{E}_{\theta_n} T|}{|A_\theta \Lambda_\theta - a_\theta|} \qquad (2.3.90) \\ &\leq \sup_{{}^c\mathcal{D}} \frac{|(\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau)(T - \operatorname{E}_\theta T)|}{|A_\theta \mathcal{J}_\zeta^\tau (T - \operatorname{E}_\theta T) - a_\theta|} \\ &\quad + \|\mathcal{J}_{\zeta_n}\|_{\mathrm{op}} \big|\operatorname{E}_{\theta_n} T - \operatorname{E}_\theta T\big| \sup_{{}^c\mathcal{D}} \frac{1}{|A_\theta \Lambda_\theta - a_\theta|} \qquad (2.3.91) \end{aligned}$$

where the second summand on the RHS tends to 0 as $n \to \infty$ since β is continuous differentiable in $\theta \in \Theta$ which implies the continuity of $\operatorname{E}_\theta T$ in θ, $\|\mathcal{J}_{\zeta_n}\|_{\mathrm{op}}$ is bounded uniformly in $n \in \mathbb{N}$ and $|A_\theta \Lambda_\theta - a_\theta| > b_\theta > 0$ on ${}^c\mathcal{D}$. Furthermore,

$$\begin{aligned} \sup_{{}^c\mathcal{D}} \frac{|(\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau)(T - \operatorname{E}_\theta T)|}{|A_\theta \mathcal{J}_\zeta^\tau (T - \operatorname{E}_\theta T) - a_\theta|} &\leq \|\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau\|_{\mathrm{op}} \sup_{{}^c\mathcal{D}} \frac{|T|}{|A_\theta \mathcal{J}_\zeta^\tau (T - \operatorname{E}_\theta T) - a_\theta|} \\ &\quad + \|\mathcal{J}_{\zeta_n}^\tau - \mathcal{J}_\zeta^\tau\|_{\mathrm{op}} |\operatorname{E}_\theta T| \sup_{{}^c\mathcal{D}} \frac{1}{|A_\theta \Lambda_\theta - a_\theta|} \qquad (2.3.92) \end{aligned}$$

where the second summand on the RHS tends to 0 as $n \to \infty$ since ζ is continuous differentiable in $\theta \in \Theta$, $\operatorname{E}_\theta T$ is bounded and $|A_\theta \Lambda_\theta - a_\theta| > b_\theta > 0$ on ${}^c\mathcal{D}$. In case T is bounded on ${}^c\mathcal{D}$, the first summand on the RHS tends to 0 as $n \to \infty$. Now, assume T is unbounded on ${}^c\mathcal{D}$ where

$$|A_\theta \mathcal{J}_\zeta^\tau (T - \operatorname{E}_\theta T) - a_\theta| \geq |A_\theta \mathcal{J}_\zeta^\tau T| - |A_\theta \mathcal{J}_\zeta^\tau \operatorname{E}_\theta T + a_\theta| \tag{2.3.93}$$

That is, we can choose some sufficiently large $M \in (0, \infty)$ such that

$$|A_\theta \mathcal{J}_\zeta^\tau T| - |A_\theta \mathcal{J}_\zeta^\tau \operatorname{E}_\theta T + a_\theta| \geq \frac{1}{2} |A_\theta \mathcal{J}_\zeta^\tau T| \tag{2.3.94}$$

for all $x \in {}^c\mathcal{D}$ such that $|T(x)| > M$. On $\mathcal{E} := {}^c\mathcal{D} \cap \big\{x \in \Omega \,\big|\, |T(x)| \leq M\big\}$ the first summand on the RHS of (2.3.92) tends to 0 as $n \to \infty$. In addition,

$$\sup_{u \in \mathbb{R}^k, u \neq 0} \frac{|u|}{|A_\theta \mathcal{J}_\zeta^\tau u|} \leq \|(A_\theta \mathcal{J}_\zeta^\tau)^{-1}\|_{\mathrm{op}} \tag{2.3.95}$$

That is, the first summand on the RHS of (2.3.92) tends to 0 as $n \to \infty$ on ${}^c\mathcal{D} \cap \{x \in \Omega \,|\, |T(x)| > M\}$, hence (2.3.11) holds.

$\boldsymbol{* = v, k = 1}$**:** The proof of condition (2.3.13) is analogous to the proof of condition (2.3.10). On ${}^c\mathcal{D}$ it holds,

$$\sup_{{}^c\mathcal{D}} \frac{|\Lambda_{\theta_n} - \Lambda_\theta|}{|\Lambda_\theta|} \le \sup_{{}^c\mathcal{D}} \frac{|(\mathcal{J}_{\zeta_n} - \mathcal{J}_\zeta)(T - \mathrm{E}_\theta\, T) + \mathcal{J}_{\zeta_n}\, \mathrm{E}_\theta\, T - \mathcal{J}_{\zeta_n}\, \mathrm{E}_{\theta_n}\, T|}{|\Lambda_\theta|} \tag{2.3.96}$$

$$\le \frac{|\mathcal{J}_{\zeta_n} - \mathcal{J}_\zeta|}{|\mathcal{J}_\zeta|} + |\mathcal{J}_{\zeta_n}|\big|\mathrm{E}_{\theta_n}\, T - \mathrm{E}_\theta\, T\big| \sup_{{}^c\mathcal{D}} \frac{1}{|\Lambda_\theta|} \tag{2.3.97}$$

which tends to 0 as $n \to \infty$ by the continuous differentiability of ζ and β, the uniform boundedness of $\mathcal{J}_{\zeta_n}$ and the fact that $|\Lambda_\theta|$ is bounded away from 0 on ${}^c\mathcal{D}$; i.e., condition (2.3.14) holds. ////

Remark 2.3.7 (a) The L_2 differentiability, respectively differentiability in quadratic mean of exponential families under the same assumptions as in part (a) of the previous lemma is mentioned in Example 7.7 of van der Vaart (1998).

(b) Strictly speaking, part (b) of the preceding lemma can be formulated to hold for L_2 differentiable parametric families with (2.1.102) and L_2 derivatives of the form

$$\Lambda_\theta(x) = C_\theta T(x) + d_\theta \tag{2.3.98}$$

where $C_\theta \in \mathbb{R}^{k\times k}$ and $d_\theta \in \mathbb{R}^k$ are continuous in $\theta \in \Theta$, C_θ is regular and $T\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^k, \mathbb{B}^k)$. However, in most cases this setup corresponds to an exponential family of full rank. ////

2.3.4 Median and Median Absolute Deviation

To apply the one-step construction, one needs some initial estimator which is strict and $\sqrt{n}$ consistent on some neighborhood system $\mathcal{U}(\theta)$ about P_θ; confer Subsection 2.3.1. The Kolmogorov(–Smirnov) minimum distance estimator, which is implemented in our `R` package `ROptEst`, is $\sqrt{n}$ consistent on neighborhoods $\mathcal{U}_\kappa(\theta)$ of Kolmogorov type (cf. Theorem 6.3.7 of Rieder (1994)) where

$$U_c(\theta, r) \subset U_v(\theta, r) \subset U_\kappa(\theta, r) \tag{2.3.99}$$

for radius $r \in (0, \infty)$; confer also Section 1.2. In robust literature the simpler median and median absolute deviation (MAD) are very often proposed as appropriate initial estimators. However, there seems to be no reference for their $\sqrt{n}$ consistency on some neighborhood system $\mathcal{U}(\theta)$. To fill this gap, we give proofs that the median and the MAD are $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta)$.

For this purpose we work with the following definition of the sample median; confer Example 2.1 in Section 3.2 of Huber (1981).

Definition 2.3.8 *Given real-valued observations $x_1, \ldots, x_n$ the sample median $M_n = M_n(x_1, \ldots, x_n)$ is any solution to*

$$\sum_{i=1}^n \psi(x_i - M_n - \varepsilon) \le 0 \le \sum_{i=1}^n \psi(x_i - M_n - \varepsilon) \qquad \forall\, \varepsilon > 0 \tag{2.3.100}$$

where $\psi(x) = \operatorname{sign}(x)$.

Let F_θ be the cumulative distribution function to P_θ. Then, the minimal median of P_θ is defined as

$$m_\theta := \inf\{x \in \mathbb{R} \,|\, F_\theta(x) \geq 0.5\} \tag{2.3.101}$$

Note, that P_θ does not necessarily belong to a location family in the following theorem.

Theorem 2.3.9 *Assume* F_θ *is continuous on some neighborhood of* m_θ *and has derivative* $f_\theta(m_\theta) > 0$ *at* m_θ. *Then, for all* $r \in (0, \infty)$

$$\lim_{T\to\infty} \limsup_{n\to\infty} \sup \left\{ Q_n^{(n)}(\sqrt{n}|M_n - m_\theta| > T) \,\middle|\, Q_{n,i} \in U_\kappa(\theta, r/\sqrt{n}\,) \right\} = 0 \tag{2.3.102}$$

where $Q_n^{(n)} = Q_{n,1} \otimes \cdots \otimes Q_{n,n}$.

Proof To simplify notation, we drop the fixed parameter $\theta \in \Theta$. Since F is differentiable in m with derivative $f(m) > 0$, $\operatorname{med}_P(X) = m$ is unique; confer Remark 2.3.10 (a). By translation equivariance of the median, we may assume $m = 0$ without loss of generality. Then, it holds for some given $t \in (0, \infty)$

$$\left\{\sqrt{n}\, M_n \leq -2t\right\} \subset \left\{ \sum_{i=1}^n \psi(x_i - M_n - t/\sqrt{n}) \leq 0 \right\} \tag{2.3.103}$$

That is,

$$\left\{\sqrt{n}\, M_n \leq -2t\right\} \subset \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(x_i + t/\sqrt{n}) \leq 0 \right\} \tag{2.3.104}$$

Similarly, we obtain

$$\left\{\sqrt{n}\, M_n \geq 2t\right\} \subset \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(x_i - t/\sqrt{n}) \geq 0 \right\} \tag{2.3.105}$$

Now, we first show

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{n,i} \right)(Q_n^{(n)}) \xrightarrow{\mathrm{w}} \mathcal{N}(0,1) \tag{2.3.106}$$

for all $r \in (0, \infty)$ and all $Q_{n,i} \in \mathcal{M}_1(\mathbb{B})$ with $d_\kappa(Q_{n,i}, P) \leq r/\sqrt{n}$ where

$$Y_{n,i} := \psi(x_i + t/\sqrt{n}\,) - \int \psi(x + t/\sqrt{n}\,)\, dQ_{n,i}(x) \tag{2.3.107}$$

and $Q_n^{(n)} = \bigotimes_{i=1}^n Q_{n,i}$. This can be done by verifying Lindeberg's condition which in this setup reads

$$\lim_{n\to\infty} \frac{1}{n s_n^2} \sum_{i=1}^n \int\limits_{\{|Y_{n,i}| > \delta\sqrt{n}\, s_n\}} Y_{n,i}\, dQ_{n,i} = 0 \qquad \forall\, \delta > 0 \tag{2.3.108}$$

where

$$s_n^2 := \frac{1}{n} \sum_{i=1}^n \int Y_{n,i}^2\, dQ_{n,i} \tag{2.3.109}$$

Since $|\psi| \le 1$, we get $|\int \psi \, dQ_{n,i}| \le 1$. Hence, it suffices to show

$$s_n^2 \longrightarrow \int \psi^2 \, dP > 0 \qquad \text{as } n \to \infty \tag{2.3.110}$$

which is implied by

$$\Big| \int \psi(x + t/\sqrt{n}\,) \, dQ_{n,i}(x) - \int \psi(x) \, dP(x) \Big| \longrightarrow 0 \tag{2.3.111}$$

and

$$\Big| \int \psi(x + t/\sqrt{n}\,)^2 \, dQ_{n,i}(x) - \int \psi(x)^2 \, dP(x) \Big| \longrightarrow 0 \tag{2.3.112}$$

as $n \to \infty$ where $\int \psi \, dP = 0$. By the triangular inequality we get

$$\begin{aligned} &\Big| \int \psi(x + t/\sqrt{n}\,) \, dQ_{n,i}(x) - \int \psi(x) \, dP(x) \Big| \\ &\qquad \le \Big| \int \psi(x + t/\sqrt{n}\,) \, dQ_{n,i}(x) - \int \psi(x + t/\sqrt{n}\,) \, dP(x) \Big| \\ &\qquad\quad + \Big| \int \psi(x + t/\sqrt{n}\,) \, dP(x) - \int \psi(x) \, dP(x) \Big| \end{aligned} \tag{2.3.113}$$

Indeed, the first summand on the RHS tends to 0, as

$$\Big| \int \psi(x + t/\sqrt{n}\,) \big[Q_{n,i} - P \big](dx) \Big|$$

$$= \Big| \int\limits_{(-t/\sqrt{n}\,,\infty)} \big[Q_{n,i} - P \big](dx) - \int\limits_{(-\infty,-t/\sqrt{n}\,)} \big[Q_{n,i} - P \big](dx) \Big| \tag{2.3.114}$$

$$= \Big| - G_{n,i}(-t/\sqrt{n}\,) + F(-t/\sqrt{n}\,) - G_{n,i}(-t/\sqrt{n} \, - 0) + F(-t/\sqrt{n} \, - 0) \Big| \tag{2.3.115}$$

$$\le 2 \frac{r}{\sqrt{n}} \tag{2.3.116}$$

where $G_{n,i}$ are the cumulative distribution functions to $Q_{n,i}$. Moreover,

$$\begin{aligned} &\Big| \int \big[\psi(x + t/\sqrt{n}\,) - \psi(x) \big] dP(x) \Big| \\ &\qquad = \Big| - F(-t/\sqrt{n}\,) - F(-t/\sqrt{n} \, - 0) + 2F(0) \Big| \longrightarrow 0 \end{aligned} \tag{2.3.117}$$

as $n \to \infty$ since F is continuous at 0. Next, we have

$$\Big| \int \psi(x + t/\sqrt{n}\,)^2 \big[Q_{n,i} - P \big](dx) \Big| = \Big| - \int\limits_{-t/\sqrt{n}} \big[Q_{n,i} - P \big](dx) \Big| \tag{2.3.118}$$

$$= \Big| P(\{-t/\sqrt{n}\,\}) - Q_{n,i}(\{-t/\sqrt{n}\,\}) \Big| \tag{2.3.119}$$

$$\le 2 \frac{r}{\sqrt{n}} \longrightarrow 0 \tag{2.3.120}$$

as $n \to \infty$ and

$$\left| \int \left[\psi(x + t/\sqrt{n}\,)^2 - \psi(x)^2\right] dP(x) \right| = \left| -P(\{-t/\sqrt{n}\,\}) + P(\{0\}) \right| \longrightarrow 0 \quad (2.3.121)$$

as $n \to \infty$ since F is continuous at 0. The previous two convergences entail (2.3.110) where

$$\int \psi(x)^2 \, dP(x) = 1 - P(\{0\}) = 1 \quad (2.3.122)$$

by the continuity of F at 0; i.e., (2.3.106) holds. Analogously, we may show

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{n,i} \right) (Q_n^{(n)}) \xrightarrow{\mathrm{w}} \mathcal{N}(0,1) \quad (2.3.123)$$

for all $r \in (0, \infty)$ and all $Q_{n,i} \in \mathcal{M}_1(\mathbb{B})$ with $d_\kappa(Q_{n,i}, P) \le r/\sqrt{n}$ where

$$Z_{n,i} := \psi(x_i - t/\sqrt{n}\,) - \int \psi(x - t/\sqrt{n}\,) \, dQ_{n,i}(x) \quad (2.3.124)$$

and $Q_n^{(n)} = \bigotimes_{i=1}^{n} Q_{n,i}$. Thus, by (2.3.104) and (2.3.105)

$$\begin{aligned} &Q_n^{(n)}\big(\{\sqrt{n}\, M_n \le -2t\}\big) \\ &\quad \le Q_n^{(n)}\left(\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{n,i} \le -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \psi(x_i + t/\sqrt{n}\,) \, dQ_{n,i}(x) \right\} \right) \end{aligned} \quad (2.3.125)$$

and

$$\begin{aligned} &Q_n^{(n)}\big(\{\sqrt{n}\, M_n \ge 2t\}\big) \\ &\quad \le Q_n^{(n)}\left(\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{n,i} \ge -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \psi(x_i - t/\sqrt{n}\,) \, dQ_{n,i}(x) \right\} \right) \end{aligned} \quad (2.3.126)$$

Analogously to (2.3.116) we obtain, uniformly in $i = 1, \ldots, n$,

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \psi(x \pm t/\sqrt{n}\,)[Q_{n,i} - P](dx) \right| \le 2 \frac{r}{\sqrt{n}} \sqrt{n} \; = 2r \quad (2.3.127)$$

Moreover,

$$\sqrt{n} \int \psi(x \pm t/\sqrt{n}\,) \, dP(x) = \sqrt{n} \left[P(\{x > \mp t/\sqrt{n}\,\}) - P(\{x < \mp t/\sqrt{n}\,\}) \right] \quad (2.3.128)$$

$$= \sqrt{n} \left[1 - F(\mp t/\sqrt{n}\,) - P(\{x < \mp t/\sqrt{n}\,\}) \right] \quad (2.3.129)$$

Since F is continuous on some neighborhood of 0, we get for sufficiently large n

$$P(\{x = \mp t/\sqrt{n}\,\}) = 0 \quad (2.3.130)$$

Consequently, it holds for sufficiently large n

$$\sqrt{n} \int \psi(x \pm t/\sqrt{n}\,) \, dP(x) = \sqrt{n} \left[1 - 2F(\mp t/\sqrt{n}\,) \right] \quad (2.3.131)$$

$$= 2\sqrt{n} \left[F(0) - F(\mp t/\sqrt{n}\,) \right] \quad (2.3.132)$$

where

$$2\sqrt{n}\,\big[F(0)-F(\mp t/\sqrt{n}\,)\big]\longrightarrow \pm 2tf(0)\qquad \text{as } n\to\infty \tag{2.3.133}$$

since F is differentiable at 0 with derivative $f(0)$. Hence,

$$\begin{aligned} Q_n^{(n)}\big(\{\sqrt{n}\,M_n\le -2t\}\big) &\le Q_n^{(n)}\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_{n,i}\le 2r-2tf(0)+\mathrm{o}(n^0)\right\}\right) &(2.3.134)\\ &\longrightarrow \Phi\big(2r-2tf(0)\big) &(2.3.135)\end{aligned}$$

as $n\to\infty$ where Φ is the cumulative distribution function of $\mathcal{N}(0,1)$. Analogously, we obtain

$$\begin{aligned} Q_n^{(n)}\big(\{\sqrt{n}\,M_n\ge 2t\}\big) &\le Q_n^{(n)}\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{n,i}\ge -2r+2tf(0)+\mathrm{o}(n^0)\right\}\right) &(2.3.136)\\ &\longrightarrow 1-\Phi\big(-2r+2tf(0)\big)=\Phi\big(2r-2tf(0)\big) &(2.3.137)\end{aligned}$$

Now, given some $\delta>0$ we may choose $T=T(\delta)\in(0,\infty)$ such large that

$$\Phi\big(2r-2Tf(0)\big)\le \delta/2 \tag{2.3.138}$$

Thus, we get, uniformly in $i=1,\dots,n$,

$$\limsup_{n\to\infty} Q_n^{(n)}\big(\{\sqrt{n}\,M_n\le -2T\}\big)\le \delta/2 \tag{2.3.139}$$

and

$$\limsup_{n\to\infty} Q_n^{(n)}\big(\{\sqrt{n}\,M_n\ge 2T\}\big)\le \delta/2 \tag{2.3.140}$$

Consequently, we have, uniformly in $i=1,\dots,n$,

$$\limsup_{n\to\infty} Q_n^{(n)}\big(\{\sqrt{n}\,|M_n|\ge 2T\}\big)\le \delta \tag{2.3.141}$$

i.e., (2.3.102) holds. ////

Remark 2.3.10 (a) We consider F_θ as given in the previous theorem. It holds

$$F_\theta(x)=0.5+(x-m_\theta)f_\theta(m_\theta)+(x-m_\theta)r(x) \tag{2.3.142}$$

and we may choose $m'<m_\theta<m''$ such that $2|r(x)|<f_\theta(m_\theta)$ for all $x\in[m',m'']$. Hence,

$$F_\theta(x)-0.5=(x-m_\theta)\big(f_\theta(m_\theta)+r(x)\big)>0 \tag{2.3.143}$$

for all $x\in(m_\theta,m'']$ and

$$F_\theta(x)-0.5=(x-m_\theta)\big(f_\theta(m_\theta)+r(x)\big)<0 \tag{2.3.144}$$

for all $x\in[m'm_\theta)$; i.e., $\mathrm{med}_{P_\theta}(X)=m_\theta$ is unique.

(b) The setup of Theorem 2.3.9 corresponds to the setup used in Section 3.2 of Huber (1981) where he derives the asymptotic normality of M estimates in the ideal model. It is also identical to the setup used in Corollary 1.5.4 of Rieder (1994) where the asymptotic linearity and normality of α-quantiles ($\alpha \in (0,1)$) in the ideal model, using compact differentiability, is shown. As a special case we may read off the asymptotic normality of the median in the ideal model which is

$$\sqrt{n}\,(M_n - m_\theta)(P_\theta^n) \xrightarrow{w} \mathcal{N}\left(0, \frac{1}{4f_\theta(m_\theta)^2}\right) \tag{2.3.145}$$

////

In scale models the median absolute deviation (MAD)

$$\mathrm{MAD}_n\,(x_1,\ldots,x_n) = \mathrm{med}\left\{|x_i - M_n(x_1,\ldots,x_n)| \,\middle|\, i = 1,\ldots,n\right\} \tag{2.3.146}$$

is frequently used as starting estimator. We denote the minimal median of $|X-m_\theta|$ under P_θ by

$$a_\theta := \inf\{x \in \mathbb{R}_+ \mid F_\theta(m_\theta + x) - F_\theta(m_\theta - x - 0) \geq 0.5\} \tag{2.3.147}$$

To prove $\sqrt{n}$ consistency on $\mathcal{U}_\kappa(\theta)$ of the MAD, we additionally use the following lemma.

Lemma 2.3.11 *Let $Q \in \mathcal{M}_1(\mathbb{B})$. Then*

$$\left|\mathrm{med}_Q\,|X-\mathrm{med}_Q(X)| - a_\theta\right| \leq \left|\mathrm{med}_Q\,|X-m_\theta| - a_\theta\right| + \left|\mathrm{med}_Q(X) - m_\theta\right| \tag{2.3.148}$$

Proof By the triangular inequality we get

$$|X(\omega) - \mathrm{med}_Q(X)| \leq |X(\omega) - m_\theta| + |\mathrm{med}_Q(X) - m_\theta| \qquad \forall\,\omega \in \mathbb{R} \tag{2.3.149}$$

That is,

$$\begin{aligned} 0.5 &\leq Q\big(|X - \mathrm{med}_Q(x)| \geq \mathrm{med}_Q\,|X - \mathrm{med}_Q(x)|\big) && (2.3.150)\\ &\leq Q\big(|X - m_\theta| + |\mathrm{med}_Q(X) - m_\theta| \geq \mathrm{med}_Q\,|X - \mathrm{med}_Q(x)|\big) && (2.3.151)\end{aligned}$$

Hence,

$$\mathrm{med}_Q\,|X - \mathrm{med}_Q(x)| \leq \mathrm{med}_Q\,|X - m_\theta| + |\mathrm{med}_Q(X) - m_\theta| \tag{2.3.152}$$

and therefore

$$\mathrm{med}_Q\,|X - \mathrm{med}_Q(x)| - a_\theta \leq \mathrm{med}_Q\,|X - m_\theta| - a_\theta + |\mathrm{med}_Q(X) - m_\theta| \tag{2.3.153}$$

Similarly, we obtain

$$\mathrm{med}_Q\,|X - \mathrm{med}_Q(x)| - a_\theta \geq \mathrm{med}_Q\,|X - m_\theta| - a_\theta - |\mathrm{med}_Q(X) - m_\theta| \tag{2.3.154}$$

Consequently, (2.3.148) holds. ////

Using the previous lemma and Theorem 2.3.9 we now verify the $\sqrt{n}$ consistency on $\mathcal{U}_\kappa(\theta)$ of the MAD.

Proposition 2.3.12 *Assume F_θ is differentiable in m_θ and in $m_\theta + a_\theta$, $m_\theta - a_\theta$ such that*

$$f_\theta(m_\theta) > 0 \qquad \text{and} \qquad f_\theta(m_\theta + a_\theta) - f(m_\theta - a_\theta) > 0 \tag{2.3.155}$$

and suppose F is continuous on some neighborhood of $\{m_\theta, m_\theta + a_\theta, m_\theta - a_\theta\}$. Then, for all $r \in (0, \infty)$

$$\lim_{T\to\infty} \limsup_{n\to\infty} \sup \left\{ Q_n^{(n)}\left(\sqrt{n}\left|\mathrm{MAD}_n - a_\theta\right| > T\right) \,\middle|\, Q_{n,i} \in U_\kappa(\theta, r/\sqrt{n}\,) \right\} = 0 \tag{2.3.156}$$

where $Q_n^{(n)} = Q_{n,1} \otimes \cdots \otimes Q_{n,n}$.

Proof To simplify notation, we drop the fixed parameter $\theta \in \Theta$. Since F is differentiable in m with derivative $f(m) > 0$, $\mathrm{med}_P(X) = m$ is unique; confer Remark 2.3.10 (a). Now, let $\hat{Q}_n$ be the empirical distribution to $x_1, \ldots, x_n$. Then,

$$\mathrm{MAD}_n(x_1, \ldots, x_n) = \mathrm{med}_{\hat{Q}_n} |X - M_n| \tag{2.3.157}$$

By Lemma 2.3.11 we obtain

$$\left|\mathrm{MAD}_n - \mathrm{med}_P |X - m|\right| \le \left|\mathrm{med}_{\hat{Q}_n} |X - m| - \mathrm{med}_P |X - m|\right| + \left|M_n - m\right| \tag{2.3.158}$$

where M_n is $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta, r/\sqrt{n}\,)$ for all $r \in (0, \infty)$ by Theorem 2.3.9. Moreover, for $t \in (0, \infty)$ we have

$$P(|X - m| \le t) = P(m - t \le X \le m + t) = F(m + t) - F(m - t - 0) \tag{2.3.159}$$

That is,

$$\tilde{F}_m(t) := \begin{cases} 0 & \text{if } t \le 0 \\ F(m + t) - F(m - t - 0) & \text{if } t > 0 \end{cases} \tag{2.3.160}$$

is the cumulative distribution function to $\mathcal{L}_P(|X - m|)$. By assumption $\tilde{F}_m$ is differentiable at a with derivative $f(m+a) - f(m-a) > 0$; i.e., $\mathrm{med}_P |X - m| = a$ is unique. In addition, it follows by the assumptions on F that $\tilde{F}_m$ is continuous on some neighborhood of a. Therefore, we get by Theorem 2.3.9, $\mathrm{med}_{\hat{Q}_n} |X - m|$ is $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\mathcal{L}_P(|X - m|), r/\sqrt{n}\,)$ for all $r \in (0, \infty)$. Now, assume $X \sim Q$ where $Q \in \mathcal{U}_\kappa(\theta, r/\sqrt{n}\,)$. Then, $|X - m|$ under Q has cumulative distribution function

$$\tilde{G}_m(t) := \begin{cases} 0 & \text{if } t < 0 \\ G(m + t) - G(m - t - 0) & \text{if } t \ge 0 \end{cases} \tag{2.3.161}$$

where G is the cumulative distribution function to Q and we obtain

$$\begin{aligned} &\sup_{t\in\mathbb{R}} |\tilde{G}_m(t) - \tilde{F}_m(t)| \\ &\quad = \sup_{t\in\mathbb{R}_+} \left|G(m + t) - G(m - t - 0) - F(m + t) + F(m - t - 0)\right| && (2.3.162) \\ &\quad \le 2\frac{r}{\sqrt{n}} && (2.3.163) \end{aligned}$$

That is,

$$\mathcal{L}_Q(|X-m|) \in \mathcal{U}_\kappa(\mathcal{L}_P(|X-m|), 2r/\sqrt{n}\,) \tag{2.3.164}$$

for all $Q \in \mathcal{U}_\kappa(\theta, r/\sqrt{n}\,)$. Consequently, the $\sqrt{n}$ consistency of $\mathrm{med}_{\hat{Q}_n}|X-m|$ on $\mathcal{U}_\kappa(\mathcal{L}_P(|X-m|), 2r/\sqrt{n}\,)$ via (2.3.158) entails the $\sqrt{n}$ consistency of MAD_n on $\mathcal{U}_\kappa(\theta, r/\sqrt{n}\,)$; i.e., (2.3.156) holds. ////

Remark 2.3.13 If $\mathcal{L}_{P_\theta}(X-m_\theta)$ is symmetric, $m_\theta - a_\theta$ and $m_\theta + a_\theta$ correspond to the lower and upper quartile of P_θ, respectively. ////

2.4 Convergence of Robust Models

Now, let

$$\mathcal{P}_\nu = \{P_{\nu,\theta} \,|\, \theta \in \Theta_\nu\} \subset \mathcal{M}_1(\mathcal{A}_\nu) \qquad (\nu \in \mathbb{N}_0) \tag{2.4.1}$$

be a sequence of parametric families of probability measures on sample spaces $(\Omega_\nu, \mathcal{A}_\nu)$, whose parameter spaces Θ_ν are open subsets of $\mathbb{R}^k$. While working on various models, we found the following useful result, whose proof goes along the lines of the proof of Theorem 2.1.11.

Theorem 2.4.1 *Assume $\mathcal{P}_\nu$ is L_2 differentiable at $\theta_\nu \in \Theta_\nu$ $(\nu \in \mathbb{N}_0)$ with L_2 derivative $\Lambda_{\nu,\theta_\nu} \in L_2^k(P_{\nu,\theta_\nu})$ and Fisher information of full rank k,*

$$\mathcal{I}_{\nu,\theta_\nu} = \mathrm{E}_{\nu,\theta_\nu} \Lambda_{\nu,\theta_\nu} \Lambda_{\nu,\theta_\nu}^\tau \tag{2.4.2}$$

where

$$\mathcal{L}_{P_{\nu,\theta_\nu}}(\gamma_\nu^{-1} G_\nu \Lambda_{\nu,\theta_\nu}) \xrightarrow{\mathrm{w}} \mathcal{L}_{P_{0,\theta_0}}(\Lambda_{0,\theta_0}) \tag{2.4.3}$$

and

$$\gamma_\nu^{-2} \operatorname{tr} \mathcal{I}_{\nu,\theta_\nu} \longrightarrow \operatorname{tr} \mathcal{I}_{0,\theta_0} \tag{2.4.4}$$

as $\nu \to \infty$ for some standardizing sequence $(\gamma_\nu)_{\nu\in\mathbb{N}} \subset (0,\infty)$ and a sequence of orthogonal matrices $(G_\nu)_{\nu\in\mathbb{N}} \subset \mathbb{R}^{k\times k}$.

(a) $* = c$: *Let $D = \mathbb{I}_k$ and $\tilde{\eta}_{\nu,\theta_\nu}$ be the solutions to the MSE problem (1.3.5) for radius $r \in (0,\infty)$ given in Theorem 1.3.11 (b). Denote the Lagrange multipliers contained in the solution $\tilde{\eta}_{\nu,\theta_\nu}$ by A_{ν,θ_ν}, a_{ν,θ_ν} and b_{ν,θ_ν}. Then,*

$$\lim_{\nu\to\infty} \gamma_\nu^2 \operatorname{tr} A_{\nu,\theta_\nu} = \operatorname{tr} A_{0,\theta_0} \qquad \lim_{\nu\to\infty} \gamma_\nu b_{\nu,\theta_\nu} = b_{0,\theta_0} \tag{2.4.5}$$

In case A_{0,θ_0} and a_{0,θ_0} are unique, then also

$$\lim_{\nu\to\infty} \gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu = A_{0,\theta_0} \qquad \text{and} \qquad \lim_{\nu\to\infty} \gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} = a_{0,\theta_0} \tag{2.4.6}$$

(b) $* = v$, $k = 1$: *Let $D = 1$ and $\tilde{\eta}_{\nu,\theta_\nu}$ be the solutions to the MSE problem (1.3.5) with radius $r \in (0,\infty)$ given in Theorem 1.3.11 (c) for the parameter $\theta_\nu \in \Theta_\nu$. Denote the Lagrange multipliers contained in the solution $\tilde{\eta}_{\nu,\theta_\nu}$ by A_{ν,θ_ν}, b_{ν,θ_ν} and c_{ν,θ_ν}. Then,*

$$\lim_{\nu\to\infty} \gamma_\nu^2 A_{\nu,\theta_\nu} = A_{0,\theta_0} \qquad \lim_{\nu\to\infty} \gamma_\nu b_{\nu,\theta_\nu} = b_{0,\theta_0} \tag{2.4.7}$$

and

$$\lim_{\nu\to\infty} \gamma_\nu c_{\nu,\theta_\nu} = \begin{cases} c_{0,\theta_0} & \text{if } G_\nu = 1 \\ -(c_{0,\theta_0} + b_{0,\theta_0}) & \text{if } G_\nu = -1 \end{cases} \tag{2.4.8}$$

(c) *Let* $D = \mathbb{I}_k$ *and* $\omega^{\min}_{*,\nu,\theta_\nu}$ *be the minimum bias for the parameter* $\theta_\nu \in \Theta_\nu$ *given in Theorems 1.3.7 (b) and 1.3.9 (b), respectively. Then,*

$$\lim_{\nu\to\infty} \gamma_\nu \omega^{\min}_{*,\nu,\theta_\nu} = \omega^{\min}_{*,0,\theta_0} \qquad (* = c, v) \tag{2.4.9}$$

Proof

(a) We fix some radius $r \in (0,\infty)$. By identifying P_{ν,θ_ν} with P_{θ_ν}, P_{0,θ_0} with P_θ, $\gamma_\nu^{-1} G_\nu \Lambda_{\nu,\theta_\nu}$ with Λ_{θ_ν}, Λ_{0,θ_0} with Λ_θ, $\mathcal{I}_{\nu,\theta_\nu}$ with $\mathcal{I}_{\theta_\nu}$ and $\mathcal{I}_{0,\theta_0}$ with $\mathcal{I}_\theta$ we are in the setup of the proof of Theorem 2.1.11 (a) which yields

$$\lim_{\nu\to\infty} \operatorname{tr} A_{G_\nu} = \operatorname{tr} A_{0,\theta_0} \qquad \lim_{\nu\to\infty} b_{G_\nu} = b_{0,\theta_0} \tag{2.4.10}$$

where A_{G_ν}, b_{G_ν}, and a_{G_ν} are the Lagrange multipliers contained in solution to the corresponding MSE problem (1.3.5). In case A_{0,θ_0} and a_{0,θ_0} are unique, then also

$$\lim_{\nu\to\infty} A_{G_\nu} = A_{0,\theta_0} \qquad \text{and} \qquad \lim_{\nu\to\infty} a_{G_\nu} = a_{0,\theta_0} \tag{2.4.11}$$

Since $G_\nu G_\nu^\tau = \mathbb{I}_k$ and

$$|A_{G_\nu} G_\nu \Lambda_{\nu,\theta_\nu} - a_{G_\nu}| = |G_\nu^\tau (A_{G_\nu} G_\nu \Lambda_{\nu,\theta_\nu} - a_{G_\nu})| \tag{2.4.12}$$

we obtain

$$\operatorname{tr} A_{G_\nu} = \gamma_\nu^2 \operatorname{tr} A_{\nu,\theta_\nu} \qquad b_{G_\nu} = \gamma_\nu b_{\nu,\theta_\nu} \tag{2.4.13}$$

and we can choose

$$A_{G_\nu} = \gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu \qquad a_{G_\nu} = \gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} \tag{2.4.14}$$

which proves part (a).

(b) We fix some radius $r \in (0,\infty)$. By the same identifications as at the beginning of the proof of part (a), we are again in the setup of the proof of Theorem 2.1.11 (a) which yields

$$\lim_{\nu\to\infty} A_{G_\nu} = A_{0,\theta_0} \qquad \lim_{\nu\to\infty} b_{G_\nu} = b_{0,\theta_0} \qquad \lim_{\nu\to\infty} c_{G_\nu} = c_{0,\theta_0} \tag{2.4.15}$$

where A_{G_ν}, b_{G_ν}, and c_{G_ν} are the Lagrange multipliers contained in solution to the corresponding MSE problem (1.3.5). Since we are in dimension $k = 1$, we can only have $G_\nu = 1$ or $G_\nu = -1$, respectively. In case $G_\nu = 1$, we at once get

$$A_{G_\nu} = \gamma_\nu^2 A_{\nu,\theta_\nu} \qquad b_{G_\nu} = \gamma_\nu b_{\nu,\theta_\nu} \qquad c_{G_\nu} = \gamma_\nu c_{\nu,\theta_\nu} \tag{2.4.16}$$

In contrast, in case $G_\nu = -1$, a look at

$$\mathrm{E}_{\nu,\theta_\nu} \left(A_{G_\nu} \gamma_\nu^{-1} (-\Lambda_{\nu,\theta_\nu}) - (c_{G_\nu} + b_{G_\nu}) \right)_+ = \mathrm{E}_{\nu,\theta_\nu} \left(c_{G_\nu} - A_{G_\nu} \gamma_\nu^{-1} (-\Lambda_{\nu,\theta_\nu}) \right)_+ \tag{2.4.17}$$

$$1 = \mathrm{E}_{\nu,\theta_\nu}\left[c_{G_\nu} \vee A_{G_\nu}\gamma_\nu^{-1}(-\Lambda_{\nu,\theta_\nu}) \wedge (c_{G_\nu} + b_{G_\nu})\right]\gamma_\nu^{-1}(-\Lambda_{\nu,\theta_\nu}) \quad (2.4.18)$$

$$= \mathrm{E}_{\nu,\theta_\nu}\left[-\gamma_\nu^{-1}(c_{G_\nu} + b_{G_\nu}) \vee A_{G_\nu}\gamma_\nu^{-2}\Lambda_{\nu,\theta_\nu} \wedge -\gamma_\nu^{-1}c_\nu\right]\Lambda_{\nu,\theta_\nu} \quad (2.4.19)$$

and

$$r^2 b_{G_\nu} = \mathrm{E}_{\nu,\theta_\nu}\left(A_{G_\nu}\gamma_\nu^{-1}(-\Lambda_{\nu,\theta_\nu}) - (c_{G_\nu} + b_{G_\nu})\right)_+ \quad (2.4.20)$$

shows, that we instead obtain

$$A_{G_\nu} = \gamma_\nu^2 A_{\nu,\theta_\nu} \qquad b_{G_\nu} = \gamma_\nu b_{\nu,\theta_\nu} \qquad -(c_{G_\nu} + b_{G_\nu}) = \gamma_\nu c_{\nu,\theta_\nu} \quad (2.4.21)$$

(c) By the same identifications as at the beginning of the proof of part (a), we are in the setup of the proof of Theorem 2.1.11 (b).

$\boldsymbol{* = c}$**:** We obtain

$$\max\left\{\frac{\operatorname{tr} A}{\mathrm{E}_{\nu,\theta_\nu}|\gamma_\nu^{-1}AG_\nu\Lambda_{\nu,\theta_\nu} - a|} \,\middle|\, a \in \mathbb{R}^k,\ A \in \mathbb{R}^{k\times k} \setminus \{0\}\right\} \longrightarrow \omega^{\min}_{c,0,\theta_0} \quad (2.4.22)$$

as $\nu \to \infty$ where $\operatorname{tr}(G_\nu^\tau A G_\nu) = \operatorname{tr}(G_\nu G_\nu^\tau A) = \operatorname{tr} A$ and

$$\mathrm{E}_{\nu,\theta_\nu}|\gamma_\nu^{-1}AG_\nu\Lambda_{\nu,\theta_\nu} - a| = \gamma_\nu^{-1}\,\mathrm{E}_{\nu,\theta_\nu}|G_\nu^\tau AG_\nu\Lambda_{\nu,\theta_\nu} - G_\nu^\tau a| \quad (2.4.23)$$

Hence,

$$\max\left\{\frac{\operatorname{tr} A}{\mathrm{E}_{\nu,\theta_\nu}|\gamma_\nu^{-1}AG_\nu\Lambda_{\nu,\theta_\nu} - a|} \,\middle|\, a \in \mathbb{R}^k,\ A \in \mathbb{R}^{k\times k} \setminus \{0\}\right\}$$

$$= \gamma_\nu \max\left\{\frac{\operatorname{tr} A}{\mathrm{E}_{\nu,\theta_\nu}|A\Lambda_{\nu,\theta_\nu} - a|} \,\middle|\, a \in \mathbb{R}^k,\ A \in \mathbb{R}^{k\times k} \setminus \{0\}\right\} \quad (2.4.24)$$

$$= \gamma_\nu \omega^{\min}_{c,\nu,\theta_\nu} \quad (2.4.25)$$

$\boldsymbol{* = v,\ k = 1}$**:** It holds

$$\left(\mathrm{E}_{\nu,\theta_\nu}(\gamma_\nu^{-1}G_\nu\Lambda_{\nu,\theta_\nu})_+\right)^{-1} \longrightarrow \omega^{\min}_{v,0,\theta_0} \quad (2.4.26)$$

as $\nu \to \infty$. Since $\mathrm{E}_{\nu,\theta_\nu} G_\nu\Lambda_\nu = 0$, we have $\mathrm{E}_{\nu,\theta_\nu}(G_\nu\Lambda_{\nu,\theta_\nu})_+ = \mathrm{E}_{\nu,\theta_\nu}(G_\nu\Lambda_{\nu,\theta_\nu})_-$. Hence,

$$\mathrm{E}_{\nu,\theta_\nu}|G_\nu\Lambda_{\nu,\theta_\nu}| = 2\,\mathrm{E}_{\nu,\theta_\nu}(G_\nu\Lambda_{\nu,\theta_\nu})_+ \quad (2.4.27)$$

In addition,

$$\mathrm{E}_{\nu,\theta_\nu}|G_\nu\Lambda_{\nu,\theta_\nu}| = \mathrm{E}_{\nu,\theta_\nu}|\Lambda_{\nu,\theta_\nu}| = 2\,\mathrm{E}_{\nu,\theta_\nu}(\Lambda_{\nu,\theta_\nu})_+ \quad (2.4.28)$$

i.e., $\mathrm{E}(G_\nu\Lambda_{\nu,\theta_\nu})_+ = \mathrm{E}(\Lambda_{\nu,\theta_\nu})_+$. Consequentially,

$$\left(\mathrm{E}_{\nu,\theta_\nu}(\gamma_\nu^{-1}G_\nu\Lambda_{\nu,\theta_\nu})_+\right)^{-1} = \gamma_\nu\left(\mathrm{E}_{\nu,\theta_\nu}(\Lambda_{\nu,\theta_\nu})_+\right)^{-1} = \gamma_\nu\omega^{\min}_{v,\nu,\theta_\nu} \quad (2.4.29)$$

////

Remark 2.4.2 (a) Theorem 2.4.1 may be regarded as a generalization of Theorem 2.1.11 by identifying $\mathcal{P}$ with $\mathcal{P}_0$ and choosing $\mathcal{P}_\nu \equiv \mathcal{P}$ for all $\nu \in \mathbb{N}$ where we have $\gamma_\nu \equiv 1$ and $G_\nu \equiv \mathbb{I}_k$.

(b) Theorem 2.4.1 entails

$$\lim_{\nu\to\infty} \gamma_\nu \omega_{*,\nu,\theta_\nu}(\tilde{\eta}_{*,\nu,\theta_\nu}) = \omega_{*,0,\theta_0}(\tilde{\eta}_{*,0,\theta_0}) \tag{2.4.30}$$

for all $r \in [0,\infty]$. Moreover, we get

$$\lim_{\nu\to\infty} \gamma_\nu^2 \, \mathrm{E}_{\nu,\theta_\nu} |\tilde{\eta}_{*,\nu,\theta_\nu}|^2 = \lim_{\nu\to\infty} \gamma_\nu^2 \operatorname{tr} A_{\nu,\theta_\nu} - r^2\gamma_\nu^2 b_{\nu,\theta_\nu}^2 \tag{2.4.31}$$

$$= \operatorname{tr} A_{0,\theta_0} - r^2 b_{0,\theta_0}^2 \tag{2.4.32}$$

$$= \mathrm{E}_{0,\theta_0} |\tilde{\eta}_{*,0,\theta_0}|^2 \tag{2.4.33}$$

for $r \in [0,\infty)$ and

$$\lim_{\nu\to\infty} \gamma_\nu^2 \, \mathrm{E}_{\nu,\theta_\nu} |\bar{\eta}_{*,\nu,\theta_\nu}|^2 = \mathrm{E}_{0,\theta_0} |\bar{\eta}_{*,0,\theta_0}|^2 \tag{2.4.34}$$

Hence,

$$\lim_{\nu\to\infty} \gamma_\nu^2 \operatorname{maxMSE}(\tilde{\eta}_{*,\nu,\theta_\nu,r}, s) = \operatorname{maxMSE}(\tilde{\eta}_{*,0,\theta_0,r}, s) \tag{2.4.35}$$

respectively

$$\lim_{\nu\to\infty} \operatorname{relMSE}(\tilde{\eta}_{*,\nu,\theta_\nu,r}, s) = \operatorname{relMSE}(\tilde{\eta}_{*,0,\theta_0,r}, s) \tag{2.4.36}$$

for all $r, s \in [0,\infty]$. That is, the relative MSE–inefficiencies converge without further standardizations.

(c) Theorem 2.4.1 might be interpreted as some kind of *convergence of robust models*. Extending Le Cam's notion of convergence of experiments (cf. Chapter 2 of Le Cam and Lo Yang (1990)), we consider neighborhoods about ideal models, where the ideal models converge in Le Cam's sense (cf. 6.5 Proposition 9 of Le Cam and Lo Yang (1990) or Corollary 9.5 of van der Vaart (1998), respectively). As we are only interested in the optimally robust estimators, however, in our notion of convergence of robust models, we do not consider arbitrary estimators. Convergence is formulated in terms of the maximum asymptotic MSE, respectively, in view of Ruckdeschel and Rieder (2004) (cf. also Remark 1.3.3), may be extended to hold for a whole class of (convex) risks, but not for arbitrary loss. Anyway, it seems necessary in many situations to restrict the class of loss functions to obtain comparable experiments; confer Lehmann (1988), pp 525.

(e) An application of the previous theorem for instance shows that in case of the binomial and the Poisson model there is a normal and Poisson approximation valid for the Lagrange multipliers contained in the corresponding optimally robust ICs. In addition, we numerically computed the "distance" between the exact and the approximated ICs in terms of the corresponding MSE–inefficiency. For more details we refer to Section 3.2 and Section 4.2, respectively.

(f) The convergence of robust models could be of special interest in situations where the optimally robust ICs are hard to compute. That is, if we find some

approximating model where the optimally robust ICs are easier to compute, the results obtained in the latter model may serve as approximations for the model of interest.

(g) Of course, Theorem 2.4.1 is also applicable if there is even equality in (2.4.3) and (2.4.4) for some constant $\gamma \in (0, \infty)$ and some orthogonal matrix $G \in \mathbb{R}^{k \times k}$. This is for instance true in case of the exponential scale and the Gumbel location model. It also holds if we compare the optimally robust ICs for different parameter values in case of the regression and scale model. For more details see Chapter 5 and Chapter 7, respectively.

(h) In view of Ruckdeschel (2005b) a generalization of Theorem 2.4.1 to arbitrary $D \in \mathbb{R}^{p \times k}$ with $\operatorname{rk} D = p$ seems to be in reach. Moreover, we conjecture that similar results may be obtained in case of regression-type models as introduced in Appendix A. Thus, it might be interesting to further investigate the convergence of robust models and, in particular, to take a closer look at the relationship between our notion and the convergence of experiments of Le Cam. However, this goes beyond the scope of this thesis. ////

Part II

Non-Standard Robust Models

The subject of this second part are well-known parametric models which in robust literature have rarely been considered. If standard robust models like location and scale are included, our distributional assumptions on the ideal model are non-standard.

Binomial and Poisson Model

Robust estimation in the binomial and Poisson models has received only little attention in robust literature. It was first mentioned in Section F.3 of Hampel (1968), who calculates the score function Λ_θ and applies his Lemma 5 to the binomial and Poisson models. His optimal ψ function $\tilde{\psi}_\theta$ in general minimizes the asymptotic variance $\mathrm{E}_\theta\,\psi_\theta^2/(\mathrm{E}_\theta\,\psi_\theta\Lambda_\theta)^2$ under the bound $b = c/\,\mathrm{E}_\theta\,\tilde{\psi}_\theta\Lambda_\theta$ on the gross error sensitivity $\sup|\psi_\theta/\,\mathrm{E}_\theta\,\psi_\theta\Lambda_\theta|$ for any $c \in (0,\infty)$. Hampel's solution is of the same form as our optimally robust influence curves in case of infinitesimal contamination neighborhoods, as specified in Subsection 1.3.3.

In his treatment of the binomial and Poisson models, as in general smooth parametric models, Hampel (1968) has no criterion for the choice of the sensitivity bound b. By considering the corresponding MSE problem, we obtain an additional equation to determine b in a unique and optimal way depending on the (starting) radius $r \in (0,\infty)$ of the infinitesimal neighborhoods; confer Subsection 1.3.4.

There are other papers on robust estimation in discrete models which, however, consider only particular aspects. Ruckstuhl and Welsh (2001) for instance propose a robust estimator which has a high breakdown point and at the same time a bounded influence curve in case of the Binomial model. Simpson et al. (1987) show asymptotic non-normality over neighborhoods of Hampel's optimal M estimators when the underlying distribution is discrete. The result, a limiting law which is pieced together by two normal distributions with different standard deviations, is analogous to the result given on p 78 of Huber (1964), respectively on p 51 of Huber (1981). Moreover, a similar result on the asymptotic non-normality of the trimmed mean has been proved by Stigler (1973). As a way out, Simpson et al. (1987) propose to replace $\tilde{\psi}_\theta$ by a smooth approximation to retain asymptotic normality.

In contrast, our optimally robust influence curves are solutions to certain optimization problems based on the MSE criterion. Asymptotic normality of our more general AL estimators on full $1/\sqrt{n}$ neighborhoods is obtained by the smoothness of the underlying parametric model and a suitable estimator construction. Most frequently in literature, robust estimators are constructed via the M principle. We prefer to construct the corresponding optimally robust estimators by means of the one-step method which, depending on a suitable initial estimator, is faster to compute and always yields a unique solution. For more details on one-step constructions we refer to Section 6.4 of Rieder (1994) and Section 2.3.

In Chapters 3 and 4 we in detail consider the binomial and Poisson model where we first briefly introduce the ideal models; confer Sections 3.1 and 4.1. In Subsubsections 3.2.1.1 and 3.2.2.1, respectively 4.2.1.1 and 4.2.2.1 we specify the MSE optimal ICs in case of contamination ($* = c$) as well as total variation neighborhoods ($* = v$) and give some numerical results for the lower case radius $\bar{r}$ introduced in Subsection 2.1.2.

Subsequently, we numerically investigate technical properties (continuity and uniqueness) of the Lagrange multipliers contained in the optimal solutions which are useful for: Determination of least favorable radii (cf. Section 2.2), one-step construction (cf. Section 2.3) and convergence of robust models (cf. Section 2.4); confer Subsubsections 3.2.1.2 and 3.2.2.2, respectively 4.2.1.2 and 4.2.2.2.

First, we study the dependence on the neighborhood radius r. The numerical results indicate that the standardizing constant A_r is smooth whereas the standardized bias b_r, the lower clipping bound c_r $(* = v)$ and the asymptotic variance $A_r - r^2 b_r^2$ may be non-differentiable at some values of r. In addition, we consider parameter values θ where $\mathrm{med}(\Lambda_\theta)$ is non-unique. As a consequence of Proposition 2.1.3, the optimal centering constant a_r is non-unique for $r \geq \bar{r}$ and those values of θ. More precisely, there is a whole interval of valid centering constants for $r \geq \bar{r}$.

Second, we treat continuity with respect to the parameter θ. The numerical results indicate that the standardizing constant A_r, the standardized bias b_r, the lower clipping bound c_r $(* = v)$ and the asymptotic variance $A_r - r^2 b_r^2$ are continuous but, not necessarily smooth functions in θ. Moreover, the centering constant a_r $(* = c)$ for radii $r \geq \bar{r}$ is even discontinuous at those values of θ for which $\mathrm{med}(\Lambda_\theta)$ is non-unique.

These numerical results confirm the continuity and uniqueness results derived in Subsections 2.1.4 and 2.1.5 and indicate that we cannot expect the Lagrange multipliers to be smooth functions neither in the radius r nor in the parameter θ, in general.

We also use the binomial and Poisson models to demonstrate the convergence of robust models derived in Section 2.4; confer Subsubsections 3.2.1.3 and 3.2.2.3, respectively Subsubsections 4.2.1.3, 4.2.2.3, 4.2.1.4 and 4.2.2.4.

For this purpose, we give a proof that the suitable standardized Lagrange multipliers in case of the binomial and Poisson models converge towards the corresponding Lagrange multipliers of one-dimensional normal location. Moreover, we show that the Lagrange multipliers in case of the Poisson model can be approximated by the corresponding Lagrange multipliers arising in the binomial model.

With these results on hand, we numerically computed the "distance" in terms of the MSE–inefficiency between the optimal IC and the corresponding approximation. In case of contamination neighborhoods these approximations work well for small radii $(r \leq 0.5)$. In case of total variation neighborhoods these approximations perform even better and we seem to get very good approximations independent of the considered neighborhood radius.

In Sections 3.3 and 4.3 we assume the (starting) radius of the infinitesimal neighborhoods is unknown. We give some numerical results for the least favorable radii and the corresponding MSE–inefficiencies in case of the binomial and Poisson models. In both models and all considered situations the efficiency loss stays below 30% and in most cases is even much smaller.

The construction problem in case of the binomial and Poisson models is solved in Sections 3.4 and 4.4. That is, we verify that we can construct the optimally robust estimator by means of one-step constructions by applying the results of Subsection 2.3.3. In particular, we investigate those parameter values for which

the centering constant a_r $(* = c)$ is non-unique for $r \geq \bar{r}$; i.e., we cannot apply Lemma 2.3.6 (b). As initial estimator we propose and also implemented the Kolmogorov(–Smirnov) minimum distance estimator.

The implementation of the binomial model by means of `S4` classes and methods (cf. Chambers (1998)) using `R` (cf. R Development Core Team (2005)) is in detail described in Section 3.5. Since the implementation of the Poisson model is very similar, we give only a very short description in Section 4.5. Both models are included in our `R` package `ROptEst` (cf. Appendix D.3) which is part of our `R` bundle `RobASt`.

To demonstrate the need of robust estimation in these two simple discrete models, we included some small simulation studies; confer Sections 3.6 and 4.6. The results indicate that the classically optimal estimator (mean) is too sensitive and already very small deviations from the ideal model may lead to a very high efficiency loss compared to the optimally robust estimator. In addition, the results of these studies point out that the radius–minimax estimator may be a good choice if the true neighborhood radius is unknown.

Exponential Scale and Gumbel Location

Hampel (1968) (cf. Section F.1) discusses robust estimation in case of the exponential model where this model arises as an important special case of the Gamma model. He proposes to use a trimmed mean and suggests that the trimmed mean has the same breakdown point as the commonly used Winsorized mean (cf. Feller (1971), Problem 17, p 41) but, in addition, has a smaller sensitivity.

Gather and Schultze (1999) consider the standardized median as robust estimator for the exponential scale model. They show (cf. Theorem 2.1, ibid.) that this estimator is most B-robust in sense of Hampel et al. (1986); i.e., has minimal gross error sensitivity. In addition, Gather and Schultze (1999) introduce two other robust estimators (RCS and Q estimators) which have been proposed by Rousseeuw and Croux (1993). All three estimators have the highest possible breakdown point which in this setup is 0.5. However, their bias curves and their asymptotic relative efficiencies are different.

As already mentioned above, our optimally robust influence curves are solutions to certain optimization problems and we obtain asymptotic normality of our more general AL estimators on full $1/\sqrt{n}$ neighborhoods by the smoothness of the underlying parametric model and a suitable estimator construction. Moreover, in case of the one-step construction global properties like breakdown can be delegated to the initial estimate. Aside from these (local and global) properties, the focus of Chapter 5 is rather the connection between location and scale models than the models themselves.

In Chapter 5 we show that certain scale and location models are connected via the transformations $\pm\log|\cdot|$ which is motivated by the treatment of the normal scale model in Section 5.6 of Huber (1981). This is for instance true in case of the exponential scale and the Gumbel location model.

We begin with a brief introduction of the one-dimensional scale and the one-dimensional location model; confer Subsections 5.1.1 and 5.1.2. Subsequently, we derive the mentioned connection (cf. Subsection 5.1.3) and show that this con-

nection entails a strong relationship between the Lagrange multipliers contained in the corresponding MSE optimal ICs. To demonstrate our results, we use the exponential scale model which is related to the Gumbel location model via the transformation $-\log|\cdot|$.

The optimally robust ICs for these two models in case of contamination $(* = c)$ as well as total variation $(* = v)$ neighborhoods are specified in Section 5.2. In both cases $(* = c, v)$ the optimal ICs can be rewritten in such a way that the contained Lagrange multipliers are identical for both models.

As a consequence of this coincidence of the Lagrange multipliers, the least favorable radii and the corresponding MSE–inefficiencies, which are given in Section 5.3, are identical for both models. In case the radius is completely unknown, the maximum efficiency loss is about 38% $(* = c)$ and 22% $(* = v)$, respectively. That is, the loss is larger than in case of normal location, respectively lognormal scale where we obtain about 18% $(* = c, v)$; confer Remark 5.1.9 (b). But, it is smaller than in case of normal scale where the subefficiency is about 50% $(* = c)$ and 25% $(* = v)$, respectively; confer Section 5.2 of Rieder et al. (2001).

The construction problem for one-dimensional location, respectively scale models is treated in Section 5.4. If the considered location, respectively scale model forms a exponential family, we can construct the optimally robust estimators by means of the one-step method; confer Lemma 2.3.6. As initial estimator we propose the Kolmogorov(–Smirnov) minimum distance estimator which has the required properties (strict and $\sqrt{n}$ consistent).

A short description of the implementation of some one-dimensional scale (exponential, normal, lognormal), respectively some one-dimensional location (Gumbel, normal) models is given in Section 5.5. All these models are included in our `R` package `ROptEst` (cf. Appendix D.3) which is part of our `R` bundle `RobASt`.

Gamma Model

In Section F.1 Hampel (1968) treats robust estimation in case of the Gamma model. However, he only considers the estimation of the scale parameter σ for known shape parameter α, respectively the estimation of the shape parameter α for known scale parameter σ and not the simultaneous estimation of scale and shape.

Hampel et al. (1986) (Section 4.4, p 256) consider the robust estimation of the shape parameter α where scale σ is regarded as nuisance. Instead of σ they use the re-parametrization $\nu = \log(\sigma)$ which has been introduced in Example 1 of Subsection 4.3d (ibid.). This re-parametrization endows the Gamma model with a certain invariance structure; confer Section 6.1.

Marazzi and Ruffieux (1996) discuss the implementation of the M estimators for the Gamma model proposed by Hampel et al. (1986). They also work with the re-parametrization ν. In addition, they consider the parametrization $\kappa = \log(\alpha) + \nu$ since their main interest is the estimation of the mean of the Gamma distribution which is $\alpha\sigma = e^{\kappa}$.

Such differentiable parameter transformations with Jacobian matrix of full rank are also allowed in case of the optimal solutions presented in Section 1.3. We use the Gamma model to demonstrate how one can estimate such transformations in our setup. Moreover, the optimality result given in Theorem 1.3.11 is clearly stronger

than the optimality provided by Theorem 4.3.1 of Hampel et al. (1986) (cf. also the discussion before Theorem 4.3.1, ibid.).

In Chapter 6 we first briefly introduce the Gamma as ideal model where we take into account the parameter transformation cited above; confer Section 6.1.

The MSE optimal IC in case of contamination neighborhoods ($* = c$) is specified in Section 6.2. We show how the re-parametrization $\nu = \log(\sigma)$ by means of Theorem 2.4.1 leads to a simplification in our setup, too. However, in contrast to Section 4.4 of Marazzi and Ruffieux (1996), where the standardizing matrices for bijective and differentiable parameter transformations can always be obtained via the corresponding Jacobian matrices, this is not possible in general for the Lagrange multipliers included in our MSE solutions. We may derive valid ICs via the corresponding Jacobian matrices, but, these ICs lead to suboptimal robust estimators which may have a quite large efficiency loss ($> 100\%$); confer Table 6.1.

In Section 6.3 we give some numerical results for the least favorable radii and the corresponding MSE–inefficiencies. In case the true neighborhood radius is completely unknown the maximum subefficiencies are about 50% in all examples considered.

Since the Gamma model forms a exponential family of full rank, we can apply the results of Subsection 2.3.3; confer Section 6.4. That is, we can construct the optimally robust estimators by means of the one-step method using the Kolmogorov(–Smirnov) minimum distance estimator as initial estimator.

A short description of the implementation of the Gamma model is given in Section 6.5. Again, the corresponding optimally robust estimators can be computed via our `R` package `ROptEst` (cf. Appendix D.3) which is part of our `R` bundle `RobASt`. So far (version 0.3-9), package `ROptEst` can be used to compute MSE optimal ICs and estimators for any L_2 differentiable parametric family which is based on a univariate distribution.

Chapter 3

Binomial Model

We first briefly introduce the binomial model where the probability of success has to be estimated; confer Section 3.1. We then specify the MSE optimal ICs provided by Theorem 1.3.11 and investigate the continuity and the smoothness of the Lagrange multipliers contained in these optimally robust ICs. In addition, we prove that there is a normal approximation available for these Lagrange multipliers; confer Section 3.2. In Section 3.3 we then present some numerical results for the least favorable radii and the corresponding MSE–inefficiencies. Subsequently, we verify that we can construct the asymptotically optimal estimators by means of one-step constructions (cf. Section 3.4) and explain how one can use our `R` package `ROptEst` (cf. Appendix D.3) to re-compute all the presented results and to estimate the unknown probability of success in an optimally robust way; confer Section 3.5. We conclude this chapter with a small simulation study which demonstrates the usage of our `R` package `ROptEst` and indicates the necessity of robust estimation in the binomial model; confer Section 3.6.

3.1 Introduction

The binomial model with unknown probability of success reads

$$\mathcal{P} = \left\{ \mathrm{Binom}(m, \theta) \,\middle|\, \theta \in (0,1) \right\} \tag{3.1.1}$$

where

$$\mathrm{Binom}(m,\theta)(\{y\}) = \binom{m}{y} \theta^y (1-\theta)^{m-y} \qquad y \in \{0,1,\ldots,m\} \tag{3.1.2}$$

and the size $m \in \mathbb{N}$ is known.

Remark 3.1.1 The binomial model (3.1.1) forms an exponential family with respect to the counting measure on $\{0,1,\ldots,m\}$ as we can rewrite,

$$\binom{m}{y} \theta^y (1-\theta)^{m-y} = \binom{m}{y} \exp\left\{ y \log \frac{\theta}{1-\theta} - m \log \frac{1}{1-\theta} \right\} \tag{3.1.3}$$

confer also Beispiel 1.152 of Witting (1985) and Example 1.5.11 of Lehmann and Casella (1998), respectively. Using the notation of Lemma 2.3.6, we get $\zeta(\theta) = \log \frac{\theta}{1-\theta}$, $\beta(\theta) = m \log \frac{1}{1-\theta}$, $T(y) = y$ and $h(y) = \binom{m}{y}$ which leads to $\mathcal{J}_\zeta = \frac{1}{\theta(1-\theta)}$, $\mathrm{E}_\theta\, T = m\theta$ and $\mathrm{Var}_\theta\, T = m\theta(1-\theta)$. ////

Lemma 3.1.2 *The binomial model (3.1.1) is L_2 differentiable at $\theta \in (0,1)$ with L_2 derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by*

$$\Lambda_\theta(y) = \frac{y - m\theta}{\theta(1-\theta)} \qquad \mathcal{I}_\theta = \frac{m}{\theta(1-\theta)} \tag{3.1.4}$$

Proof A consequence of Lemma 2.3.6 (a) in connection with Remark 3.1.1. ////

Remark 3.1.3 **(a)** Some applications of the binomial model are for instance given in Section VI.2 of Feller (1968) or Section 4.8 of Pfanzagl (1988) and of course, in many other text books on probability theory.

(b) Since y ranges only over a bounded set, the IC

$$\eta_h(y) = \mathcal{I}_\theta^{-1}\Lambda_\theta = \frac{y}{m} - \theta \tag{3.1.5}$$

of the maximum likelihood estimator (MLE) is bounded and therefore is robust in the sense that one observation can only have a bounded influence on the estimation. In particular, for $m = 1$ and $\theta \in (0,1)$, respectively for $m = 2$ and $\theta = 0.5$ the optimally robust IC with respect to the maximum asymptotic MSE for contamination $(* = c)$ as well as total variation neighborhoods $(* = v)$ is

$$\tilde\eta_{*,r}(y) \equiv \eta_h(y) = \frac{y}{m} - \theta \tag{3.1.6}$$

for all $r \in [0,\infty]$; confer also Remark 2.1.6 (a). Consequentially, we obtain $\mathrm{relMSE}(\tilde\eta_{*,r_0}, r) \equiv 1$ for all $r_0, r \in [0,\infty]$, hence the MLE is also radius–minimax, as defined in Section 2.2, in these cases. But, nevertheless, as the examples on page 98 of Hampel (1968) (data generated by a binomial distribution with very large size containing outliers) and page 1119 of Ruckstuhl and Welsh (2001) (data generated by an equal mixture of two binomial distributions) as well as our small simulation study in Section 3.6 indicate, there is also the need for robust estimators for $m > 1$.

(c) Robust estimation in generalized linear models (logistic regression) is quite different from the simple binomial model and goes beyond the scope of this chapter. Some references for robust estimation in logistic regression are for instance given in Ruckstuhl and Welsh (2001). ////

3.2 Optimally Robust Influence Curves

3.2.1 Contamination Neighborhoods

3.2.1.1 Mean Square Error Solution

The unique MSE optimal IC $\tilde\eta_{c,r}$ for infinitesimal contamination neighborhoods (1.2.4) and radius $r \in (0,\infty)$ may be read off from Theorem 1.3.7 (a) and Theo-

rem 1.3.11 (b). For some given $D \in \mathbb{R}\setminus\{0\}$ we can rewrite the solution as

$$\tilde{\eta}_{c,r}(y) = A_r(\Lambda(y) - z_r)\min\left\{1, \frac{c_r}{|\Lambda(y) - z_r|}\right\} \tag{3.2.1}$$

where

$$0 = \mathrm{E}(\Lambda - z_r)\min\left\{1, \frac{c_r}{|\Lambda - z_r|}\right\} \tag{3.2.2}$$

$$D = A_r\,\mathrm{E}\,|\Lambda - z_r|\min\left\{|\Lambda - z_r|, c_r\right\} \tag{3.2.3}$$

and

$$r^2\,c_r = \mathrm{E}\left(|\Lambda - z_r| - c_r\right)_+ \tag{3.2.4}$$

For $r = \infty$ we obtain by Theorem 1.3.7 (b)

$$\tilde{\eta}_{c,\infty}(y) = \omega_c^{\min}\,\mathrm{sign}(D)\Big[\mathrm{I}(y > M) - \mathrm{I}(y < M) + \beta\,\mathrm{I}(y = M)\Big] \tag{3.2.5}$$

where

$$\beta = \Big[P(y < M) - P(y > M)\Big]\Big/P(y = M) \tag{3.2.6}$$

with any $M = \mathrm{med}(y)$ and $\tilde{\eta}_{c,\infty}$ attains the minimum bias

$$\omega_c^{\min} = \frac{|D|\theta(1-\theta)}{\mathrm{E}\,|y - M|} \tag{3.2.7}$$

confer also Remark 1.3.8.
For a plot of the optimally robust ICs in case $m = 25$, $\theta = 0.25$ and for different values of r see Figure 3.1.

Remark 3.2.1 **(a)** Since ICs are defined with respect to the corresponding parametric model (cf. Definition 1.1.1), the ICs of the binomial model are defined on $\{0, 1, \ldots, m\}$. However, if we consider neighborhoods of the ideal binomial model, we may allow distributions whose support is no longer restricted to $\{0, 1, \ldots, m\}$ but can be any subset T of $\mathbb{N}$, $\mathbb{Z}$ or even $\mathbb{R}$, respectively. In view of the construction of the corresponding optimally robust estimator one has to choose an extension such that $|\tilde{\eta}_r| \le b_r = A_r c_r$ on T. Otherwise the bias would increase if we pass over from the neighborhood submodel to full neighborhoods. In view of Theorem 2.3.3 and Lemma 2.3.6 one possible choice is to extend $\tilde{\eta}_{c,r}$ to $T \setminus \{0, 1, \ldots, m\}$ simply by regarding Λ as function on T.

(b) We only need to consider $\theta \in (0, 0.5]$. Since

$$\mathrm{Binom}\,(m, \theta)(\{y\}) = \mathrm{Binom}\,(m, 1-\theta)(\{m - y\}) \tag{3.2.8}$$

it is equivalent to use $\bar{\theta} := 1 - \theta$ and $\bar{y}_1, \ldots, \bar{y}_n := m - y_1, \ldots, m - y_n$ for $\theta > 0.5$ and estimate $\bar{\theta}$. Hence, the results (e.g., for the Lagrange multipliers and the lower case radius $\bar{r}$) we obtain are symmetric to $\theta = 0.5$. This equivariance of the binomial model is in more detail considered in Example 3.1.1 of Lehmann and Casella (1998).

(c) As for each m, $\mathrm{med}(\Lambda)$ is non-unique for some $\theta \in (0, 1)$ and the assumptions of Proposition 2.1.3 hold, z_r is non-unique if $r \ge \bar{r}$ where the lower case

radius $\bar{r}$ is defined in (2.1.12). For a plot of $\bar{r}$ on $(0, 0.5]$ for different values of m confer Figure 3.2. The upper peaks correspond to those values of θ for which the median of Λ is non-unique.

(d) Since $\mathcal{L}(\Lambda)$ is symmetric around zero for $\theta = 0.5$, we can choose $z_r = 0$ for all $r \in (0, \infty)$, respectively $M = 0$ for $r = \infty$; i.e., in this case, we only have to determine the clipping bound c_r and the standardizing constant A_r.

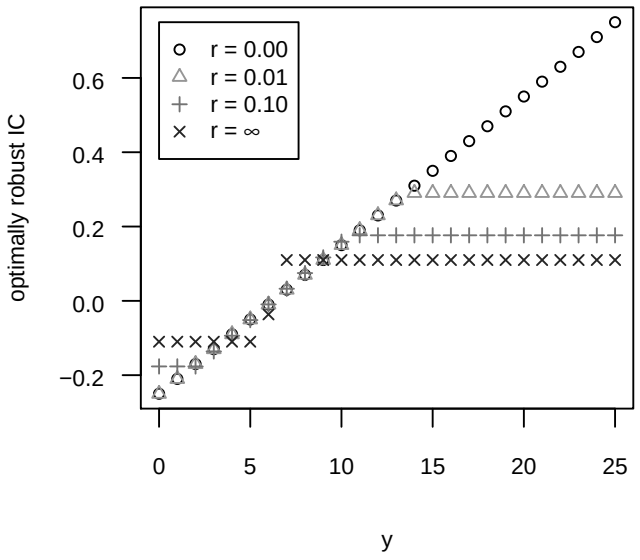

Figure 3.1: Optimally robust ICs for $\text{Binom}(25, 0.25)$ in case of contamination neighborhoods $(* = c)$ with (starting) radius $r = 0, 0.01, 0.10, \infty$.

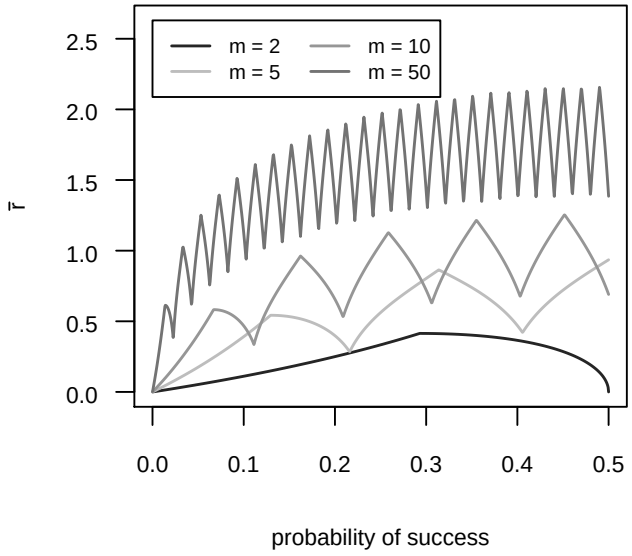

Figure 3.2: Lower case radius $\bar{r}$ for $\theta \in (0, 0.5]$ (results symmetric to $\theta = 0.5$) and sizes $m = 2, 5, 10, 50$ in case of contamination neighborhoods $(* = c)$.

3.2.1.2 Continuity and Uniqueness of Lagrange Multipliers

The continuity of the Lagrange multipliers A_r, and b_r in r, stated in Proposition 2.1.9, is visualized in Figure 3.3. Since $\text{med}(\Lambda)$ is non-unique for $\theta = 1 - \sqrt[25]{0.5}$, there is a whole interval of valid centering constants a_r for $r \geq \bar{r} \approx 0.606$. The boundaries of this interval can be determined via (2.1.17) and are given in Figure 3.3. In contrast to the standardized bias b_r and the asymptotic variance $A_r - r^2 b_r^2$ which seem to be non-differentiable at some values of r, A_r looks very smooth in r.

The Lagrange multipliers A_r and $b_r = A_r c_r$ and therefore also the minimax asymptotic MSE and the asymptotic variance are continuous in θ as proven in Theorem 2.1.11. However, as part (c) of Remark 3.2.1 already indicates this is not necessarily true in case of the corresponding centering constant $a_r = A_r z_r$. This fact is illustrated in Figure 3.4 where we choose $r \geq \bar{r}$ to demonstrate the extreme case. The discontinuity points coincide with those values of θ for which the median of Λ is non-unique. In addition, this plot indicates that the Lagrange multipliers and hence the standardized asymptotic bias, the asymptotic variance and the maximum asymptotic MSE are not necessarily smooth functions in θ. ////

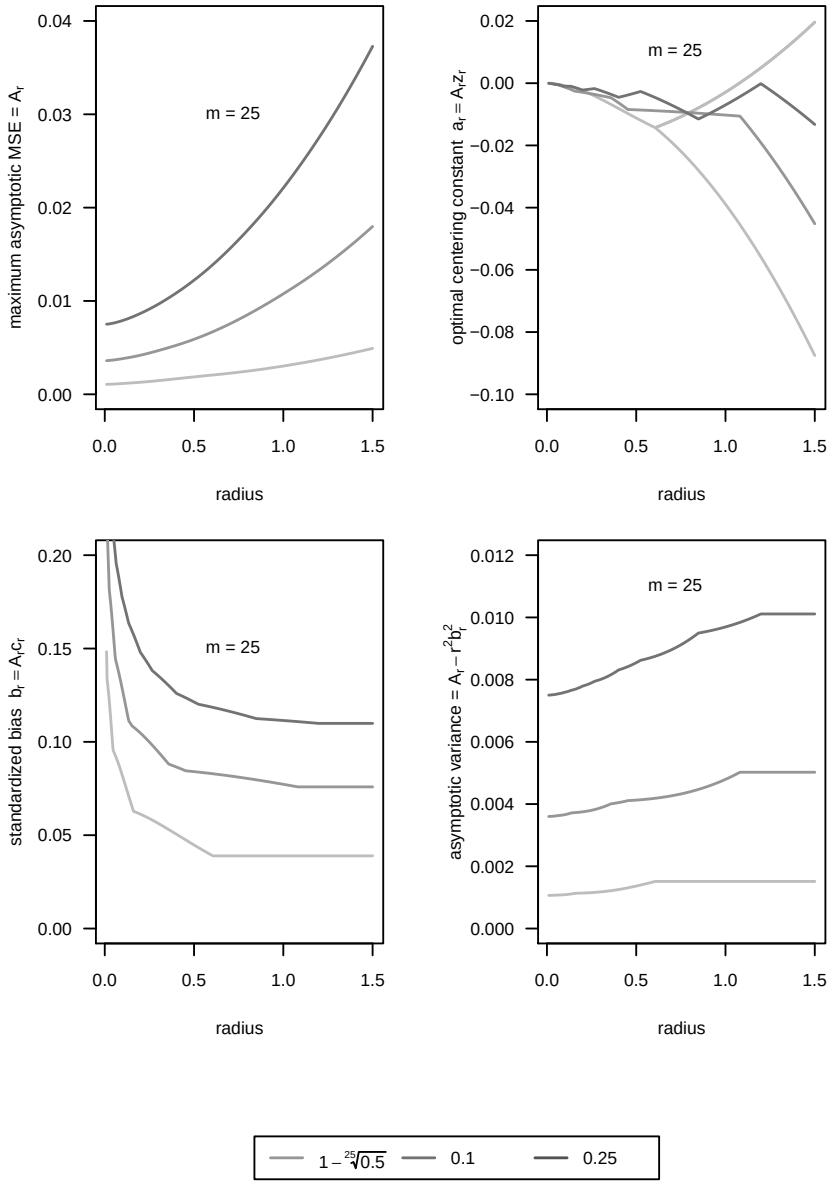

Figure 3.3: Continuity in the radius r of the Lagrange multipliers contained in the MSE optimal ICs for $r \in (0, 1.5]$, $\theta = 1 - \sqrt[25]{0.5}, 0.1, 0.25$ and $m = 25$ in case of contamination neighborhoods $(* = c)$.

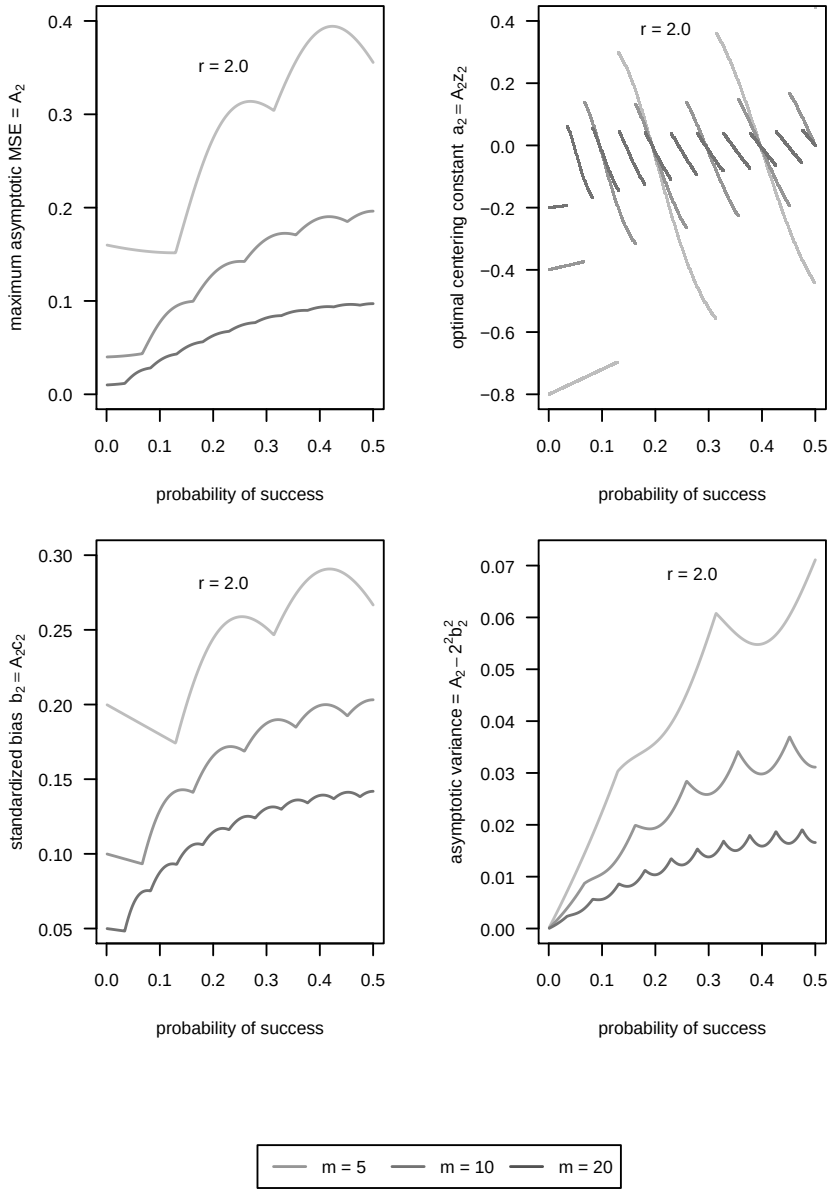

Figure 3.4: Continuity in θ of the Lagrange multipliers contained in the MSE optimal ICs for $\theta \in (0, 0.5]$ (results symmetric to $\theta = 0.5$), $r = 2.0$ and $m = 5, 10, 20$ in case of contamination neighborhoods ($* = c$).

3.2.1.3 Normal Approximation

The following lemma states the normal approximation for the optimally robust ICs, respectively for the Lagrange multipliers contained in these ICs which is a consequence of Theorem 2.4.1. The corresponding optimally robust IC in case of one-dimensional normal location is

$$\tilde{\eta}_{c,r}^{1.\mathrm{loc}}(y) = A_r^{1.\mathrm{loc}} y \min\left\{1, c_r^{1.\mathrm{loc}}|y|^{-1}\right\} \tag{3.2.9}$$

with

$$1 = A_r^{1.\mathrm{loc}} \,\mathrm{E}_{\mathcal{N}(0,1)}\, |y| \min\left\{|y|,\, c_r^{1.\mathrm{loc}}\right\} \tag{3.2.10}$$

and

$$r^2\, c_r^{1.\mathrm{loc}} = \mathrm{E}_{\mathcal{N}(0,1)}\left(|y| - c_r^{1.\mathrm{lco}}\right)_+ \tag{3.2.11}$$

For $r = \infty$ we obtain

$$\tilde{\eta}_{c,\infty}^{1.\mathrm{loc}}(y) = \sqrt{\tfrac{\pi}{2}}\,\mathrm{sign}(y) \tag{3.2.12}$$

which attains the minimum bias $\omega_c^{\min,\,1.\mathrm{loc}} = \sqrt{\frac{\pi}{2}}$; confer also Subsection 2.2.2 of Rieder et al. (2001).

Lemma 3.2.2 *Let* $D = 1$ *and* $\gamma_m = \sqrt{\frac{m}{\theta(1-\theta)}}$. *Then,*

$$\lim_{m\to\infty} \gamma_m^2 A_r = A_r^{1.\mathrm{loc}} \qquad \lim_{m\to\infty} \gamma_m^{-1} z_r = z_r^{1.\mathrm{loc}} = 0 \qquad \lim_{m\to\infty} \gamma_m^{-1} c_r = c_r^{1.\mathrm{loc}} \tag{3.2.13}$$

for all $r \in (0,\infty)$ *and*

$$\lim_{m\to\infty} \gamma_m \omega_c^{\min} = \omega_c^{\min,\,1.\mathrm{loc}} \tag{3.2.14}$$

Proof We have

$$\gamma_m^{-1}\Lambda_m(y) = \tfrac{y - m\theta}{\sqrt{m\theta(1-\theta)}} \tag{3.2.15}$$

and the central limit theorem of de Moivre-Laplace yields $\mathcal{L}(\gamma_m^{-1}\Lambda_m) \xrightarrow{\mathrm{w}} \mathcal{N}(0,1)$ as $m \to \infty$ where $\mathrm{E}\,\gamma_m^{-1}\Lambda_m = 0$ and $\mathrm{E}\left(\gamma_m^{-1}\Lambda_m\right)^2 = 1$ for all $m \in \mathbb{N}$; i.e.,

$$\lim_{m\to\infty} \mathrm{E}\left(\gamma_m^{-1}\Lambda_m\right)^2 = 1 = \mathrm{E}_{\mathcal{N}(0,1)}\, y^2 \tag{3.2.16}$$

Hence, we can apply Theorem 2.4.1 which yields (3.2.13) and (3.2.14). ////

Remark 3.2.3 **(a)** The convergence of the standardized optimal clipping bounds $\gamma_m^{-1} z_r$, respectively of the standardized bias terms $\gamma_m \omega_c(\tilde{\eta}_{c,r})$ is illustrated in Figure 3.5 and Figure 3.6, respectively. In case $r = \infty$, the discontinuity points of the centering constant z_r coincide with those values of θ for which the median of Λ is non-unique. At these non-uniqueness points the standardized (infinitesimal) bias terms attain local minima in case $r = \infty$.

(b) Some examples for the convergence of the standardized minimax asymptotic MSE $\gamma_m^2 A_r$, respectively the MSE–inefficiencies are given in Figure 3.7 and Subsection 3.3.1, respectively.

(c) Figure 3.8 shows the MSE–inefficiency of the normal approximated IC. That is, we use the Lagrange multipliers $A_r^{1.\mathrm{loc}}/\gamma_m^2$, $z_r^{1.\mathrm{loc}}\gamma_m$ and $c_r^{1.\mathrm{loc}}\gamma_m$ instead of the optimal A_r, z_r and c_r; confer Lemma 3.2.2. To make sure that the resulting function is indeed an IC (with respect to the binomial model), we additionally centered and standardized this function. In case the radius is not too large ($r \leq 0.5$), the MSE–inefficiency of this normal approximation is very small independent of the size m. To get a good approximation for larger radii, we need moderate to large sizes m depending on the parameter θ. Thus, these numerical results indicate that the "distance" between the optimal IC and its normal approximation also depends on the radius r as well as on the parameter θ. In particular, for radii larger than the lower case radius $\bar{r}$ the approximation seems to work best for those values of θ where $\mathrm{med}(\Lambda)$ is non-unique, respectively $\theta \leq 1 - \sqrt[m]{0.5}$ or $\theta \geq \sqrt[m]{0.5}$. This result, the results in case of total variation neighborhoods (cf. Subsubsection 3.2.2.3) and further numerical investigations indicate that this behavior is caused by the fact that the lower case solution in case of the binomial model and contamination neighborhoods in all cases, except $\mathrm{med}(\Lambda)$ is non-unique, respectively $\theta \leq 1 - \sqrt[m]{0.5}$ or $\theta \geq \sqrt[m]{0.5}$, attains three values ($\pm\omega_c^{\min}$, β) and the third value β in most cases is only badly approximated by the normal approximation. ////

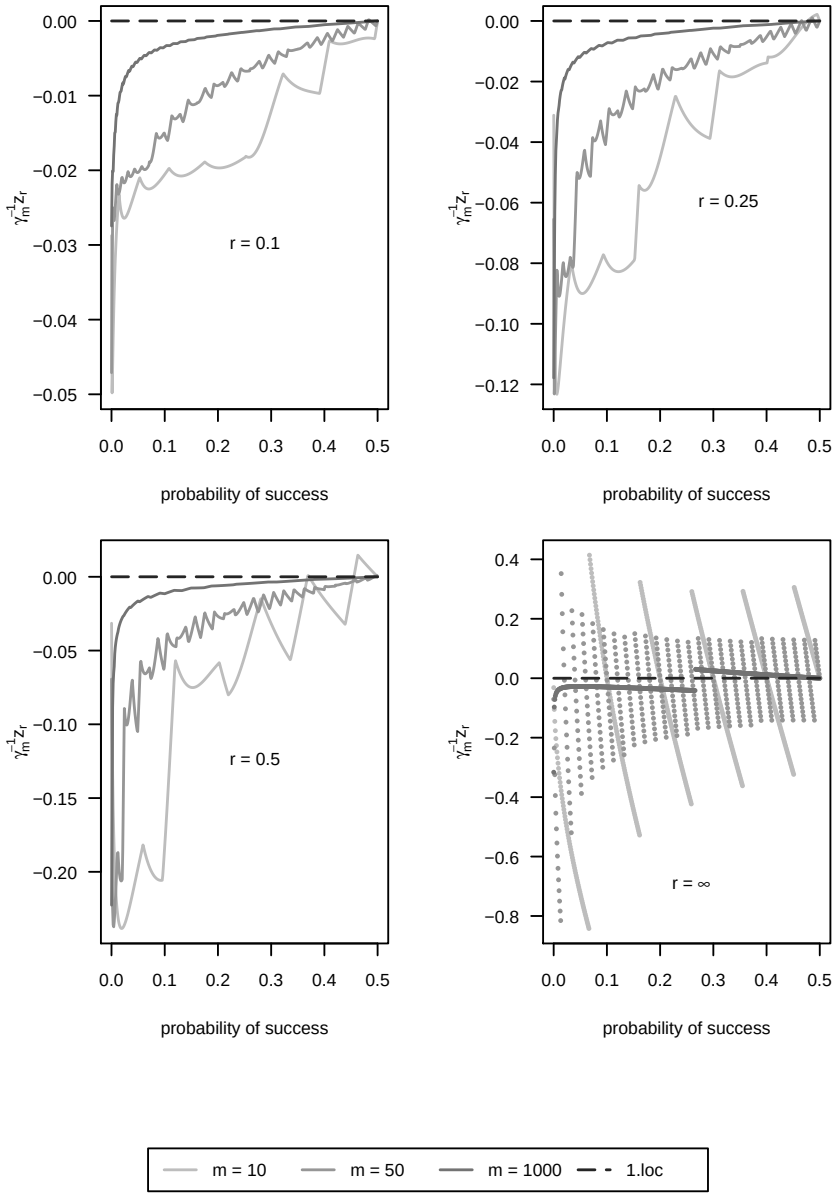

Figure 3.5: Normal approximation of the standardized centering constant $\gamma_m^{-1} z_r$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) in case of contamination neighborhoods ($* = c$) with radius $r = 0.1, 0.25, 0.5, \infty$.

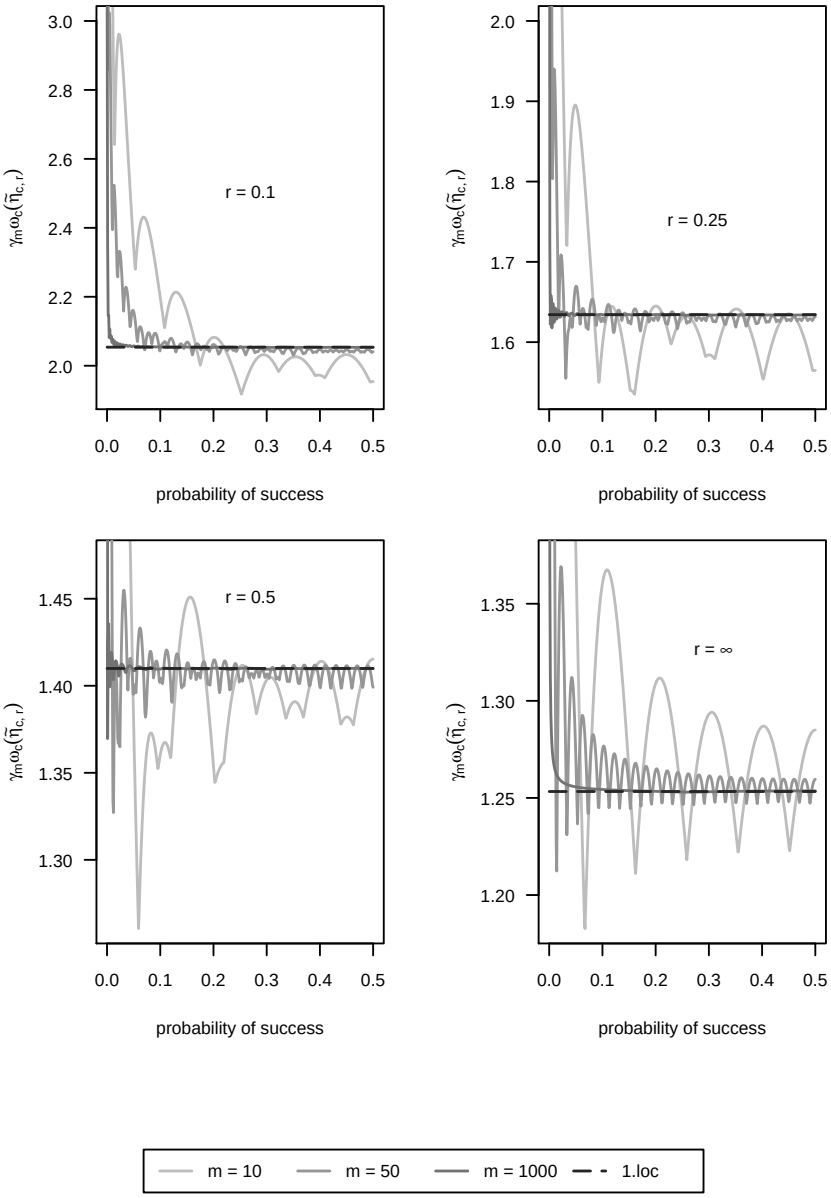

Figure 3.6: Normal approximation of the standardized infinitesimal bias terms $\gamma_m \omega_c(\tilde{\eta}_{c,r})$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) in case of contamination neighborhoods ($* = c$) with radius $r = 0.1, 0.25, 0.5, \infty$.

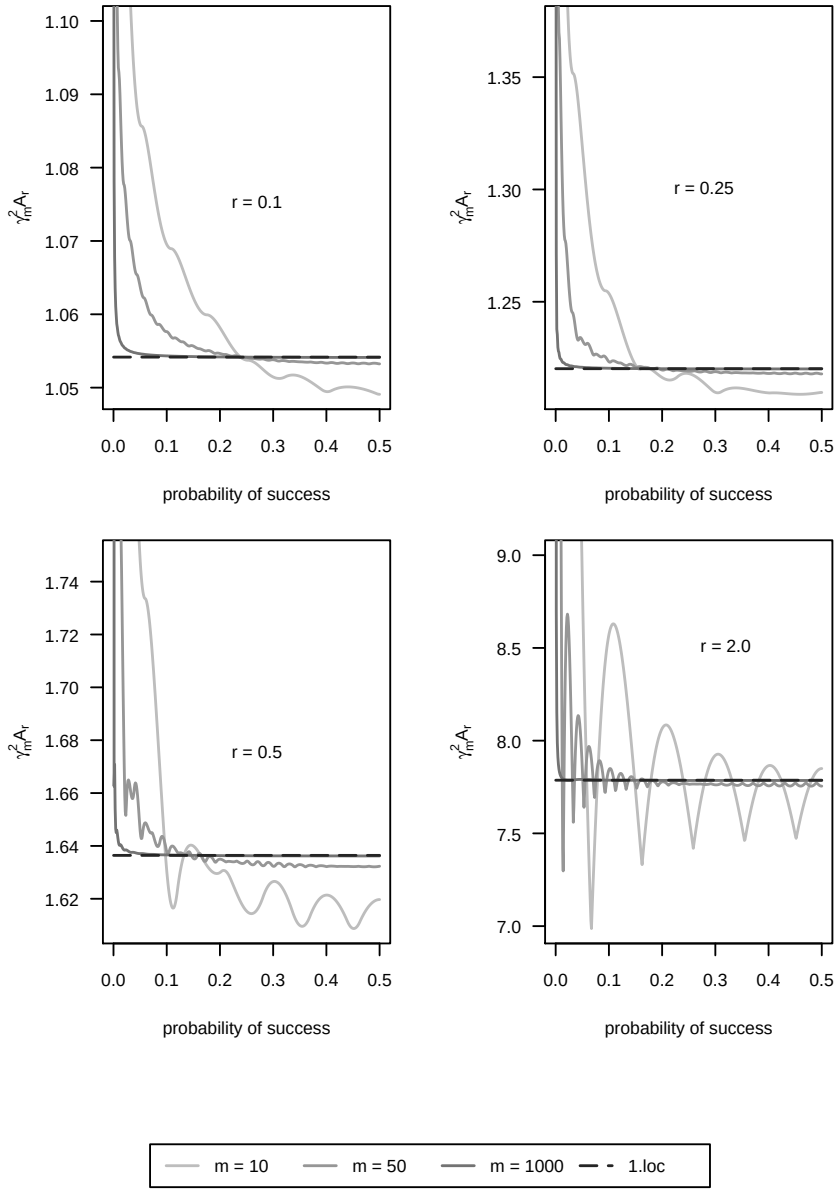

Figure 3.7: Normal approximation of the standardized maximum asymptotic MSE $\gamma_m^2 A_r$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) for contamination neighborhoods ($* = c$) with radius $r = 0.1, 0.25, 0.5, 2.0$.

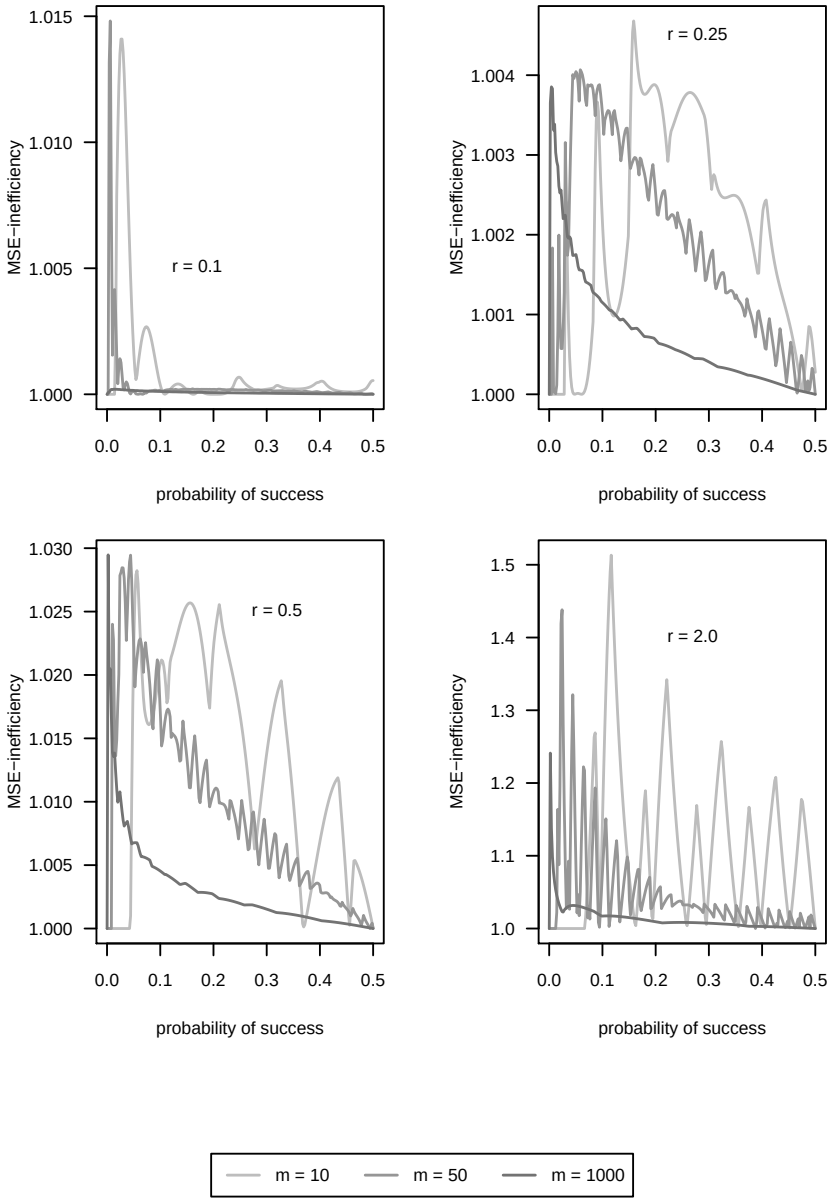

Figure 3.8: MSE–inefficiency of the normal approximated IC for $\theta \in (0, 0.5]$ (results symmetric to 0.5) in case of contamination neighborhoods with radius $r = 0.1, 0.25, 0.5, 2.0$.

3.2.2 Total Variation Neighborhoods

3.2.2.1 Mean Square Error Solution

The optimally robust IC $\tilde{\eta}_{v,r}$ for infinitesimal total variation neighborhoods (1.2.5) provided by Theorem 1.3.9 (a) and Theorem 1.3.11 (c) for some given $D \in \mathbb{R} \setminus \{0\}$ can be rewritten as

$$\tilde{\eta}_{v,r}(y) = A_r\big[\, g_r \vee \Lambda(y) \wedge (g_r + c_r)\big] \tag{3.2.17}$$

where

$$0 = \mathrm{E}\,\big(g_r - \Lambda\big)_+ - \mathrm{E}\,\big(\Lambda - (g_r + c_r)\big)_+ \tag{3.2.18}$$

$$D = A_r\,\mathrm{E}\,\Lambda\big[\, g_r \vee \Lambda \wedge (g_r + c_r)\big] \tag{3.2.19}$$

and

$$r^2\, c_r = \mathrm{E}\,\big(\Lambda - (g_r + c_r)\big)_+ \tag{3.2.20}$$

For $r = \infty$ we get by Theorem 1.3.9 (b)

$$\tilde{\eta}_{v,\infty}\,(y) = \omega_v^{\min}\,\mathrm{sign}(D)\left(\frac{P(y < m\theta)}{P(y \neq m\theta)}\,\mathrm{I}(y > m\theta) - \frac{P(y > m\theta)}{P(y \neq m\theta)}\,\mathrm{I}(y < m\theta)\right) \tag{3.2.21}$$

with minimum bias

$$\omega_v^{\min} = \frac{|D|\theta(1-\theta)}{\mathrm{E}(y - m\theta)_+} \tag{3.2.22}$$

For a plot of the optimally robust ICs in case $m = 25$, $\theta = 0.25$ and for different values of r see Figure 3.9.

Remark 3.2.4 **(a)** As in case of contamination neighborhoods, it might be necessary to extend the optimal ICs to $\mathbb{R} \setminus \{0, 1, \ldots, m\}$. Moreover, we again only consider $\theta \in (0, 0.5]$ by symmetry. For more details we refer to Remark 3.2.1.

(b) Since $\mathcal{L}(\Lambda)$ is symmetric around zero for $\theta = 0.5$, we obtain $\tilde{\eta}_{v,r} = \tilde{\eta}_{c,2r}$ for all $r \in [0, \infty]$. In particular, $\omega_c^{\min} = 2\omega_v^{\min}$.

(c) The map $\theta \mapsto \bar{r}$ is discontinuous at $\big\{\theta \in (0,1)\,\big|\, m\theta \in \{1, \ldots, m-1\}\big\}$. By choosing $\theta \in (0,1)$ such that $m\theta$ is arbitrarily close to some integer value in $\{1, \ldots, m-1\}$, the gap $\gamma = \inf_{P_\theta}\big\{|\Lambda|\,\big|\,|\Lambda| > 0\big\}$ becomes arbitrarily small, respectively M and $\bar{r}$ as defined in (2.1.34) and (2.1.33) become arbitrarily large. But, if we choose $\theta \in (0,1)$ such that $m\theta \in \{1, \ldots, m-1\}$ we get $\gamma = 1/(\theta(1-\theta))$ $(= \gamma_1 = -\gamma_2)$; confer also Figure 3.10. ////

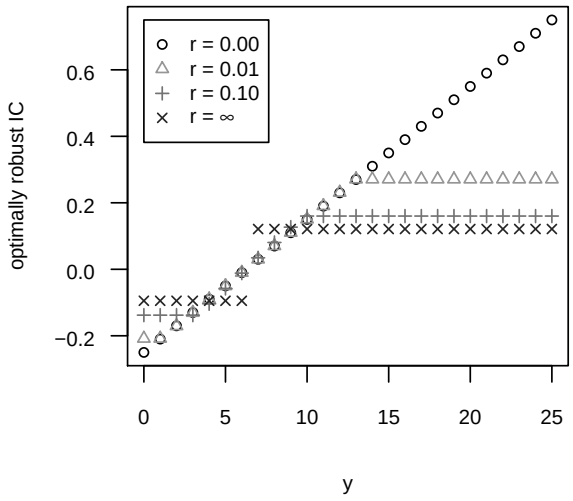

Figure 3.9: Optimally Robust ICs for $\text{Binom}(25, 0.25)$ in case of total variation neighborhoods $(* = v)$ with radius $r = 0, 0.01, 0.10, \infty$.

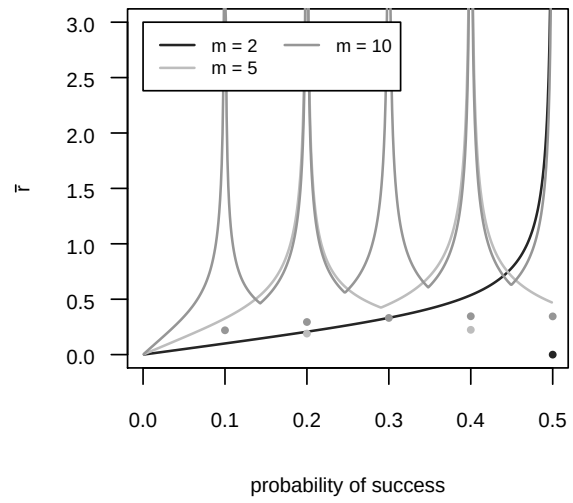

Figure 3.10: Lower case radius $\bar{r}$ for sizes $m = 2, 5, 10$ in case of total variation neighborhoods $(* = v)$.

3.2.2.2 Continuity and Uniqueness of Lagrange Multipliers

In contrast to contamination neighborhoods, all Lagrange multipliers contained in the optimally robust ICs are unique and are continuous in the radius r for $\theta \in (0,1)$ as well as in the parameter θ for $r \in (0,\infty)$; confer Subsection 2.1.4, Proposition 2.1.9 and Theorem 2.1.11, respectively. This is illustrated in Figure 3.11 and Figure 3.12, respectively. In Figure 3.12 we use the `R` function `points` (cf. R Development Core Team (2005)) to illustrate that the lower clipping bound and the asymptotic variance indeed have no jumps but attain values between the local extrema, too. However, again the Lagrange multipliers and hence the standardized asymptotic bias, the asymptotic variance and the maximum asymptotic MSE are not necessarily smooth functions in r, respectively θ. Only the maximum asymptotic MSE A_r seems to be smooth in r. Moreover, as the form of the lower case solution $\tilde{\eta}_{v,\infty}$ already indicates, the lower and the upper clipping bound in case $r = \infty$ are discontinuous for those values of θ such that $m\theta \in \mathbb{N}$; confer Figure 3.13.

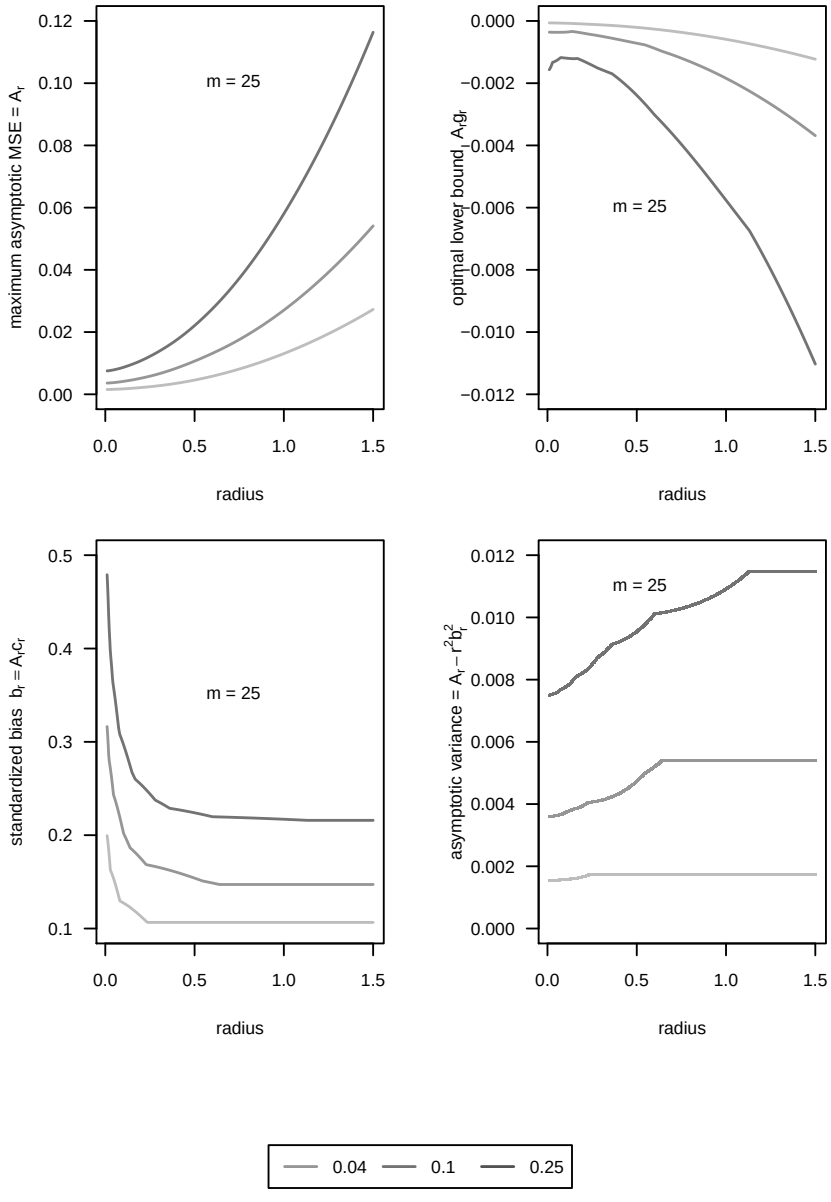

Figure 3.11: Continuity in the radius r of the Lagrange multipliers contained in the MSE optimal ICs for $r \in (0, 1.5]$, $\theta = 0.04, 0.1, 0.25$ and $m = 25$ in case of total variation neighborhoods $(* = v)$.

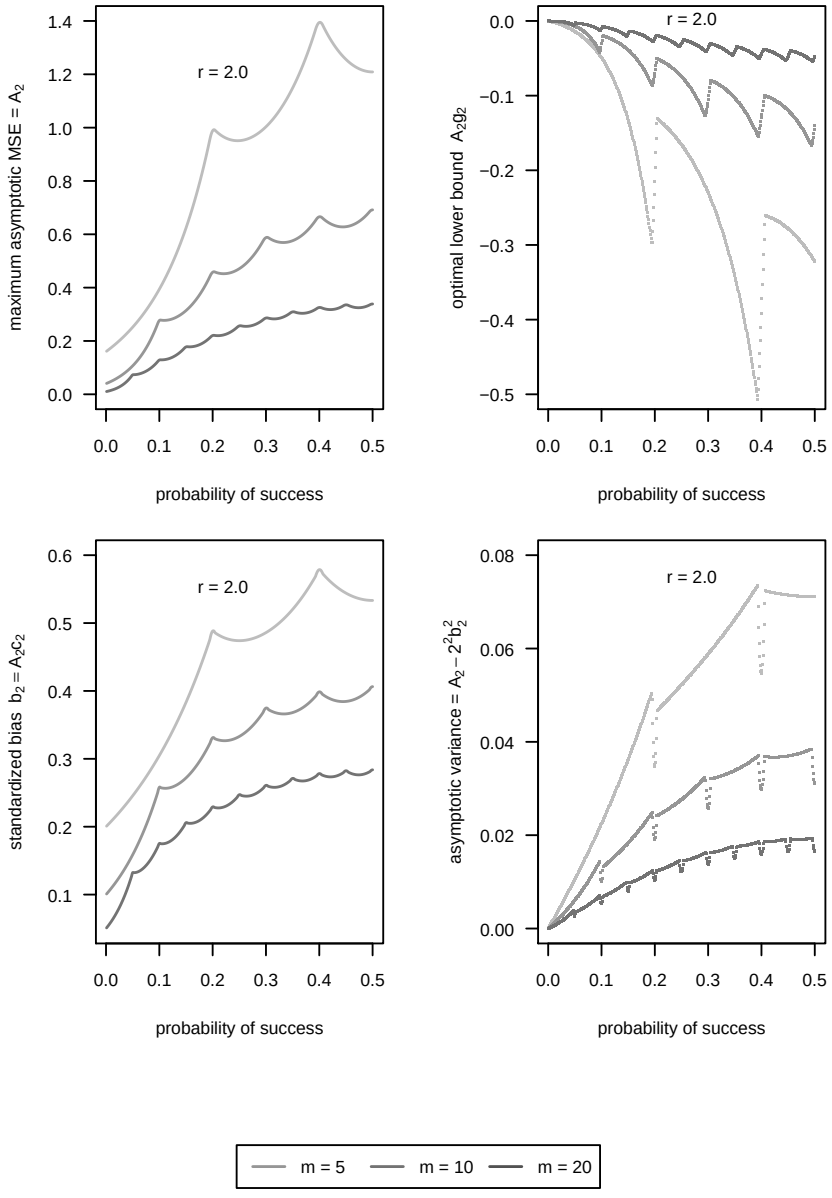

Figure 3.12: Continuity in θ of the Lagrange multipliers contained in the MSE optimal ICs for $\theta \in (0, 0.5]$ (results symmetric to $\theta = 0.5$), $r = 2.0$ and $m = 5, 10, 20$ in case of total variation neighborhoods ($* = v$).

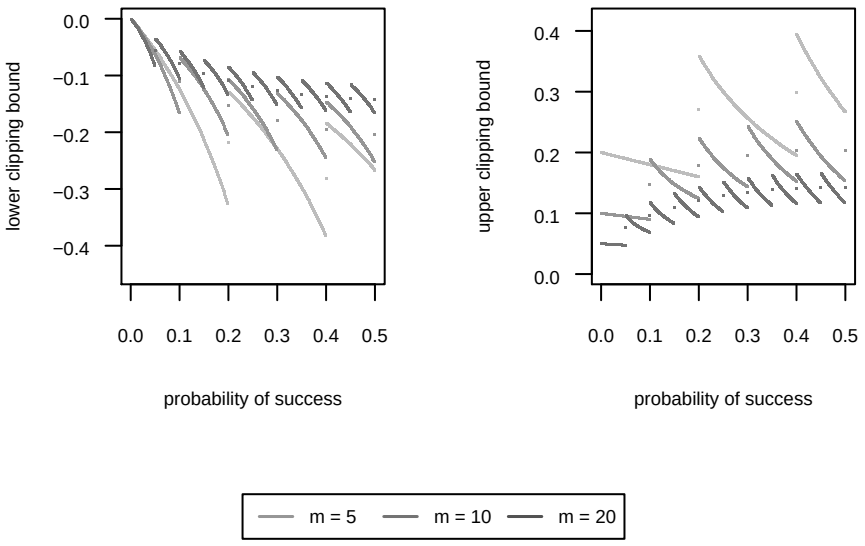

Figure 3.13: Discontinuity of the lower and upper clipping bound for $\theta \in (0, 0.5]$ (results symmetric to $\theta = 0.5$) and sizes $m = 5, 10, 20$ in case of total variation neighborhoods ($* = v$) with radius $r = \infty$.

3.2.2.3 Normal Approximation

In case of one-dimensional normal location and total variation neighborhoods the corresponding optimally robust IC reads

$$\tilde{\eta}_{v,r}^{1.\mathrm{loc}}(y) = A_r\big[g_r \vee y \wedge (g_r + c_r)\big] \qquad g_r = -c_r/2 \tag{3.2.23}$$

where

$$1 = A_r \,\mathrm{E}\, y\big[g_r \vee y \wedge (g_r + c_r)\big] \tag{3.2.24}$$

and

$$r^2\, c_r = \mathrm{E}(g_r - y)_+ \tag{3.2.25}$$

It follows $\tilde{\eta}_{v,r}^{1.\mathrm{loc}} = \tilde{\eta}_{c,2r}^{1.\mathrm{loc}}$; confer also Remark 1.3 of Rieder et al. (2001). For $r = \infty$ we obtain

$$\tilde{\eta}_{v,\infty}^{1.\mathrm{loc}}(y) = \sqrt{\tfrac{\pi}{2}}\ \mathrm{sign}(y) = \tilde{\eta}_{c,\infty}^{1.\mathrm{loc}}(y) \tag{3.2.26}$$

which attains the minimum bias $\omega_v^{\min,\,1.\mathrm{loc}} = \sqrt{2\pi} = 2\omega_c^{\min,\,1.\mathrm{loc}}$. The following lemma states the normal approximation of the Lagrange multipliers contained in

the optimally robust ICs in case of total variation neighborhoods.

Lemma 3.2.5 *Let* $D = 1$ *and* $\gamma_m = \sqrt{\frac{m}{\theta(1-\theta)}}$. *Then,*

$$\lim_{m\to\infty} \gamma_m^2 A_r = A_r^{1.\mathrm{loc}} \qquad \lim_{m\to\infty} \gamma_m^{-1} g_r = g_r^{1.\mathrm{loc}} \qquad \lim_{m\to\infty} \gamma_m^{-1} c_r = c_r^{1.\mathrm{loc}} \tag{3.2.27}$$

for all $r \in (0,\infty)$ *and*

$$\lim_{m\to\infty} \gamma_m \omega_v^{\min} = \omega_v^{\min,\,1.\mathrm{loc}} \tag{3.2.28}$$

PROOF Consequence of Theorem 2.4.1; confer also the proof of Lemma 3.2.2. ////

Remark 3.2.6 (a) The convergence of the standardized lower clipping bounds $\gamma_m A_r g_r$, respectively of the standardized infinitesimal bias terms $\gamma_m \omega_v(\tilde{\eta}_{v,r})$ is visualized in Figure 3.14 and Figure 3.15, respectively. In case $r = \infty$, we again see that the discontinuity points of the lower clipping bound $A_r g_r$ coincide with those values of θ for which $m\theta \in \mathbb{N}$. At these non-uniqueness points the standardized infinitesimal bias terms attain local maxima in case $r = \infty$.

(b) Some examples for the convergence of the standardized maximum asymptotic MSE $\gamma_m^2 A_r$, respectively the MSE–inefficiencies are given in Figure 3.16 and Subsection 3.3.2, respectively.

(c) Figure 3.16 shows the MSE–inefficiency of the normal approximated IC. That is, we use the Lagrange multipliers $A_r^{1.\mathrm{loc}}/\gamma_m^2$, $g_r^{1.\mathrm{loc}}\gamma_m$ and $c_r^{1.\mathrm{loc}}\gamma_m$ instead of the optimal A_r, g_r and c_r; confer Lemma 3.2.5. To make sure that the resulting function is indeed an IC (with respect to the binomial model), we additionally centered and standardized this function. We get very small MSE–inefficiencies in case of total variation neighborhoods and the approximation even seems to become better with increasing radius. This is contrary to contamination neighborhoods where the MSE–inefficiencies may be quite large for radii $r > \bar{r}$ and the quality of the approximation decreases with increasing radius. Interestingly, for small radii r ($r < 0.5$) the approximation in case of contamination neighborhoods seem to be better than in case of total variation neighborhoods. ////

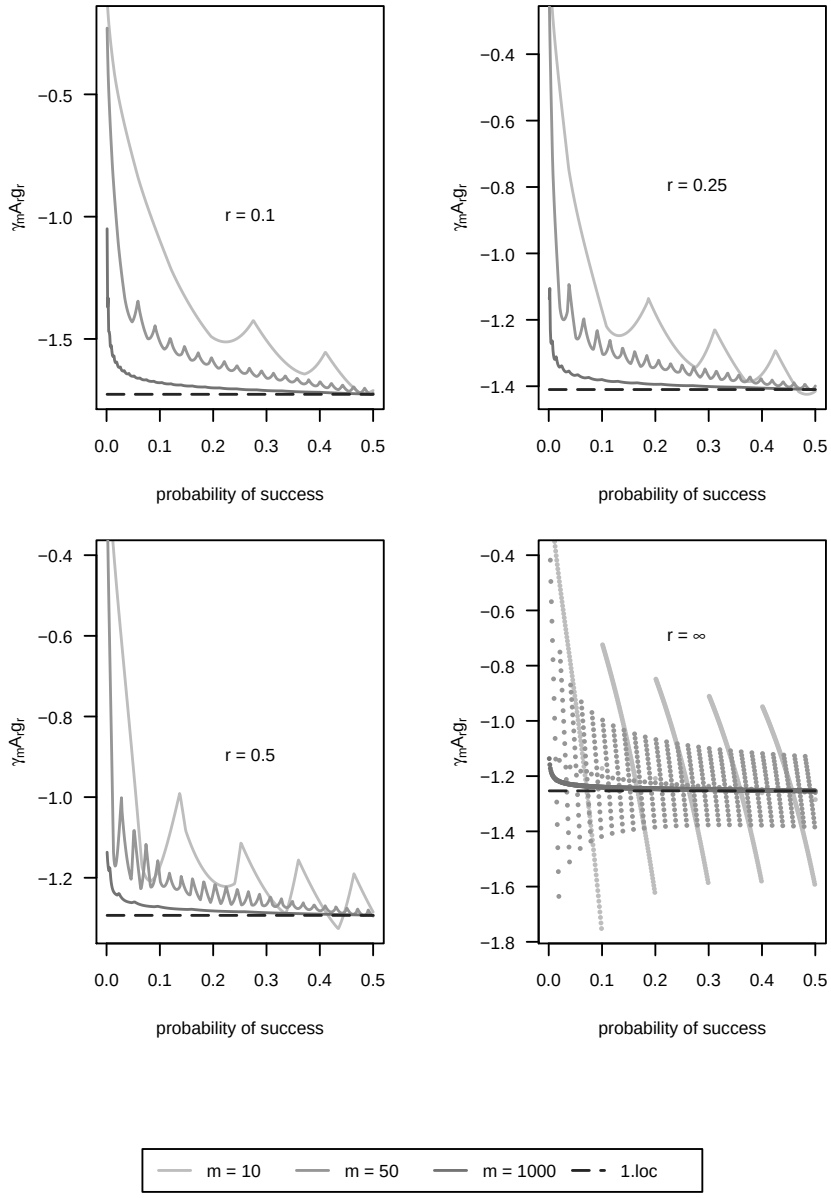

Figure 3.14: Normal approximation of the standardized lower clipping bound $\gamma_m A_r g_r$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) in case of total variation neighborhoods ($* = v$) with radius $r = 0.1, 0.25, 0.5, \infty$.

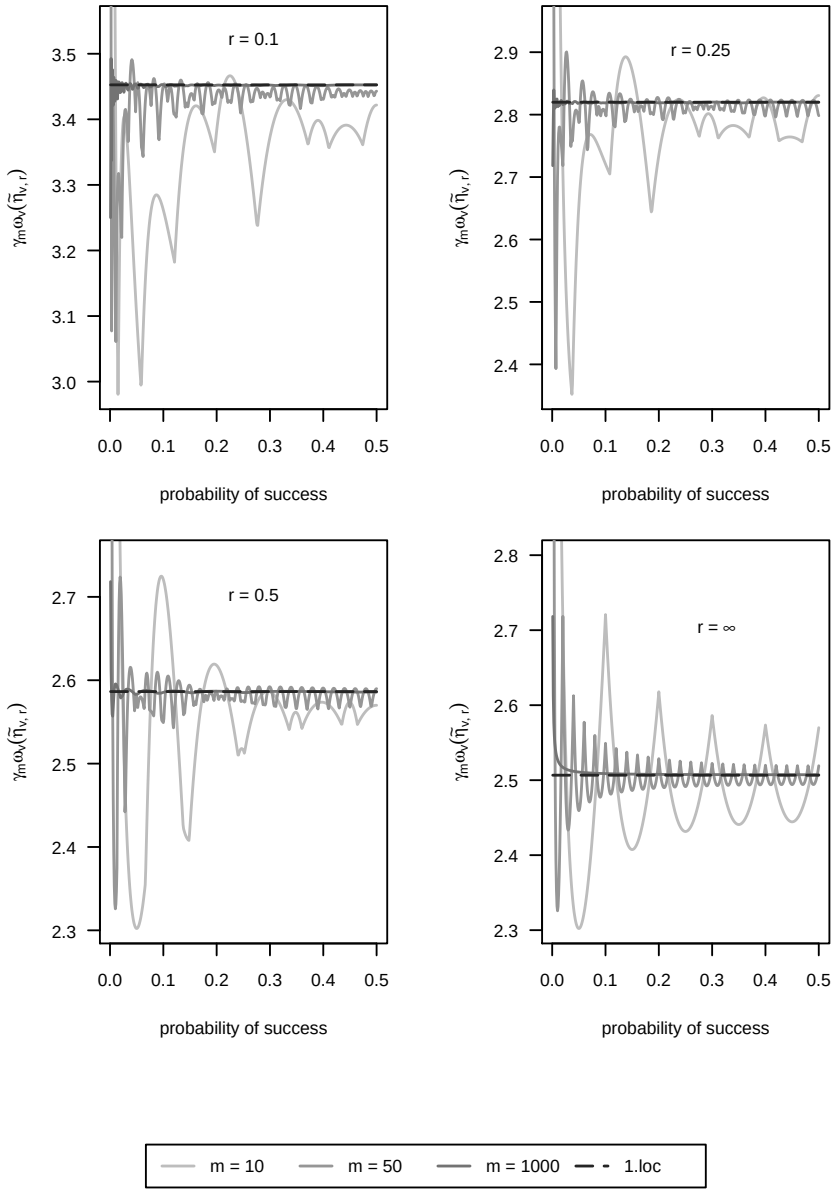

Figure 3.15: Normal approximation of the standardized infinitesimal bias terms $\gamma_m \omega_v(\tilde{\eta}_{v,r})$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) for total variation neighborhoods ($* = v$) with radius $r = 0.1, 0.25, 0.5, \infty$.

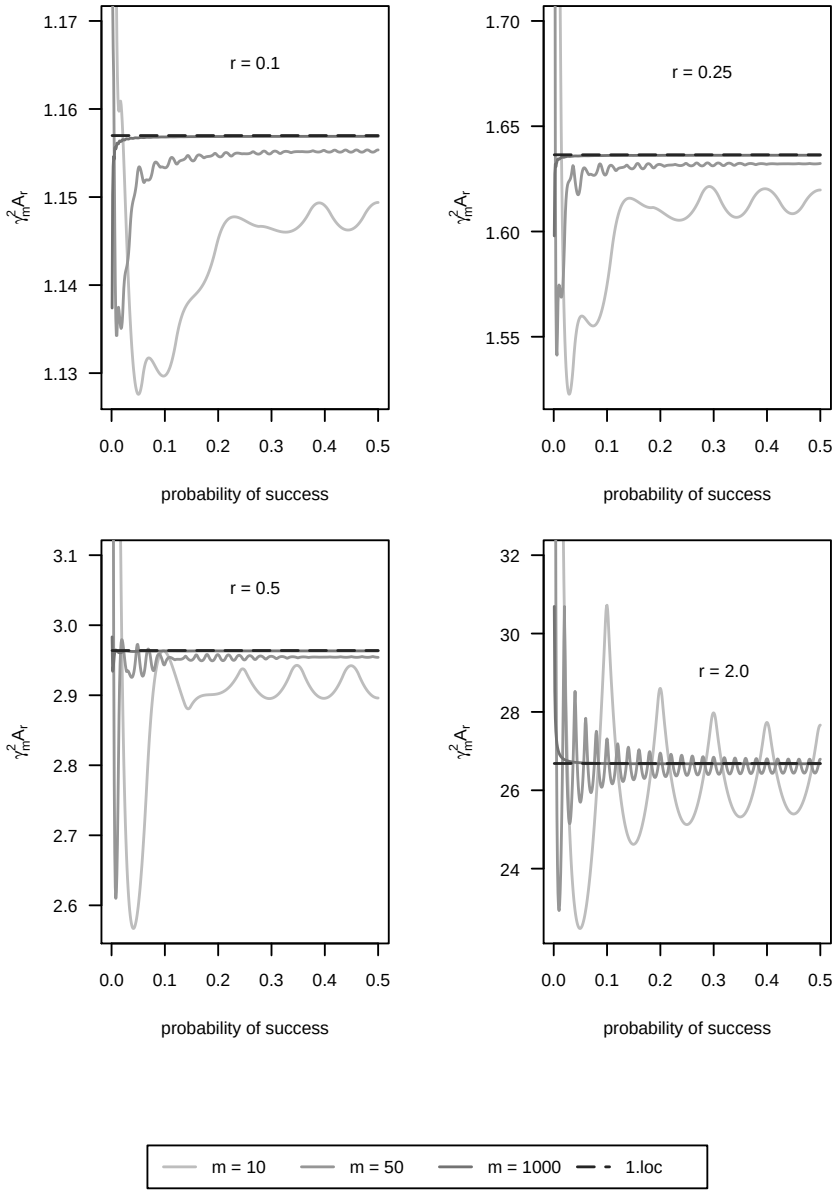

Figure 3.16: Normal approximation of the standardized maximum asymptotic MSE $\gamma_m^2 A_r$ for $\theta \in (0, 0.5]$ (results symmetric to 0.5) for total variation neighborhoods ($* = v$) with radius $r = 0.1, 0.25, 0.5, 2.0$.

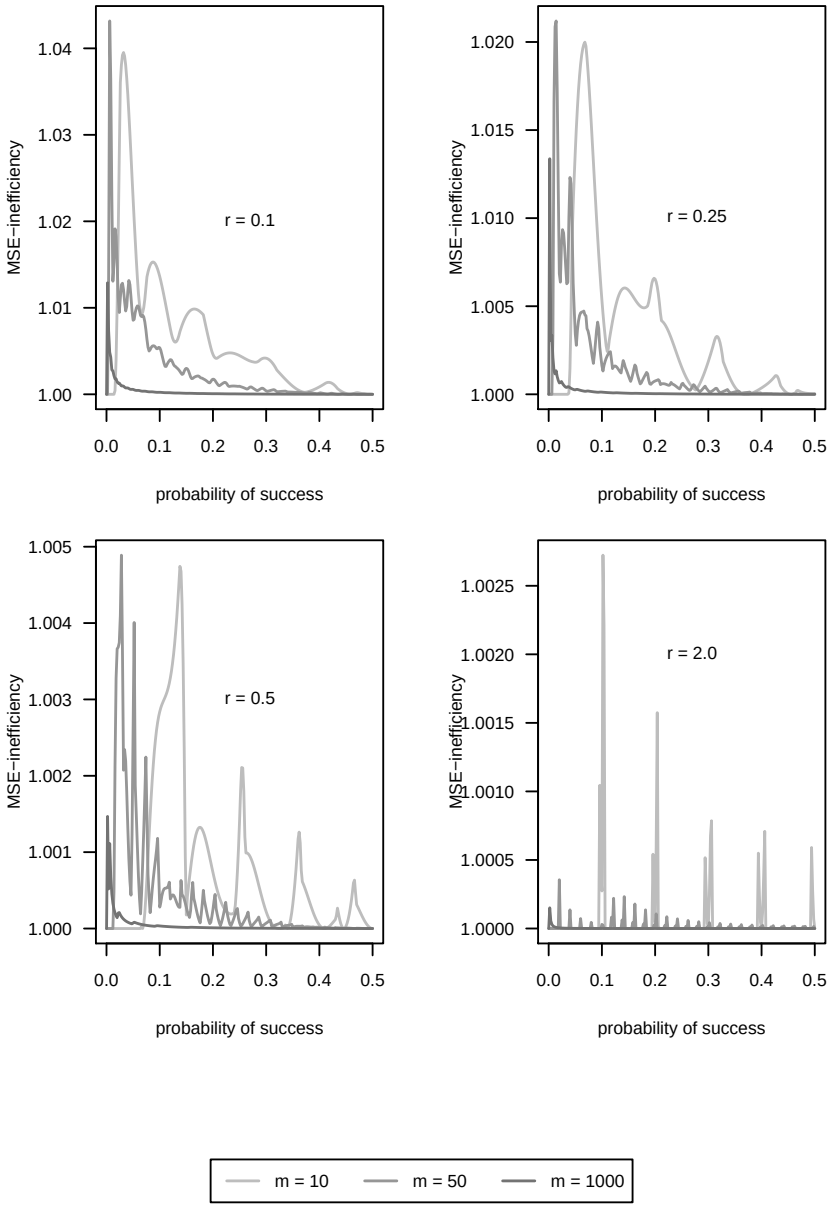

Figure 3.17: MSE–inefficiency of the normal approximated IC for $\theta \in (0, 0.5]$ (results symmetric to 0.5) in case of total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.25, 0.5, 2.0$.

3.3 Least Favorable Radius

In this section we specify the least favorable radii r_0, r_2 and r_3 for different values of $m \in \mathbb{N}$ and $\theta \in (0, 0.5]$ (results are symmetric to 0.5). For the corresponding definitions we refer to Section 2.2. In case $m = 2$, the largest value of θ which we consider is 0.499, because at $\theta = 0.5$ there exists only one IC for all $r \in [0, \infty]$ (cf. Remark 3.1.3); i.e., the inefficiency is identical to 1 for any arbitrary radius r. For $\theta \to 0$ the algorithm becomes less stable since the inefficiency curves are very flat and the least favorable radii are very small. Hence, the smallest value of θ we consider is 0.01. In addition, the inefficiency curves are also very flat near the least favorable radii. That is, the computations lead to accurate numerical approximations of the inefficiencies but are less accurate concerning the least favorable radii. Finally, in case of contamination neighborhoods the behavior at those values of θ for which the median of Λ is non-unique is different from the other values of θ. They seem to lead to the largest inefficiencies and the largest least favorable radii. However, it is very difficult to generate a grid of θ values which covers all these non-uniqueness points, respectively exact numerical approximations of these non-uniqueness points, especially if the size m is large; i.e., the plots probably contain this further inexactness in case of contamination neighborhoods. But, nevertheless, we get a good impression of the range of the least favorable radii and the corresponding inefficiencies.
Moreover, we included the corresponding inefficiencies and least favorable radii in case of one-dimensional normal location in view of the normal approximation mentioned in Remark 3.2.3 (b) and Remark 3.2.6 (b), respectively. If we use the rule of thumb $m\theta(1-\theta) \geq 9$ (i.e., $m = 50$ leads to $\theta \in [0.236, 0.764]$), the values of one-dimensional normal location provide us good approximations to the corresponding inefficiencies in case of the binomial model. In addition, as the inefficiency curves are very flat near the least favorable radii, also the corresponding least favorable radii of one-dimensional normal location may serve as acceptable approximations in case $m\theta(1-\theta) \geq 9$.

3.3.1 Contamination Neighborhoods

The minimax subefficiencies are small, respectively very small. In all considered cases ($m = 2, 5, 10, 20, 50, 1000$) they stay below about 30%, 16% and 10% in case of r_0, r_3 and r_2, respectively where the corresponding least favorable radii are small. In all these cases they lie in the interval $[0, 0.8]$. The results are plotted in Figures 3.18–3.20. But, as already noted, the considered grid of θ values does not include exact numerical approximations of all non-uniqueness points of the median of Λ. Thus, the given inefficiencies and least favorable radii at these non-uniqueness points and therefore over the whole range ($(0, 0.5]$) (resp. $(0, 1)$) of θ may be slightly larger.

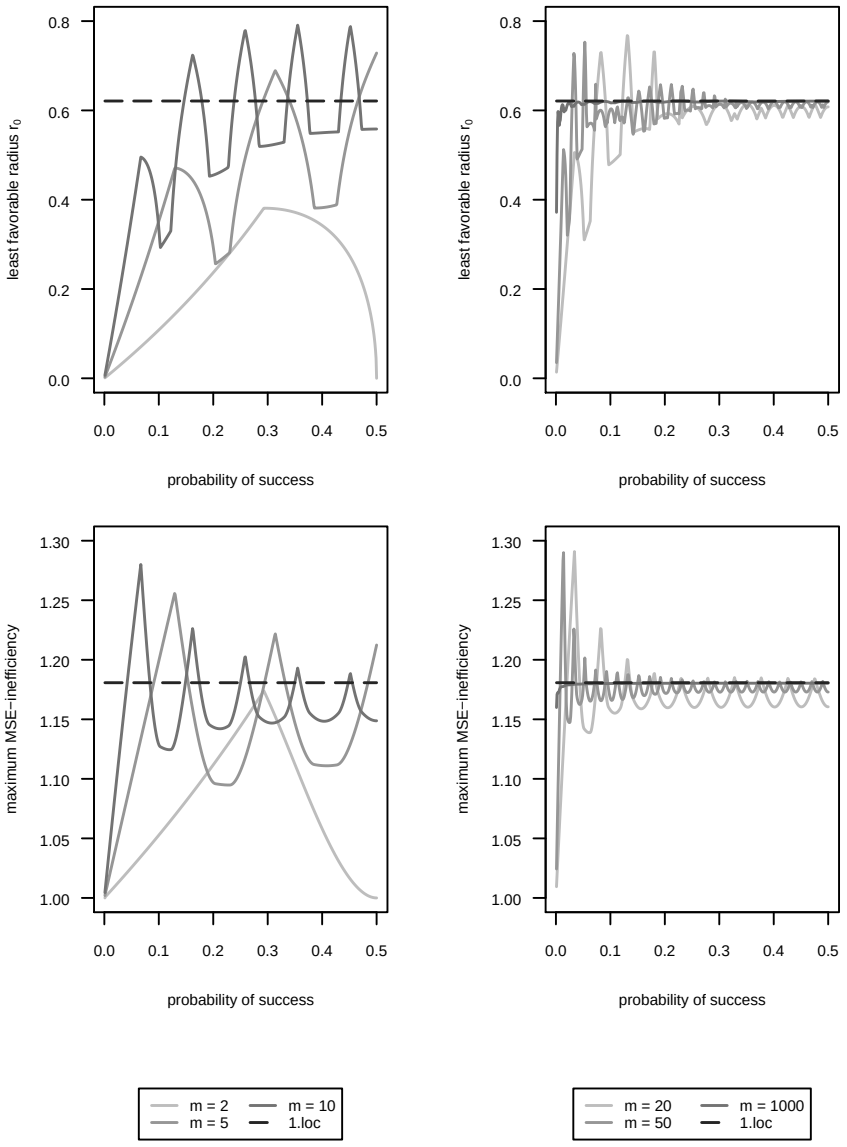

Figure 3.18: Least favorable radius r_0 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and contamination neighborhoods ($* = c$).

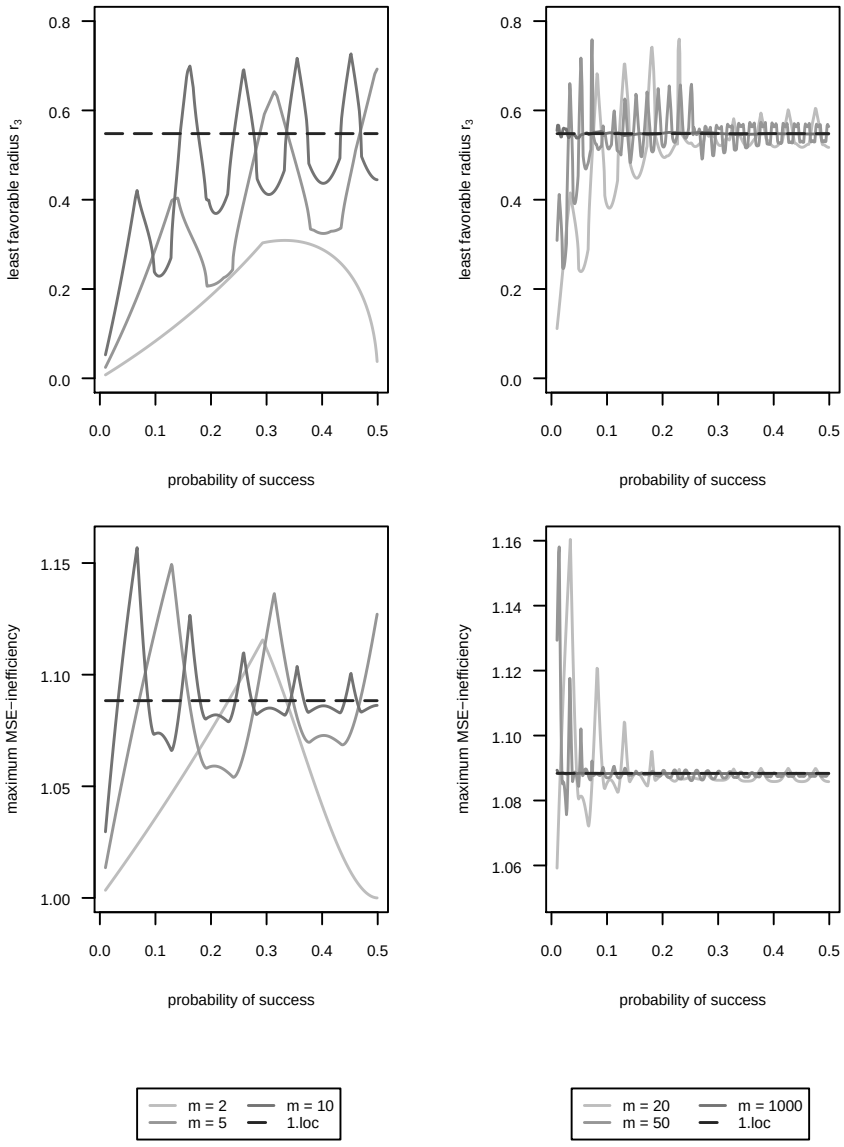

Figure 3.19: Least favorable radius r_3 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and contamination neighborhoods ($* = c$).

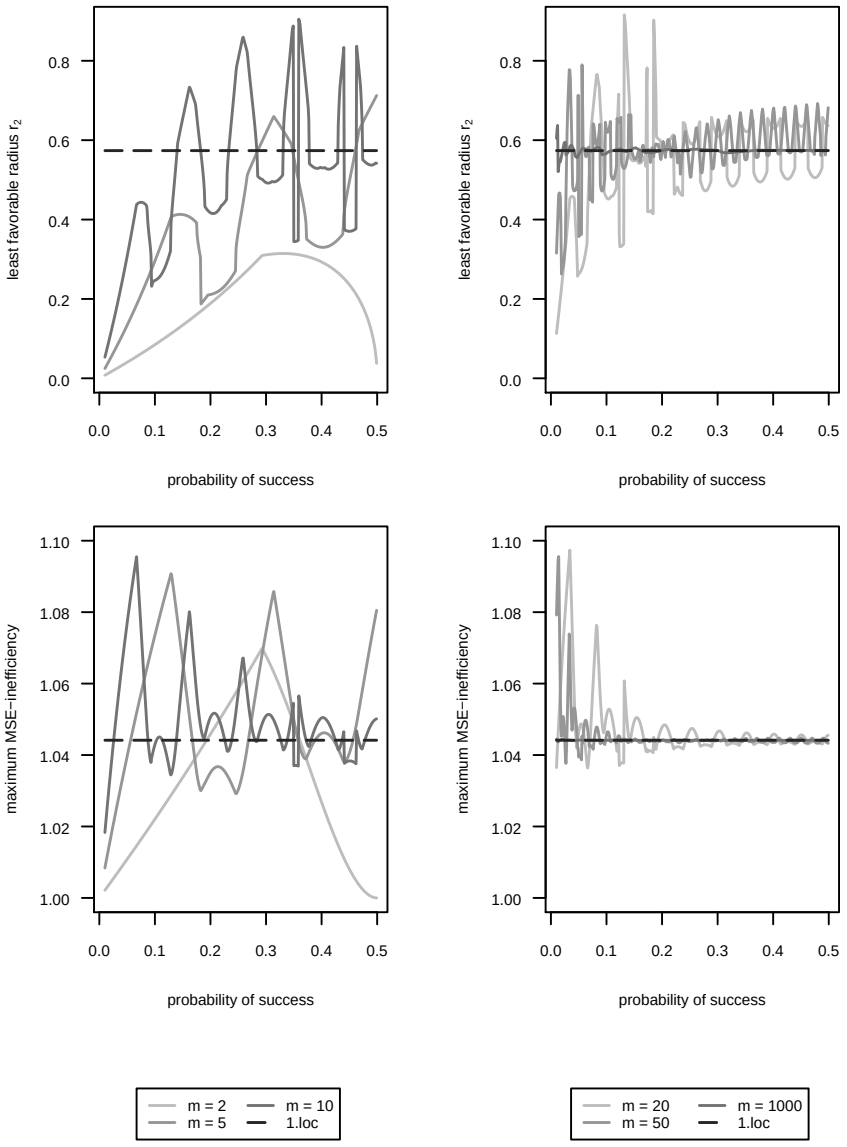

Figure 3.20: Least favorable radius r_2 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and contamination neighborhoods ($* = c$).

3.3.2 Total Variation Neighborhoods

The minimax subefficiencies are even a little bit smaller than in case of contamination neighborhoods with respect to the whole range of θ values. They stay below about 22%, 14% and 9% in case of r_0, r_3 and r_2, respectively. However, there are values of θ where the subefficiency is larger than in case of contamination neighborhoods. In particular, the least favorable radii are larger than in case of contamination neighborhoods. For the most part, they lie in the interval $[0, 1.0]$. In addition, there are some peaks near values of $\theta \in (0, 1)$ such that $m\theta \in \mathbb{N}$. In case $m = 2$ the least favorable radii show a similar behavior as the lower case radii $\bar{r}$ (cf. Figure 3.10); i.e., they get larger and larger for $\theta \to 0.5$. But, the inefficiencies for $m = 2$ and $\theta \to 0.5$ are very small. That is, in practice it does not matter which radius we choose in these cases. The results are plotted in Figures 3.21–3.23.

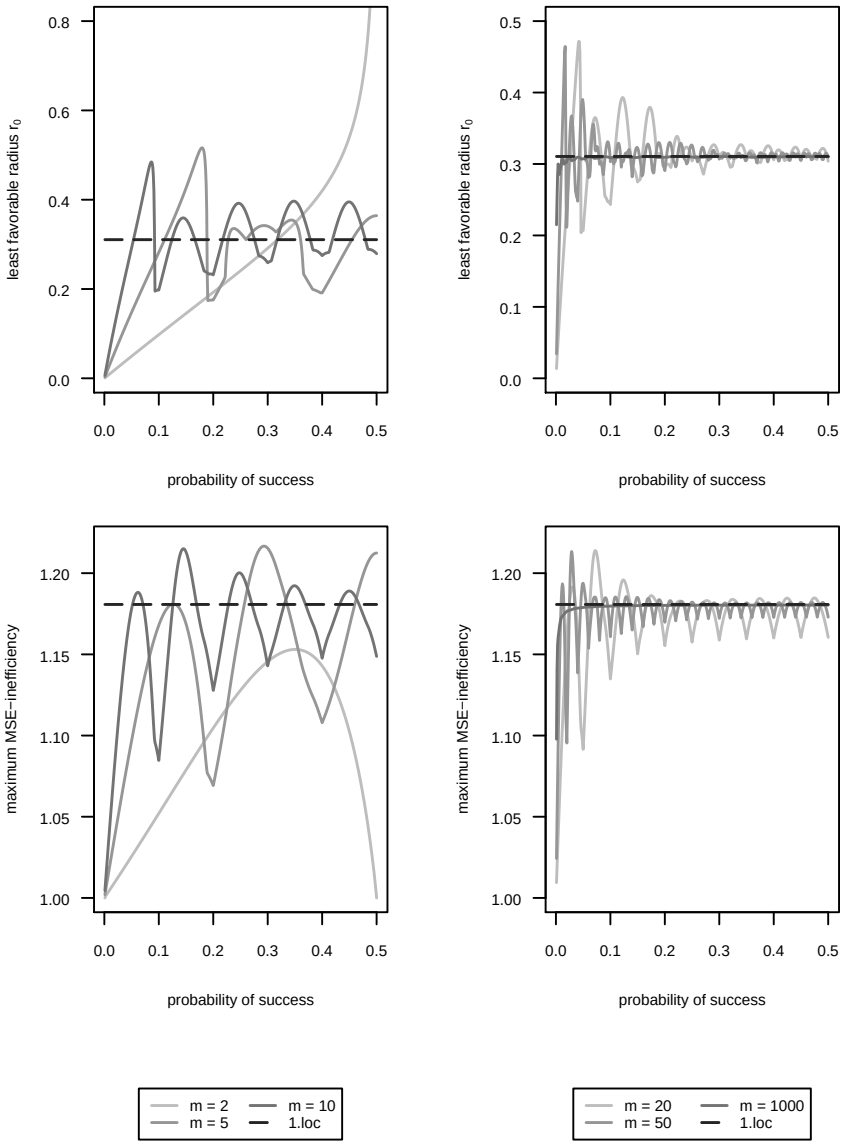

Figure 3.21: Least favorable radius r_0 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and total variation neighborhoods ($* = v$).

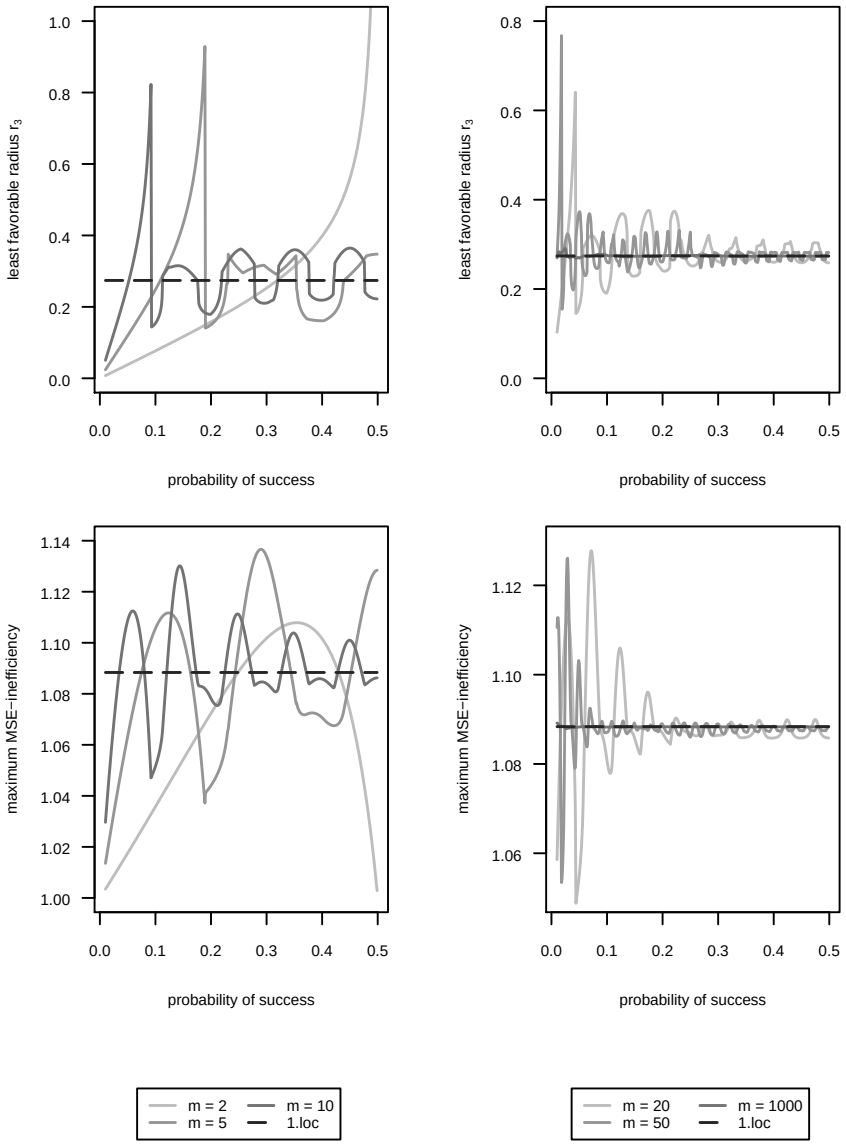

Figure 3.22: Least favorable radius r_3 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and total variation neighborhoods ($* = v$).

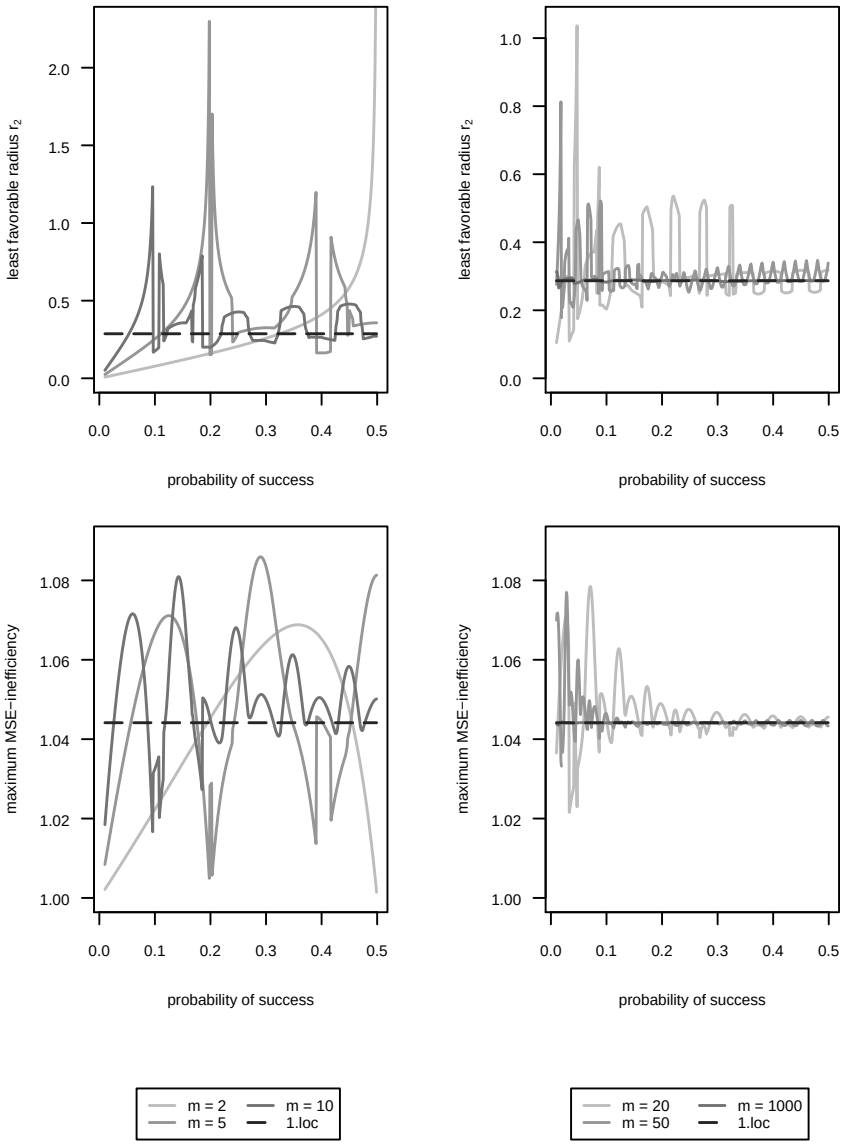

Figure 3.23: Least favorable radius r_2 and maximum MSE–inefficiency for $m = 2, 5, 10, 20, 50, 1000$ and total variation neighborhoods ($* = v$).

3.4 One-Step Construction

We can apply Lemma 2.3.6 and therefore construct the optimally robust estimator as an one-step estimator at least in case of total variation neighborhoods. In case of contamination neighborhoods the centering constant z_r may be non-unique; confer Remark 3.2.1 (c). However, this happens only if the median of Λ is non-unique and $r \geq \bar{r}$. In addition, already for very small deviations ($\pm 1e-07$ or even $\pm 1e-08$) from the exact values of θ the numerical computations yield unique centering constants. Moreover, for $\text{Binom}(m,\theta)$ there are only m values of $\theta \in (0,1)$ for which the median is non-unique, namely the roots of the following polynomial in θ

$$F_l(\theta) := \sum_{k=0}^{l} \binom{m}{k} \theta^k (1-\theta)^{m-k} - 0.5 \qquad l = 0, 1, \ldots, m-1 \tag{3.4.1}$$

If $X \sim \text{Binom}(m,\theta)$ and $Y \sim \text{Binom}(m,\theta')$ with $\theta' > \theta$, then Y is stochastically strictly larger than X; i.e., $P(Y > l) > P(X > l)$ for $l = 0, 1, \ldots, m-1$. Thus, $F_l(\theta)$ is strictly decreasing in $\theta \in (0,1)$ for each $l = 0, 1, \ldots, m-1$ with $\lim\limits_{\theta \to 0} F_l(\theta) = 0.5$ and $\lim\limits_{\theta \to 1} F_l(\theta) = -0.5$. Hence, $F_l(\theta)$ has a unique zero for each $l = 0, 1, \ldots, m-1$. Furthermore, if $\theta_l \in (0,1)$ is the zero for some l, then $1-\theta_l$ is the zero for $m-1-l$ (by symmetry). In particular, we obtain $\theta_0 = 1 - \sqrt[m]{0.5}$ ($l = 0$) and $\theta_{m-1} = \sqrt[m]{0.5}$ ($l = m-1$) are the roots of F_0 and F_{m-1} in $(0,1)$, respectively. In case m is odd, the median is also non-unique for $\theta = 0.5$ ($l = (m-1)/2$). But, in this case the distribution of Λ is symmetric to 0.5 and we can neglect the centering condition; confer Remark 3.2.1 (d).
In view of the previous considerations, the non-uniqueness of the centering constant z_r is only a minor problem for the one-step construction of the corresponding optimally robust estimators and we may neglect this non-uniqueness in practice.
Since the minimum Kolmogorov(–Smirnov) distance estimator has the necessary properties (strict and $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta) \supset \mathcal{U}_v(\theta) \supset \mathcal{U}_c(\theta)$) by Subsection 6.3.2 of Rieder (1994) and is also well computable in this simple discrete model, we use it as our starting estimator.

3.5 Implementation using R

For a detailed description of the R package ROptEst, which is part of our R bundle RobASt, including the definition of the classes and methods mentioned below we refer to Appendix D.3. To generate a member B of the binomial family with size $m \in \mathbb{N}$ and probability of success $\theta \in (0,1)$, we provide the generating function BinomFamily; i.e.,

```
> B <- BinomFamily(size = m, prob = θ)
```

One can also specify a transformation $D \in \mathbb{R}$ of the parameter θ by adding trafo=as.matrix(D) to the call of BinomFamily. The classical optimal (partial) IC (3.1.5) corresponding to B can then be computed using

```
> IC0 <- optIC(model = B, risk = asCov())
```

That is, calling the method `optIC` with an object of class `L2ParamFamily` and an object of class `asCov` returns an object `IC0` of class `IC`.
For the computation of optimally robust ICs we provide the class `InfRobModel` which in addition to an object of class `L2ParamFamily` includes an unconditional infinitesimal neighborhood. The call

```
> RobB1 <- InfRobModel(center = B,
+                      neighbor = ContNeighborhood(radius = r))
```

respectively

```
> RobB2 <- InfRobModel(center = B,
+                      neighbor=TotalVarNeighborhood(radius = r))
```

generates an instance `RobB1`, respectively `RobB2` of a binomial family with contamination, respectively total variation neighborhood and radius $r \in [0, \infty]$. In case we want to know, if the specified radius r is larger than the corresponding lower case radius $\bar{r}$, we can use

```
> lowerCaseRadius(L2Fam = B, neighbor = ContNeighborhood(),
+                 risk = asMSE())
```

respectively

```
> lowerCaseRadius(L2Fam = B, neighbor = TotalVarNeighborhood(),
+                 risk = asMSE())
```

The optimally robust IC with respect to the maximum asymptotic MSE (cf. (3.2.1), (3.2.17)) can then be computed via

```
> IC1 <- optIC(model = RobB1, risk = asMSE())
```

respectively

```
> IC2 <- optIC(model = RobB2, risk = asMSE())
```

That is, an object of class `ContIC`, respectively `TotalVarIC` is returned. In case $r = 0$, the classical optimal IC and in case $r = \infty$, the minimum bias solution (cf. (3.2.5), (3.2.21)) are computed, respectively. The minimum bias solution can also be obtained by calling

```
> IC3 <- optIC(model = RobB1, risk = asBias())
```

respectively

```
> IC4 <- optIC(model = RobB2, risk = asBias())
```

In addition, it is also possible to determine the solutions to the corresponding Hampel type problem for a given clipping bound $b \in \left(\omega_*^{\min}, \omega_*(\eta_h)\right]$ $(* = c, v)$ by using

```
> IC5 <- optIC(model = RobB1, risk = asHampel(bound = b))
```

respectively

```
> IC6 <- optIC(model = RobB2, risk = asHampel(bound = b))
```

In case $b \geq \omega_*(\eta_h)$, the classical optimal IC and in case $b \leq \omega_*^{\min}$, the minimum bias solution is returned, respectively.
If the radius is unknown, respectively unknown except to belong to some interval $[a, b]$ with $a \in [0, \infty)$ and $b \in (a, \infty]$, one can call

```
> IC7 <- radiusMinimaxIC(L2Fam = B, neighbor = ContNeighborhood(),
+                        risk = asMSE(), loRad = a, upRad = b)
```

respectively

```
> IC8 <- radiusMinimaxIC(L2Fam = B,
+                        neighbor = TotalVarNeighborhood(),
+                        risk = asMSE(), loRad = a, upRad = b)
```

which computes the IC of the AL estimator which is radius–minimax in sense of Section 2.2. Finally, one can determine the least favorable radius r_ρ ($\rho \in (0, 1)$) and the corresponding MSE–inefficiency using

```
> r.rho1 <- leastFavorableRadius(L2Fam = B,
+                                neighbor = ContNeighborhood(),
+                                risk = asMSE(), rho = ρ)
```

respectively

```
> r.rho2 <- leastFavorableRadius(L2Fam = B,
+                       neighbor = TotalVarNeighborhood(),
+                       risk = asMSE(), rho = ρ)
```

which returns a `list` with members `rho`, `leastFavorableRadius` and `ineff`. As pointed out in the following remark, the computation of the least favorable radius may take quite some time; in particular, in case of total variation neighborhoods. Hence, intermediate results are printed to bridge this time gap.

Remark 3.5.1 The convergence of the algorithm for the computation of the optimally robust ICs is clearly faster in case of contamination neighborhoods. More precisely, the computation time for the optimally robust IC is about $5-10$ times larger in case of total variation neighborhoods. For instance, in case $\text{Binom}(25, 0.25)$ and $* = c$, the computation time is about 0.25 seconds whereas in case $* = v$ it is about 1.65 seconds on an AMD Athlon with 2.5 GHz and 512 MB RAM using `R 2.0.1`; confer R Development Core Team (2005). This, of course, also affects the computation of the radius–minimax IC ($* = c$: approx. 2.1 sec. $* = v$: approx. 33 sec.) and of the least favorable radius ($* = c$: approx. 25 sec. $* = v$: approx. 8 min.). ////

Having some (contaminated) sample X of $\mathrm{Binom}(m, \theta)$ distributed data where $m \in \mathbb{N}$ is known, we can perform a robust estimation using the one-step construction. As starting estimator we provide the Kolmogorov(–Smirnov) minimum distance estimator which can be computed via

```
> est0 <- ksEstimator(x = X, distribution = Binom(size = m),
+                     param = "prob")
```

With this initial estimator we then determine the corresponding optimally robust one-step estimator. If we are sure about the neighborhood radius $r \in (0, \infty)$, we can use

```
> B0 <- BinomFamily(size = m, prob = est0)
> RobB3 <- InfRobModel(center = B0,
+                      neighbor = ContNeighborhood(radius = r))
> IC9 <- optIC(model = RobB3, risk = asMSE())
> est1 <- oneStepEstimator(x = X, IC = IC9, start = est0)
```

respectively

```
> RobB4 <- InfRobModel(center = B0,
+                      neighbor=TotalVarNeighborhood(radius = r))
> IC10 <- optIC(model = RobB4, risk = asMSE())
> est2 <- oneStepEstimator(x = X, IC = IC10, start = est0)
```

However, if the neighborhood radius is unknown except to lie in some interval $[a, b]$ ($a \in [0, \infty)$, $b \in (a, \infty]$), we instead can proceed as follows

```
> IC11 <- radiusMinimaxIC(L2Fam = B0,
+                         neighbor = ContNeighborhood(),
+                         risk = asMSE(), loRad = a, upRad = b)
> est3 <- oneStepEstimator(x = X, IC = IC11, start = est0)
```

respectively

```
> IC12 <- radiusMinimaxIC(L2Fam = B0,
+                         neighbor = TotalVarNeighborhood(),
+                         risk = asMSE(), loRad = a, upRad = b)
> est4 <- oneStepEstimator(x = X, IC = IC12, start = est0)
```

Remark 3.5.2 After installing our `R` bundle `RobASt` one can find the `R` script `BinomialModel.R`, which contains some examples for the binomial model, in the directory " . . . /`RHome`/library/`ROptEst`/scripts/" where `RHome` stands for the local home directory of `R`. ////

3.6 A Small Simulation Study

To illustrate the use of our `R` package `ROptEst` (cf. Appendix D.3) and the need of robust estimation in case of the binomial model, we conclude this chapter with the presentation of a small simulation study.
We consider three different situations. First, the data is generated by the ideal model $\mathrm{Binom}(25, 0.25)$, we then replace 2% of the data by $\mathrm{Binom}(25, 0.75)$; i.e., we consider the (realistic) gross error model

$$0.98\,\mathrm{Binom}(25, 0.25) + 0.02\,\mathrm{Binom}(25, 0.75) \tag{3.6.1}$$

Hence, we could say that in 2% of the considered cases the failures were noted instead of the successes – a rather realistic situation; confer Subsection 1.2c of Hampel et al. (1986) for the frequency of gross errors in real data sets. In the third situation, we replace the same 2% of the data by $\mathrm{I}_{\{25\}}$; i.e., we also study the ("extreme") gross error model

$$0.98\,\mathrm{Binom}(25, 0.25) + 0.02\,\mathrm{I}_{\{25\}} \tag{3.6.2}$$

We computed 1000 samples of size 100 and determined the mean (classical optimal) and the Kolmogorov(–Smirnov) minimum distance (ksMD) estimator. The ksMD estimator then serves as initial estimator for two robust one-step estimators. The first one-step estimator ($r = 0.2$) is based on the optimally robust IC for radius $r = 0.2$; i.e., the amount of contamination is known ($r/\sqrt{100} = 0.02$). The second one-step estimator ($r = r_0$) was calculated using the radius–minimax IC for the least favorable radius r_0; i.e., the radius is completely unknown. For a boxplot of the results see Figure 3.24.

Remark 3.6.1 If we only admit contaminating distributions which are also concentrated on $\{0, 1, \ldots, m\}$, any value in $\mathbb{R} \setminus \{0, 1, \ldots, m\}$ is identified as outlier with probability 1 and could therefore be omitted from the sample. Thus, in this situation contaminating with $\mathrm{I}_{\{m\}}$ if $\theta < 0.5$, respectively $\mathrm{I}_{\{0\}}$ if $\theta > 0.5$ heuristically seems to have the largest effect on the estimation of θ — at least for the mean. As our results indicate, this is not necessarily true for the ksMD estimator; confer Tables 3.1 and 3.2. In addition, in case of the Poisson model (cf. Section 4.6) the ksMD estimator performs even better in the "extreme" than in the realistic situation and hence, this probably also holds in case of the binomial model for larger values of m. Moreover, the robust estimators which use the ksMD estimator as initial estimator show a similar behavior. ////

As we are dealing with the estimation of a parameter that has a very limited range ($\theta \in (0, 1)$), there is no large difference in absolute values between the empirical MSEs as well as the asymptotic MSEs of the considered estimators in any case. The results are given in Table 3.1 where we also provide 95% confidence intervals based on the central limit theorem.

However, with respect to MSE–inefficiency we see clear differences between the considered estimators. In the ideal case, the subefficiency of the ksMD and the first

one-step estimator ($r = 0.2$) with respect to the mean is clearly below 10% and also the second one-step estimator ($r = r_0$) loses only about 20% efficiency; confer Table 3.2. Somewhat surprisingly, the first one-step estimator ($r = 0.2$) performs even better than the ksMD estimator although we are in the ideal model.
In the contaminated samples the ksMD estimator ($\approx 18\%$) and, even more, the radius–minimax estimator ($\approx 5\%$) do not lose much efficiency. The mean, however,

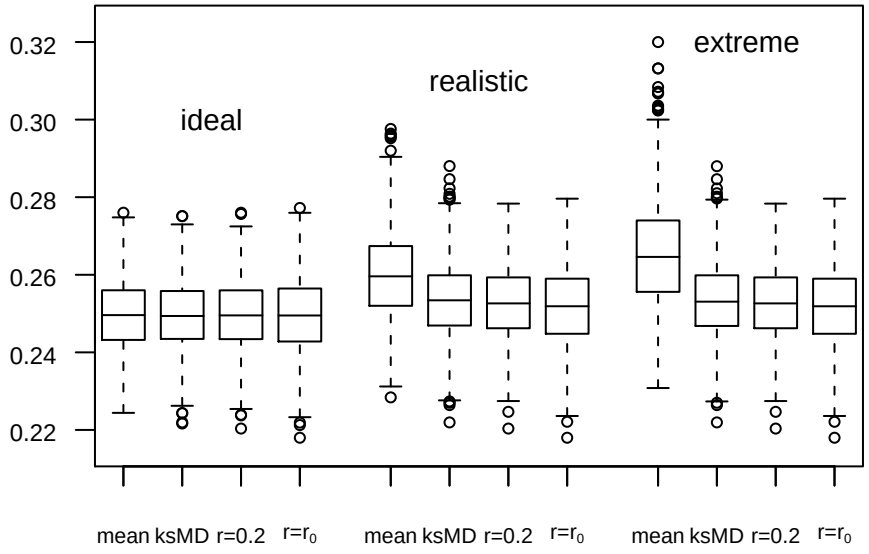

Figure 3.24: Boxplot for a small simulation study in case of contamination neighborhoods ($* = c$).

situation	mean	ksMD	$r = r_0$	$r = 0.2$
emp. ideal	0.0078 ± 0.0021	0.0085 ± 0.0023	0.0093 ± 0.0024	0.0082 ± 0.0021
as. ideal	0.0075	—	0.0087	0.0078
realistic	0.0230 ± 0.0061	0.0114 ± 0.0032	0.0102 ± 0.0026	0.0097 ± 0.0025
extreme	0.0426 ± 0.0111	0.0114 ± 0.0032	0.0102 ± 0.0026	0.0097 ± 0.0025
asympt.	0.0300	—	0.0093	0.0087

Table 3.1: Empirical and asymptotic MSEs for a small simulation study in case of contamination neighborhoods ($* = c$).

situation	mean	ksMD	$r = r_0$	$r = 0.2$
emp. ideal	1.000	1.090	1.192	1.051
as. ideal	1.000	—	1.160	1.040
realistic	2.371	1.175	1.052	1.000
extreme	4.392	1.175	1.052	1.000
asymptotic	3.448	—	1.069	1.000

Table 3.2: MSE–inefficiencies for a small simulation study in case of contamination neighborhoods ($* = c$).

is strongly affected by the contamination in any case and already in the rather harmless realistic situation has a subefficiency of about 140%. This efficiency loss increases up to about 340% (!) in the "extreme" case. Thus, even if the amount of contamination is small and the contaminating distribution is rather harmless, the mean cannot be regarded as an appropriate estimator. In contrast, the robust estimators do not lose much efficiency in the ideal case and also the computational effort is not much larger compared to the mean. Therefore, we recommend to use the robust estimators in any case.
Furthermore, we see that the empirical MSE and the empirical MSE–inefficiency are already in good agreement with the corresponding asymptotic values – at least in case of the robust estimators. However, the empirical values and probably also the exact finite-sample values appear to be rather larger than the asymptotic values especially in case of the mean. This points in the same direction as the results which we obtain in Part V of this thesis and in Ruckdeschel and Kohl (2005) as well as with the results of the higher order studies of Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c) and Ruckdeschel (2005e).

Remark 3.6.2 **(a)** We do not give the asymptotic values for the ksMD estimator as we are not sure about its asymptotic distribution in this setup.

(b) In view of this small simulation study, it might be of interest to take a closer look at the finite-sample and asymptotic behavior of the Kolmogorov(–Smirnov) minimum distance estimator. In addition, one could perhaps investigate how much a one-step estimator is influenced by the initial estimator and if it is worth to use two- or even k-step estimators ($k > 2$). The higher order studies for one-step estimators in Ruckdeschel (2005e) indeed confirm that one can gain some efficiency by using k-step estimators.

(c) Of course, a similar simulation study can be made in case of total variation neighborhoods (cf. also Section 4.6) where the "extreme" situation is of the form

$$k \mapsto \left(\mathrm{Binom}\,(m,\theta)(\{k\}) - r/\sqrt{n}\right) \vee 0 + r/\sqrt{n}\ \mathrm{I}_{\{m\}}(k) \qquad \text{if } \theta < 0.5 \qquad (3.6.3)$$

respectively

$$k \mapsto \left(\mathrm{Binom}\,(m,\theta)(\{k\}) + r/\sqrt{n}\right) \wedge 1 + r/\sqrt{n}\ \mathrm{I}_{\{0\}}(k) \qquad \text{if } \theta > 0.5 \qquad (3.6.4)$$

similar to Rieder (1994), p 175. In particular, a larger simulation study may be of interest to compare empirical, respectively finite-sample and asymptotic results for different (small and medium) sample sizes and to investigate if there is a difference concerning the speed of convergence towards the asymptotic values between contamination and total variation neighborhoods as encountered in Part V. ////

Chapter 4

Poisson Model

This chapter contains various results about the Poisson model which is defined in Section 4.1. In Section 4.2 we specialize the solutions to the asymptotic MSE problem (1.3.5) given in Theorem 1.3.11 to this model and investigate the continuity and the smoothness of the Lagrange multipliers contained in the optimally robust ICs. We then derive a normal approximation for these multipliers and show that they can also be approximated by the Lagrange multipliers arising in the optimally robust ICs in case of the binomial model. In the subsequent section (Section 4.3) we present numerical results for the least favorable radii and the corresponding MSE–inefficiencies in case of the Poisson model. Finally, we verify that we can construct the optimally robust estimator by means of the one-step construction (cf. Section 4.4) and describe how one can use our `R` package `ROptEst` (cf. Section 4.5). We conclude this chapter with a small simulation study to demonstrate the use of robust estimation in case of the Poisson model; confer Section 4.6.

4.1 Introduction

The Poisson model with unknown mean is,

$$\mathcal{P} = \left\{ \mathrm{Pois}(\theta) \,\middle|\, \theta \in (0,\infty) \right\} \tag{4.1.1}$$

where

$$\mathrm{Pois}(\theta)(\{y\}) = \frac{\theta^y}{y!}\,\exp(-\theta) \qquad y \in \mathbb{N}_0 \tag{4.1.2}$$

Remark 4.1.1 The Poisson model (4.1.1) forms an exponential family with respect to the counting measure on $\mathbb{N}_0$, since

$$\frac{\theta^y}{y!}\,\exp(-\theta) = \frac{1}{y!}\exp\{y\log\theta - \theta\} \tag{4.1.3}$$

confer also Example 1.5.12 of Lehmann and Casella (1998). With the notation of Lemma 2.3.6 we obtain, $\zeta(\theta) = \log\theta$, $\beta(\theta) = \theta$, $T(y) = y$ and $h(y) = (y!)^{-1}$ which leads to $\mathcal{J}_\zeta = \theta^{-1}$, $\mathrm{E}_\theta\, T = \theta$ and $\mathrm{Var}_\theta\, T = \theta$. ////

Lemma 4.1.2 *The Poisson model (4.1.1) is L_2 differentiable at $\theta \in (0,\infty)$ with L_2 derivative Λ_θ and Fisher information!Poisson model $\mathcal{I}_\theta$ given by*

$$\Lambda_\theta(y) = \theta^{-1}y - 1 \qquad \mathcal{I}_\theta = \theta^{-1} \tag{4.1.4}$$

PROOF A consequence of Lemma 2.3.6 (a) in connection with Remark 4.1.1. ////

Remark 4.1.3 (a) There is quite a number of examples where a Poisson distribution fits well like radioactive disintegrations, chromosome interchanges in cells, bacteria and blood counts or catches in fishery; confer Section VI.7 of Feller (1968) and Section 1 of Cadigan and Chen (2001). Hence, robust estimation in the Poisson model is an important topic. This is also confirmed by the small simulation study which is presented at the end of this chapter (cf. Section 4.6).

(b) Similarly to the binomial model (cf. Remark 3.1.3 (c)), we do not consider robust estimation in Poisson generalized linear models. For an example we refer to Rieder (1996). ////

4.2 Optimally Robust Influence Curves

4.2.1 Contamination Neighborhoods

4.2.1.1 Mean Square Error Solution

For some given $D \in \mathbb{R} \setminus \{0\}$ we can rewrite the unique MSE optimal IC $\tilde{\eta}_{c,r}$ for infinitesimal contamination neighborhoods (1.2.4) and radius $r \in (0,\infty)$ supplied by Theorem 1.3.7 (a) and Theorem 1.3.11 (b) as

$$\tilde{\eta}_{c,r}(y) = A_r(\Lambda(y) - z_r)\min\left\{1, \frac{c_r}{|\Lambda(y) - z_r|}\right\} \tag{4.2.1}$$

where

$$0 = \mathrm{E}(\Lambda - z_r)\min\left\{1, \frac{c_r}{|\Lambda - z_r|}\right\} \tag{4.2.2}$$

$$D = A_r\,\mathrm{E}\,|\Lambda - z_r|\min\left\{|\Lambda - z_r|, c_r\right\} \tag{4.2.3}$$

and

$$r^2\,c_r = \mathrm{E}\left(|\Lambda - z_r| - c_r\right)_+ \tag{4.2.4}$$

For $r = \infty$ Theorem 1.3.7 (b) yields

$$\tilde{\eta}_{c,\infty}(y) = \omega_c^{\min}\,\mathrm{sign}(D)\Big[\mathrm{I}(y > M) - \mathrm{I}(y < M) + \beta\,\mathrm{I}(y = M)\Big] \tag{4.2.5}$$

where

$$\beta = \Big[P(y < M) - P(y > M)\Big]\Big/P(y = M) \tag{4.2.6}$$

with any $M = \mathrm{med}(y)$ and $\tilde{\eta}_{c,\infty}$ achieves the minimum bias

$$\omega_c^{\min} = \frac{|D|\theta}{\mathrm{E}\,|y - M|} \tag{4.2.7}$$

confer also Remark 1.3.8.
For a plot of the optimally robust ICs in case $\theta = 5$ and for different values of r see Figure 4.1.

Remark 4.2.1 **(a)** Like in the binomial model, it might be necessary to extend the optimally robust ICs to $\mathbb{R} \setminus \mathbb{N}_0$. For more details see Remark 3.2.1.

(b) Since $\mathrm{med}(\Lambda)$ is non-unique for some $\theta \in (0, \infty)$ and the assumptions of Proposition 2.1.3 hold, we get z_r is non-unique in case $r \geq \bar{r}$ $(\bar{r} < \infty)$ where the lower case radius $\bar{r}$ is defined in (2.1.12). For a plot of the lower case radius $\bar{r}$ for $\theta \in (0, 10]$ confer Figure 4.2. The upper peaks correspond to values of θ for which the median of Λ is non-unique.

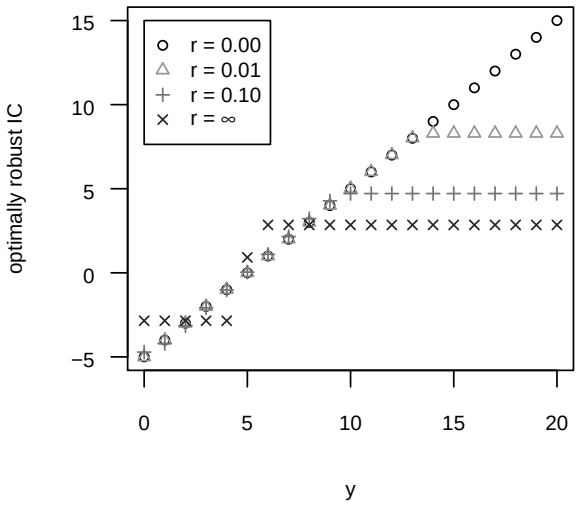

Figure 4.1: Optimally Robust ICs for $\mathrm{Pois}(5)$ in case of contamination neighborhoods with radius $r = 0, 0.01, 0.10, \infty$.

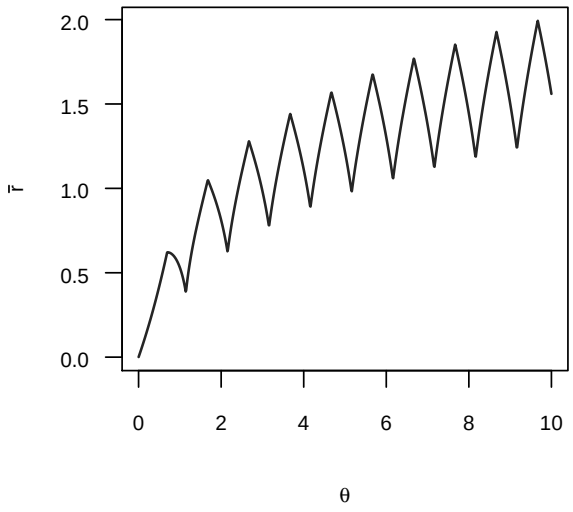

Figure 4.2: Lower case radius $\bar{r}$ for $\theta \in (0, 10]$ in case of contamination neighborhoods.

4.2.1.2 Continuity and Uniqueness of Lagrange Multipliers

The continuity of the Lagrange multipliers A_r, and b_r in r, stated in Proposition 2.1.9, is visualized in Figure 4.3. Since $\text{med}(\Lambda)$ is non-unique for $\theta \approx 3.67206$, there is a whole interval of valid centering constants a_r for $r \geq \bar{r} \approx 1.443$. The boundaries of this interval can be determined via (2.1.17) and are given in Figure 4.3. In contrast to the standardized asymptotic bias b_r and the asymptotic variance $A_r - r^2 b_r^2$, which seem to be non-differentiable at some values of r, the maximum asymptotic MSE A_r seems to be very smooth in r.
Similarly to the binomial model the centering constant $a_r = A_r z_r$ can have discontinuity points coinciding with those values of θ for which the median of Λ is non-unique. This fact is illustrated in Figure 4.4 where we choose a relatively large radius ($r = 2.0$) to demonstrate the extreme case. In addition, we again see that the Lagrange multipliers and hence the standardized asymptotic bias, the asymptotic variance and the maximum asymptotic MSE are continuous (cf. Theorem 2.1.11) but are not necessarily smooth functions in the parameter θ. ////

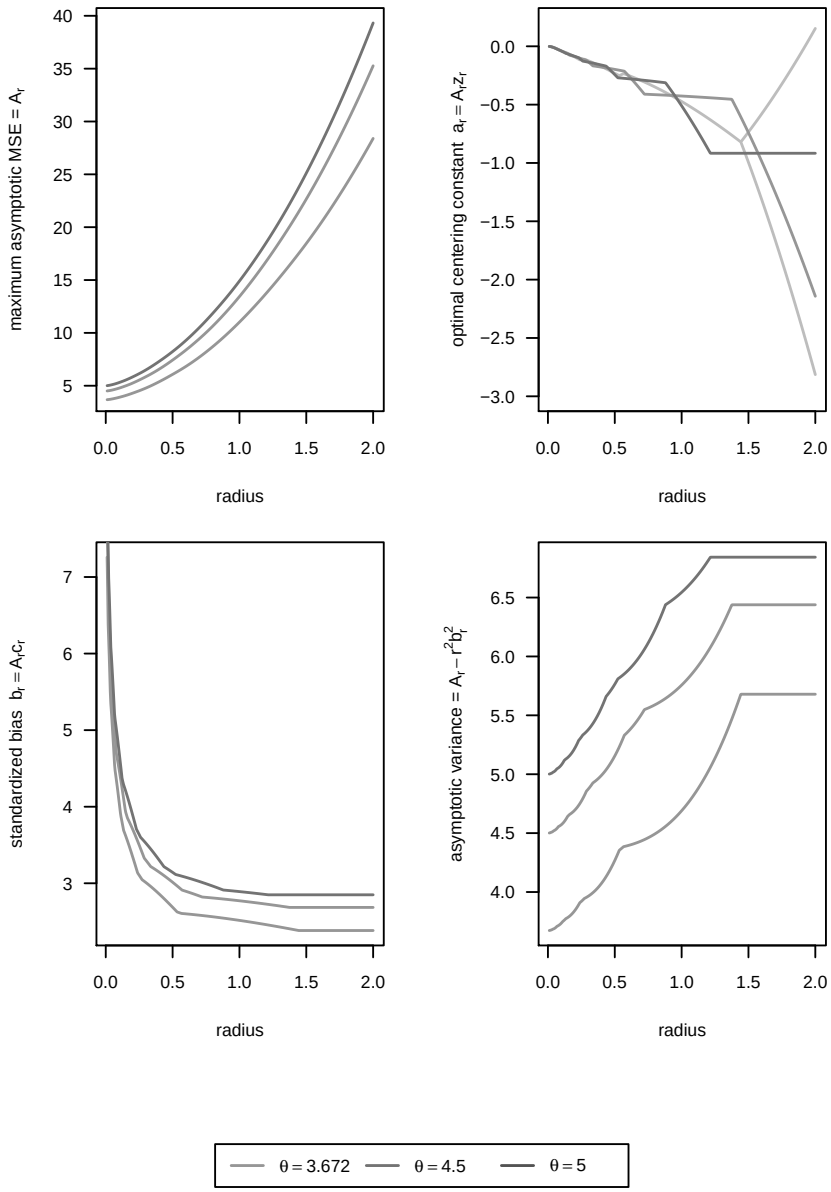

Figure 4.3: Continuity in the radius r of the Lagrange multipliers contained in the MSE optimal ICs for $r \in (0, 2.0]$ and $\theta = 3.672, 4.5, 5.0$ in case of contamination neighborhoods $(* = c)$.

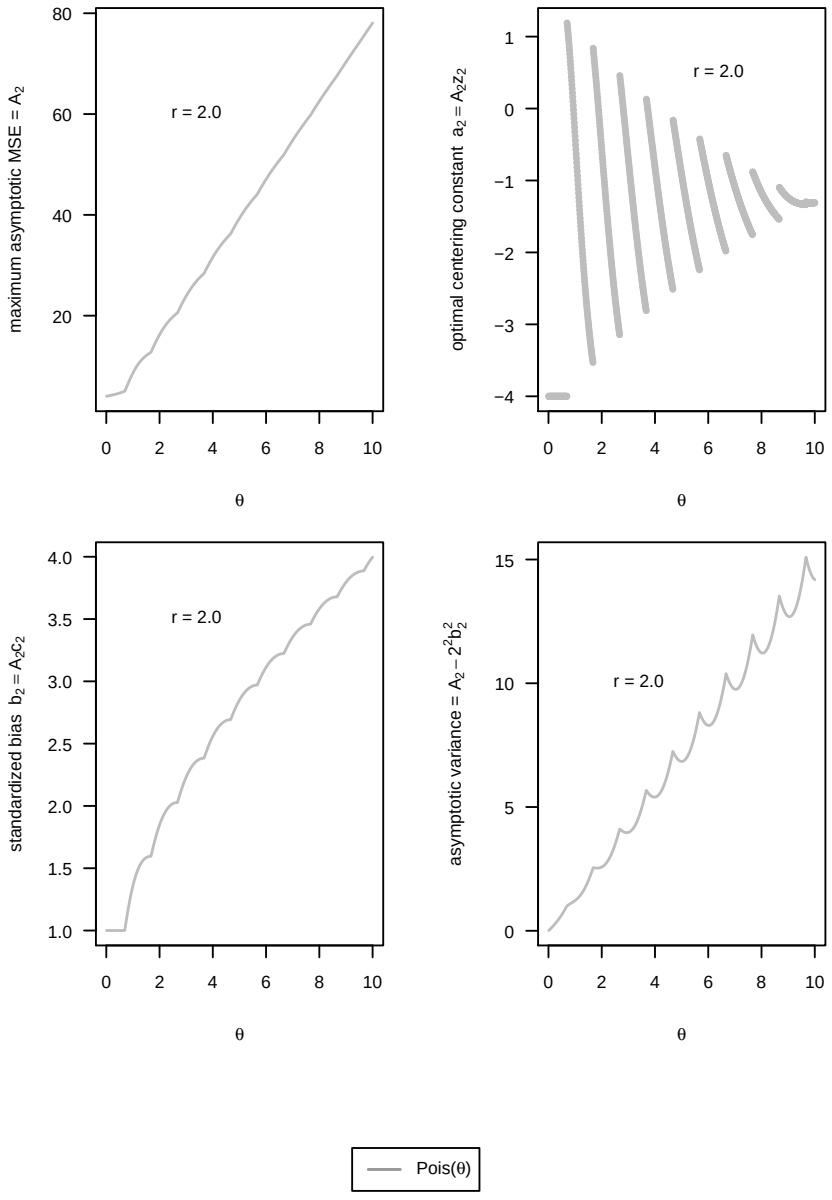

Figure 4.4: Continuity in the parameter θ of the Lagrange multipliers contained in the MSE optimal ICs for $\theta \in (0, 10]$ in case of contamination neighborhoods $(* = c)$ with radius $r = 2.0$.

4.2.1.3 Normal Approximation

Similar to the binomial model there is a normal approximation for the Lagrange multipliers contained in the optimally robust ICs where the corresponding optimally robust IC $\tilde{\eta}_{c,r}^{1.\mathrm{loc}}$ in case of one-dimensional normal location is specified in (3.2.9)–(3.2.11). The lower case solution can be read off from (3.2.12) and attains minimum bias $\omega_c^{\min,\,1.\mathrm{loc}} = \sqrt{\frac{\pi}{2}}$.

Lemma 4.2.2 *Let* $D = 1$. *It holds,*

$$\lim_{\theta\to\infty} \theta^{-1} A_r = A_r^{1.\mathrm{loc}} \qquad \lim_{\theta\to\infty} \sqrt{\theta}\, z_r = z_r^{1.\mathrm{loc}} = 0 \qquad \lim_{\theta\to\infty} \sqrt{\theta}\, c_r = c_r^{1.\mathrm{loc}} \tag{4.2.8}$$

for all $r \in (0,\infty)$ *and*

$$\lim_{\theta\to\infty} \theta^{-1/2} \omega_c^{\min} = \omega_c^{\min,\,1.\mathrm{loc}} \tag{4.2.9}$$

Proof We have

$$\sqrt{\theta}\,\Lambda(y) = \theta^{-1/2}(y-\theta) \tag{4.2.10}$$

By the convolution property of the Poisson distribution we can regard $\mathrm{Pois}\,(\theta)$ as the N-fold convolution of $\mathrm{Pois}\,(\theta/N)$. Hence, by the central limit theorem of Lindeberg-Lévy $\mathcal{L}(\sqrt{\theta}\,\Lambda) \xrightarrow{\mathrm{w}} \mathcal{N}(0,1)$ as $\theta \to \infty$ where $\mathrm{E}\sqrt{\theta}\,\Lambda = 0$ and $\mathrm{E}\,(\sqrt{\theta}\,\Lambda)^2 = 1$ for all $\theta \in (0,\infty)$. Thus, we can apply Theorem 2.4.1 and obtain (4.2.8) and (4.2.9). ////

Remark 4.2.3 **(a)** The convergence of the standardized optimal clipping bounds $\sqrt{\theta}\, z_r$, respectively of the standardized bias terms $\theta^{-1/2}\omega_c(\tilde{\eta}_{c,r})$ are illustrated in Figure 4.5 and Figure 4.6, respectively. In case $r = \infty$, the discontinuity points of the centering constant z_r coincide with those values of θ for which the median of Λ is non-unique. At these non-uniqueness points the standardized infinitesimal bias terms attain local minima in case $r = \infty$. These results are analogously to the binomial model; confer Remark 3.2.3 (b).

(b) The convergence of the standardized maximum asymptotic MSE, respectively the MSE–inefficiencies is visualized in Figure 4.7 and Subsection 4.3.1, respectively.

(c) Figure 4.8 shows the MSE–inefficiency of the normal approximated IC. That is, we used the Lagrange multipliers $A_r^{1.\mathrm{loc}}\theta$, $z_r^{1.\mathrm{loc}}/\sqrt{\theta}$ and $c_r^{1.\mathrm{loc}}/\sqrt{\theta}$ instead of the optimal A_r, z_r and c_r; confer Lemma 4.2.2. To make sure that the resulting function is indeed an IC (with respect to the Poisson model), we additionally centered and standardized this function. The results are similar to the binomial model (cf. Subsubsection 3.2.1.3). For small to moderate radii we get very small, respectively small efficiency losses. However, for large radii ($r > \bar{r}$) the subefficiency can become quite large where we get the smallest MSE–inefficiencies for $\theta \le \log(2)$ and for such values of θ at which $\mathrm{med}(\Lambda)$ is non-unique. A possible explanation for this behavior is given in Remark 3.2.3 (c). ////

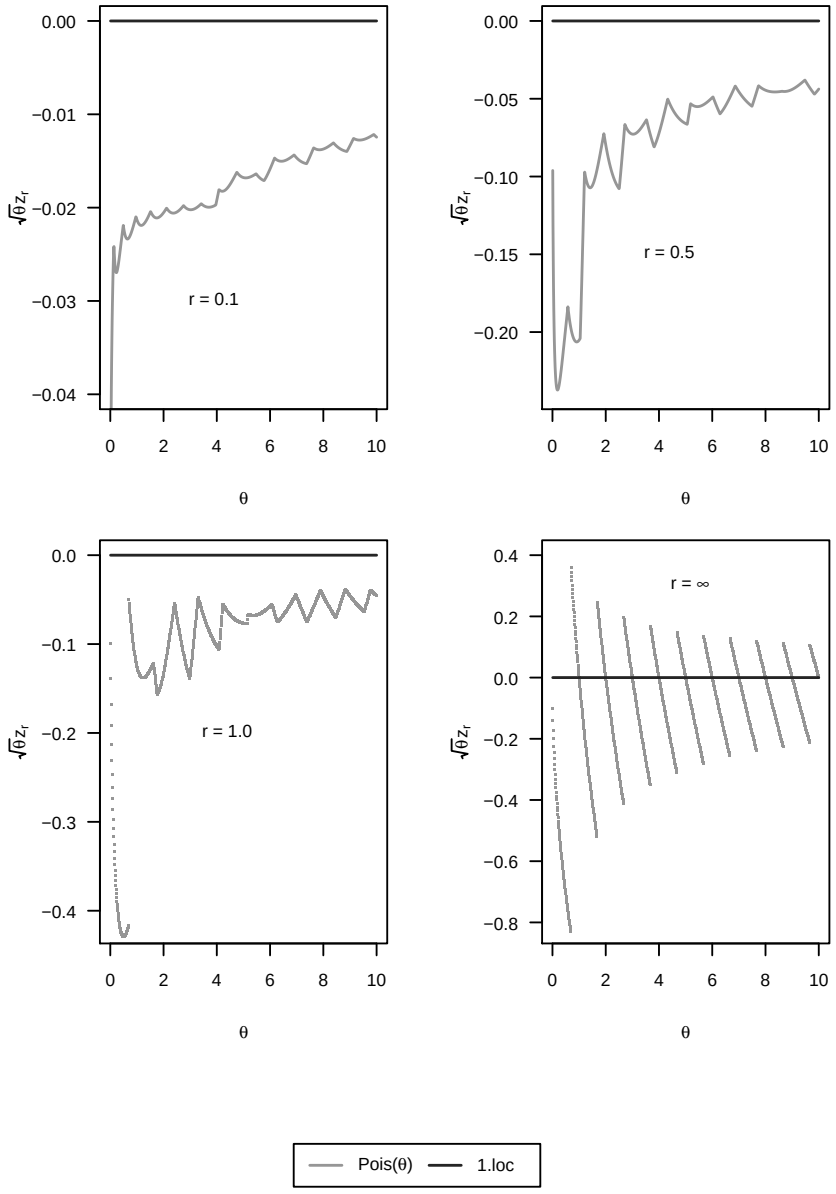

Figure 4.5: Normal approximation of the standardized centering constant $\sqrt{\theta}\, z_r$ for $\theta \in (0, 10]$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.25, 0.5, \infty$.

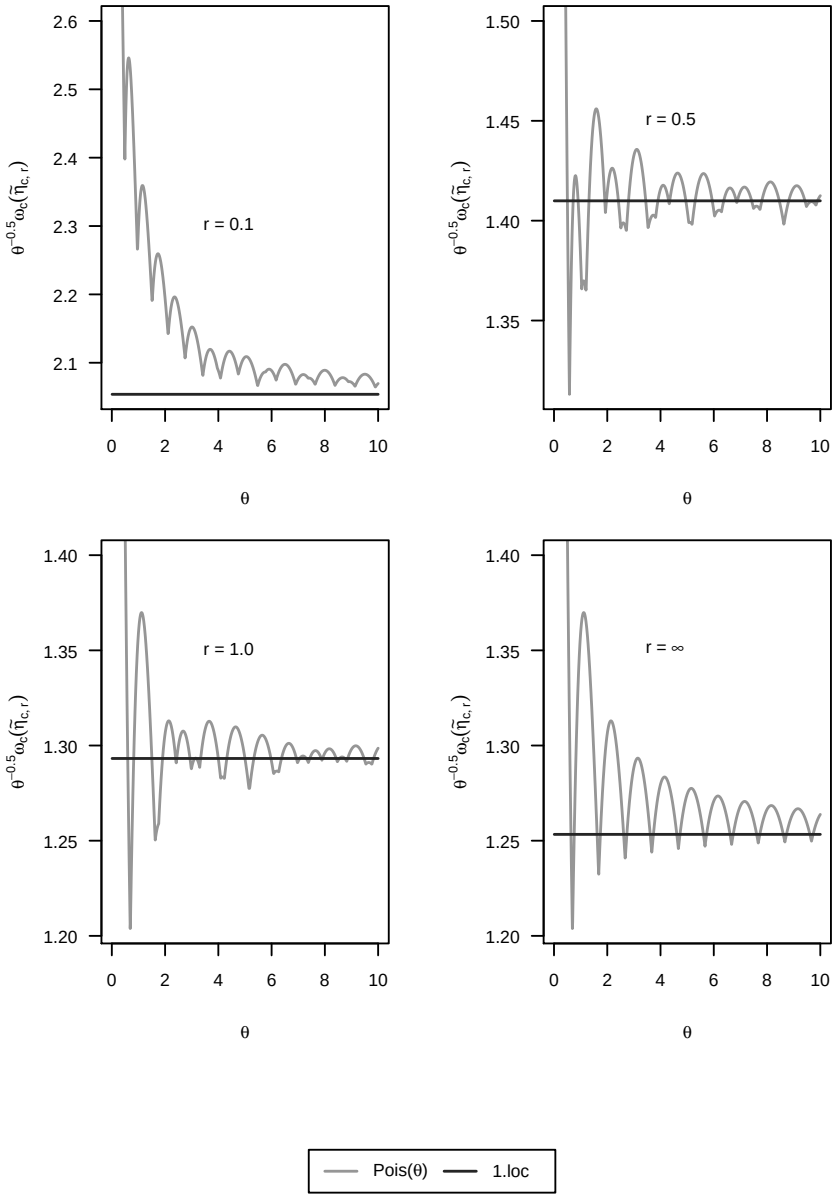

Figure 4.6: Normal approximation of the standardized infinitesimal bias terms $\theta^{-1/2}\omega_c(\tilde{\eta}_{c,r})$ for $\theta \in (0, 10]$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

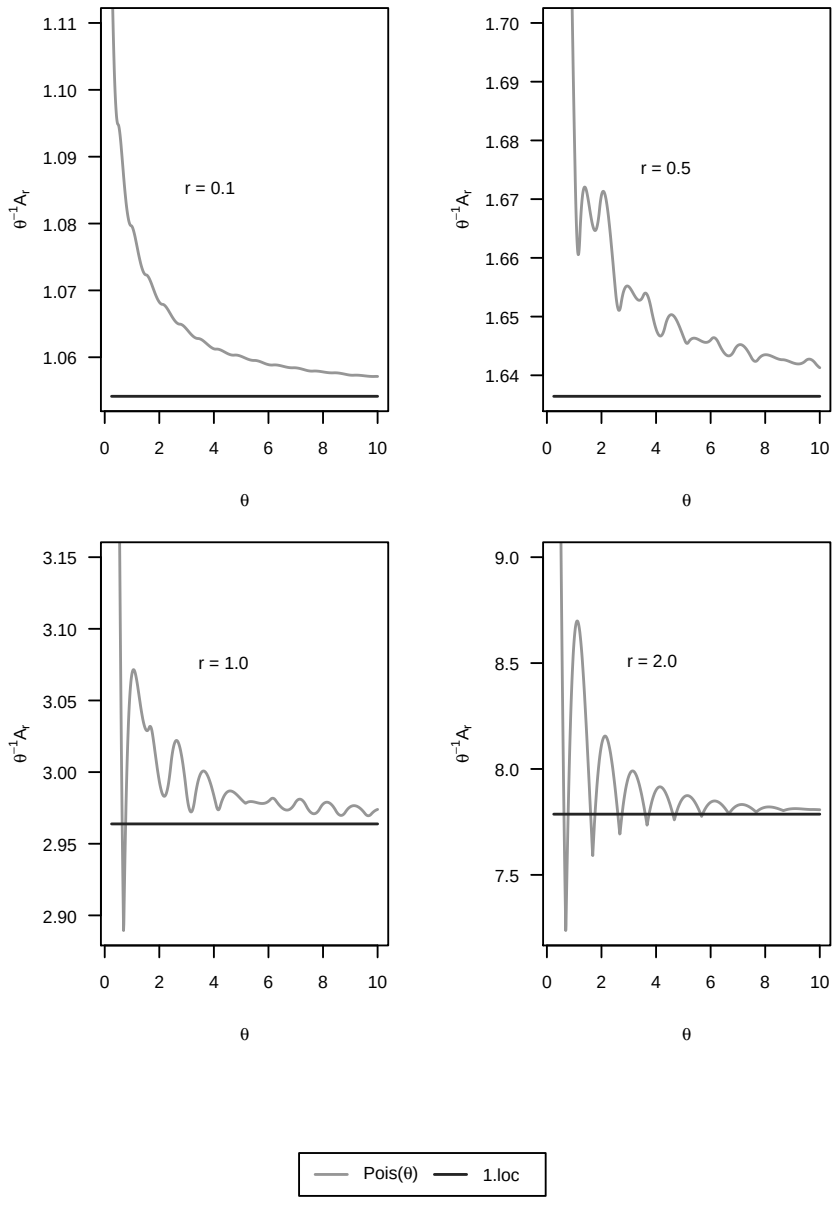

Figure 4.7: Normal approximation of the standardized maximum asymptotic MSE $\theta^{-1}A_r$ for $\theta \in (0, 10]$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, 2.0$.

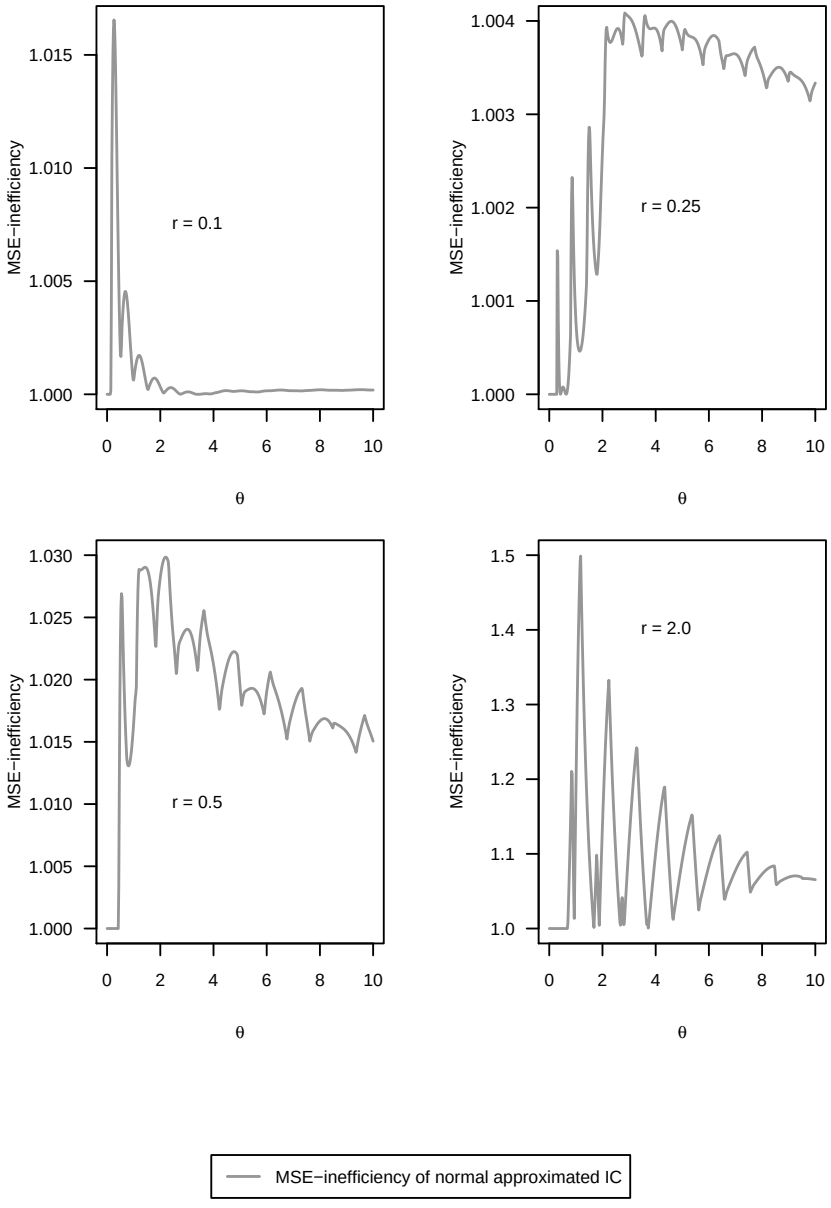

Figure 4.8: MSE–inefficiency of the normal approximated IC for $\theta \in (0, 10]$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.25, 0.5, 2.0$.

4.2.1.4 Poisson Approximation

The Lagrange multipliers contained in the optimally robust ICs in case of the Poisson model are also approximated by the corresponding Lagrange multipliers in case of the binomial model. This is stated in the following lemma. To distinguish between the two models, we introduce the additional superscripts "bin" and "pois", respectively.

Lemma 4.2.4 *Let $D = 1$. Consider the binomial model $(* = c)$ with size $m \in \mathbb{N}$ and probability of success $p_m \in (0,1)$ and the Poisson model $(* = c)$ with parameter $\theta \in (0,\infty)$ where $mp_m \to \theta \in (0,\infty)$ as $m \to \infty$. Then,*

$$\lim_{m\to\infty} \gamma_m^2 A_r^{\text{bin}} = A_r^{\text{pois}} \qquad \lim_{m\to\infty} \gamma_m^{-1} c_r^{\text{bin}} = c_r^{\text{pois}} \tag{4.2.11}$$

for all $r \in (0,\infty)$ where $\gamma_m = \frac{m}{1-p_m}$ and

$$\lim_{m\to\infty} \gamma_m \omega_c^{\min,\,\text{bin}} = \omega_c^{\min,\,\text{pois}} \tag{4.2.12}$$

If $\text{med}(\Lambda^{\text{pois}})$ is unique, respectively $\text{med}(\Lambda^{\text{pois}})$ non-unique and $r < \bar{r}$, then also

$$\lim_{m\to\infty} \gamma_m^{-1} z_r^{\text{bin}} = z_r^{\text{pois}} \tag{4.2.13}$$

Proof We have

$$\gamma_m^{-1}\Lambda^{\text{bin}}(y) = \frac{y - mp_m}{mp_m} \tag{4.2.14}$$

By the Poisson approximation of the binomial distribution and $mp_m \to \theta$ we get $\mathcal{L}(\gamma_m^{-1}\Lambda^{\text{bin}}) \xrightarrow{w} \mathcal{L}(\Lambda^{\text{pois}})$ as $m \to \infty$ where $\mathrm{E}\,\gamma_m^{-1}\Lambda^{\text{bin}} = 0$ for all $m \in \mathbb{N}$ and

$$\mathrm{E}\left(\gamma_m^{-1}\Lambda^{\text{bin}}\right)^2 = \tfrac{1-p_m}{mp_m} \longrightarrow \theta^{-1} = \mathrm{E}\left(\Lambda^{\text{pois}}\right)^2 \qquad \text{as} \qquad m \to \infty \tag{4.2.15}$$

Hence, we can apply Theorem 2.4.1 which yields (4.2.11)–(4.2.13). ////

Remark 4.2.5 **(a)** If the median of Λ^{pois} is non-unique, the standardized centering constant $\lambda_m^{-1} z_r^{\text{bin}}$ does not necessarily converge. As we see in Figure 4.9, different choices of p_m lead to different accumulation points. In case $p_m = \theta/m$, the median of $\gamma_m^{-1}\Lambda^{\text{bin}}$ is identical to the maximum median of Λ^{pois}. However, if the median of Λ^{pois} is unique, respectively non-unique and $r < \bar{r}$, then the preceding lemma implies (4.2.13) for all $r \in [0,\infty]$; an example is given in Figure 4.10.

(b) The convergences of the standardized infinitesimal bias terms $\gamma_m \omega_c(\tilde{\eta}_{c,r}^{\text{bin}})$ and the standardized minimax asymptotic MSE $\gamma_m^2 A_r$ are visualized in Figure 4.11 and 4.12, respectively. An example for the Poisson approximation of the MSE–inefficiencies is given in Section 4.3.

(c) Figure 4.13 shows the MSE–inefficiency of the Poisson approximated IC. That is, we used the Lagrange multipliers $A_r^{\text{pois}}/\gamma_m^2$, $z_r^{\text{pois}}\gamma_m$ and $c_r^{\text{pois}}\gamma_m$ instead of the optimal A_r^{bin}, z_r^{bin} and c_r^{bin}; confer Lemma 4.2.4. To make sure that the resulting function is indeed an IC (with respect to the binomial model), we additionally centered and standardized this function. As in case of the normal approximation the Poisson approximation seem to get worse for increasing radius which is

contrary to the results in case of total variation neighborhoods where the quality of the approximation increases with increasing radius; confer Subsubsection 4.2.2.4.
////

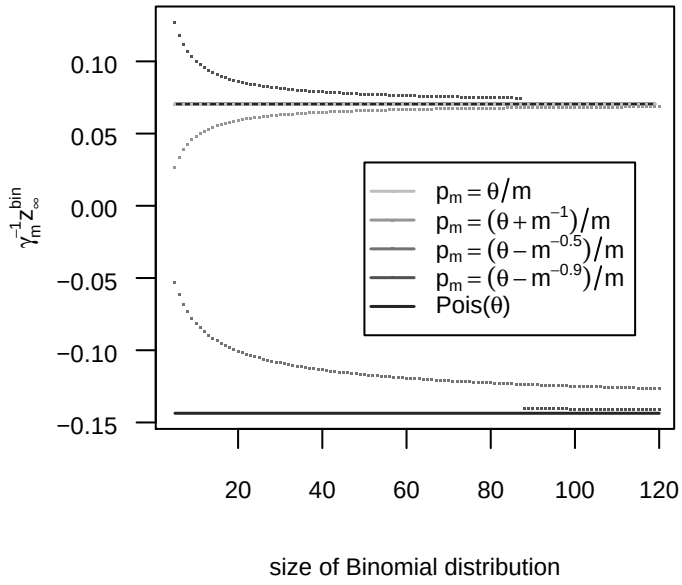

Figure 4.9: Poisson approximation of the standardized lower case centering constant $\gamma_m^{-1} z_\infty^{\text{bin}}$ for $m \in [5, 120]$ and $\theta = 4.670909$ (median of Λ^{pois} non-unique) in case of contamination neighborhoods ($* = c$); confer also Remark 4.2.5 (a)

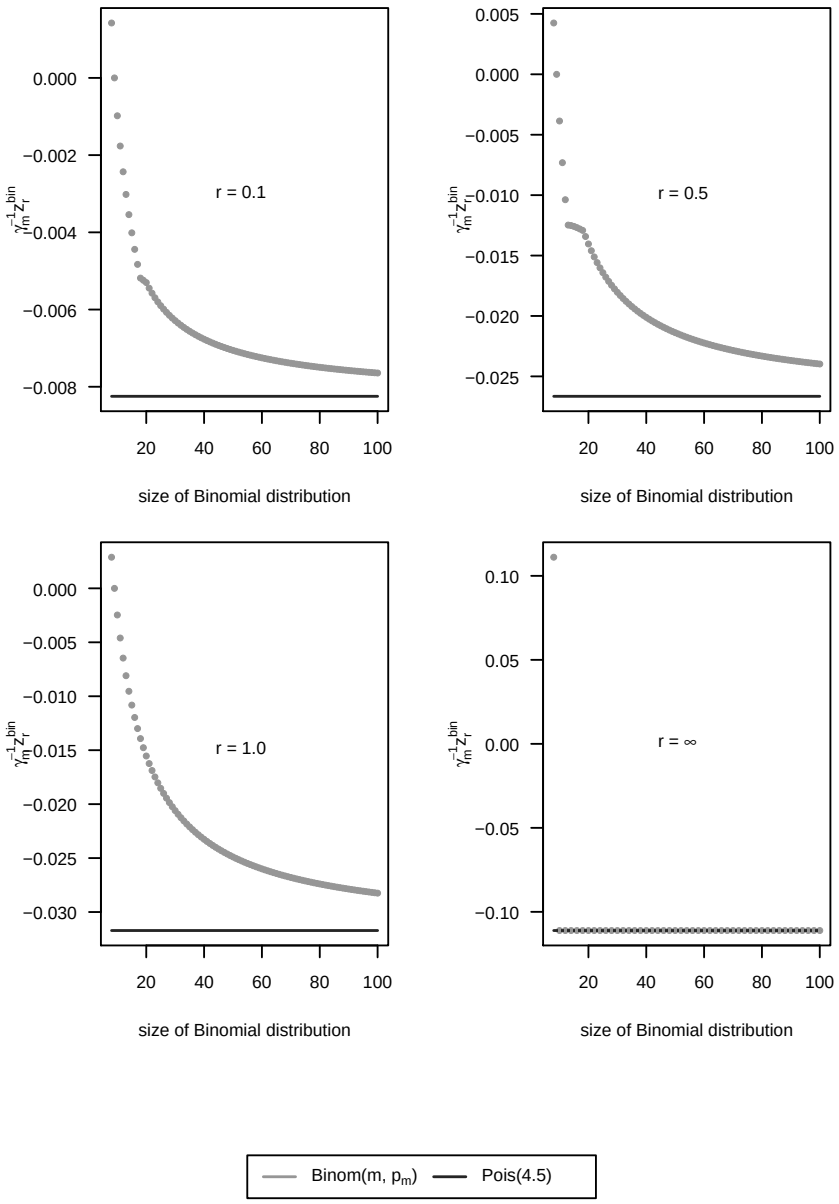

Figure 4.10: Poisson approximation of the standardized centering constant $\gamma_m^{-1} z_r^{\text{bin}}$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

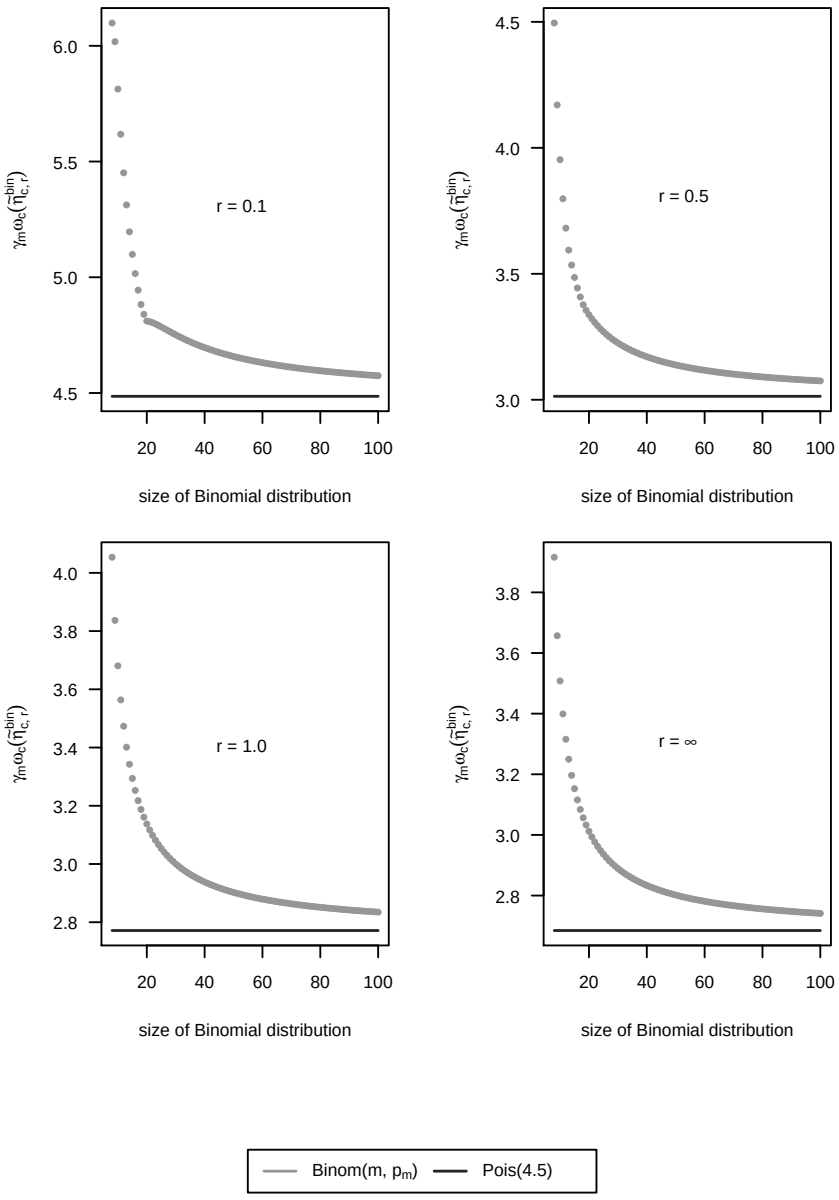

Figure 4.11: Poisson approximation of the standardized infinitesimal bias terms $\gamma_m \omega_c(\tilde{\eta}_{c,r}^{\mathrm{bin}})$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

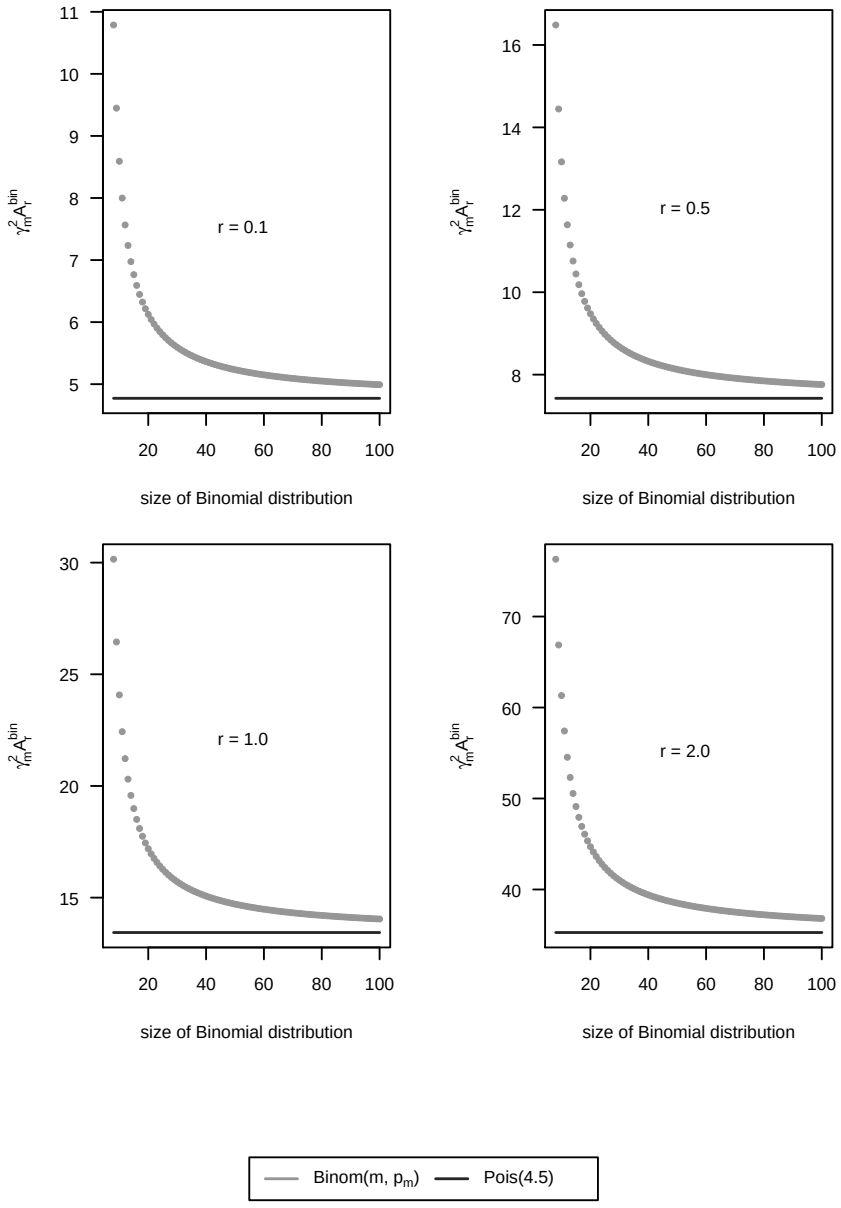

Figure 4.12: Poisson approximation of the standardized maximum asymptotic MSE $\gamma_m^2 A_r^{\text{bin}}$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ for contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, 2.0$.

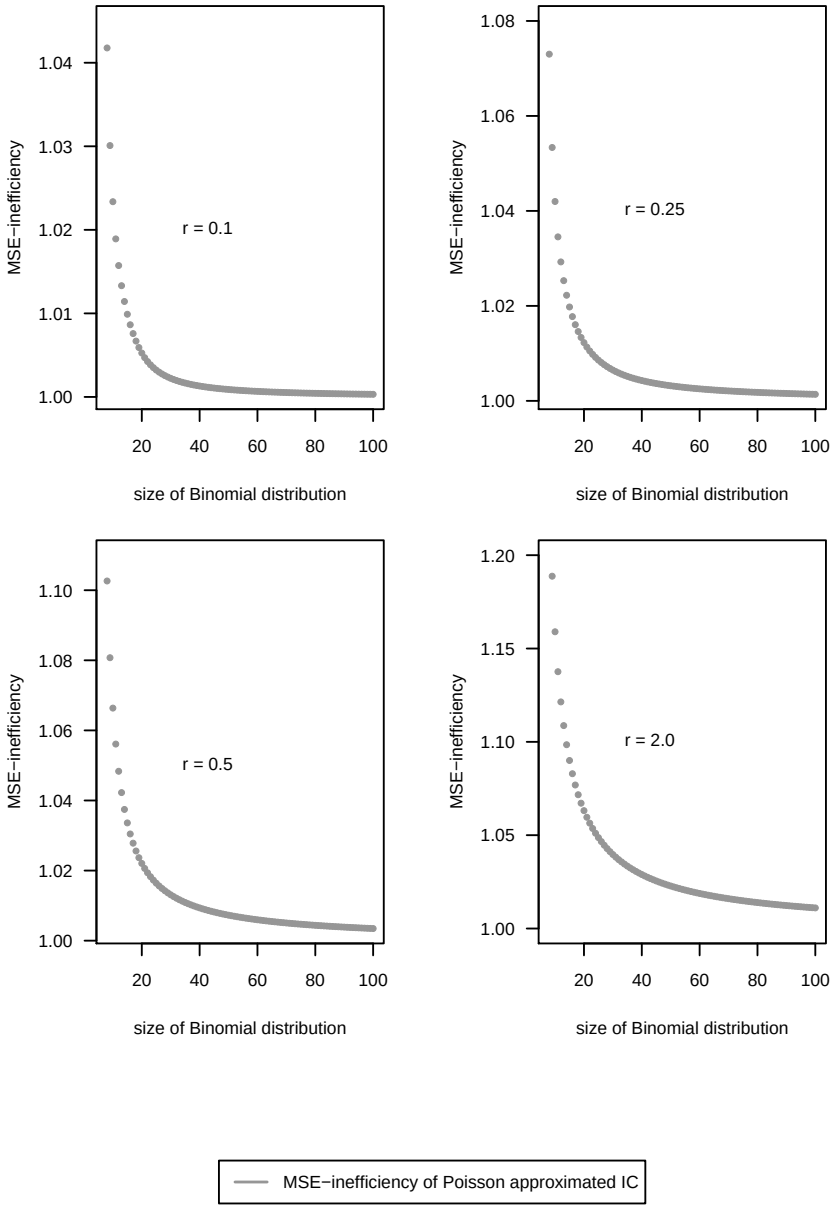

Figure 4.13: MSE–inefficiency of the Poisson approximated IC for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0.1, 0.5, 1.0, 2.0$.

4.2.2 Total Variation Neighborhoods

4.2.2.1 Mean Square Error Solution

Second, we consider the Poisson model (4.1.1) with infinitesimal total variation neighborhoods (1.2.5). The optimally robust IC $\tilde{\eta}_{v,r}$ supplied by Theorem 1.3.9 (a) and Theorem 1.3.11 (c) for some given $D \in \mathbb{R} \setminus \{0\}$ can be rewritten as

$$\tilde{\eta}_{v,r}(y) = A_r\big[\, g_r \vee \Lambda(y) \wedge (\, g_r + c_r)\big] \tag{4.2.16}$$

where

$$0 = \mathrm{E}\,\big(g_r - \Lambda\big)_+ - \mathrm{E}\,\big(\Lambda - (g_r + c_r)\big)_+ \tag{4.2.17}$$

$$D = A_r\,\mathrm{E}\,\Lambda\big[\, g_r \vee \Lambda \wedge (\, g_r + c_r)\big] \tag{4.2.18}$$

and

$$r^2\, c_r = \mathrm{E}\,\big(g_r - \Lambda\big)_+ \tag{4.2.19}$$

For $r = \infty$ Theorem 1.3.9 (b) provides

$$\tilde{\eta}_{v,\infty}\,(y) = \omega_v^{\min}\,\mathrm{sign}(D)\left(\frac{P(y<\theta)}{P(y\neq\theta)}\,\mathrm{I}(y>\theta) - \frac{P(y>\theta)}{P(y\neq\theta)}\,\mathrm{I}(y<\theta)\right) \tag{4.2.20}$$

with minimum bias

$$\omega_v^{\min} = \frac{|D|\theta}{\mathrm{E}(y-\theta)_+} \tag{4.2.21}$$

For a plot of the optimally robust ICs in case $\theta = 5$ and for different values of r see Figure 4.14.

Remark 4.2.6 **(a)** Similarly to the binomial model it might be necessary to extend the optimally robust ICs to $\mathbb{R} \setminus \mathbb{N}_0$. For more details see Remark 3.2.1.

(b) The map $\theta \mapsto \bar{r}$ has discontinuity points at $\theta \in \mathbb{N}$ since we may choose $\theta \in (0,\infty)$ such that θ is arbitrarily close to some integer value in $\mathbb{N}$, the gap $\gamma = \inf_{P_\theta}\big\{|\Lambda|\,\big|\,|\Lambda| > 0\big\}$ becomes arbitrarily small, respectively M and the lower case radius $\bar{r}$ defined in (2.1.34) and (2.1.33) become arbitrarily large. However, if we choose $\theta \in (0,\infty)$ such that $\theta \in \mathbb{N}$ we get $\gamma = 1/\theta$ $(= \gamma_1 = -\gamma_2)$. For a plot of the lower case radius $\bar{r}$ for $\theta \in (0, 10]$ see Figure 4.15. ////

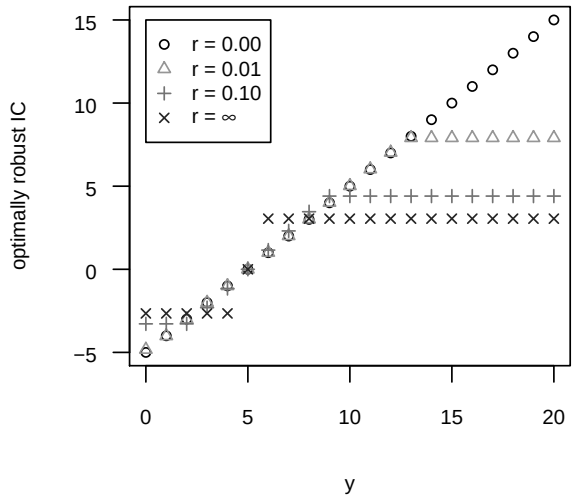

Figure 4.14: Optimally Robust ICs for $\mathrm{Pois}(5)$ in case of total variation neighborhoods $(* = v)$ with radius $r = 0, 0.01, 0.10, \infty$.

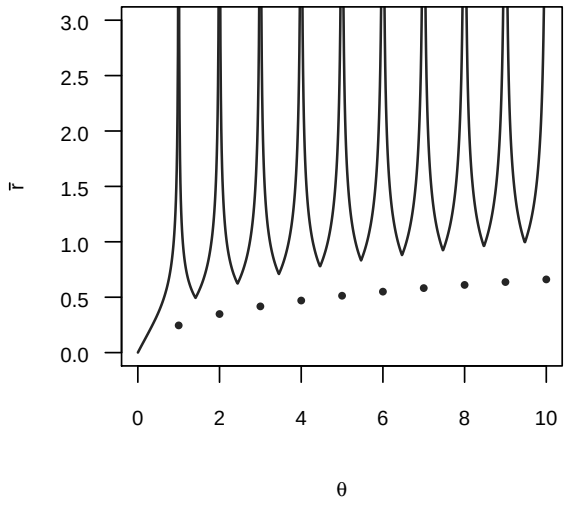

Figure 4.15: Lower case radius $\bar{r}$ for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$.

4.2.2.2 Continuity and Uniqueness of Lagrange Multipliers

The Lagrange multipliers contained in the optimally robust ICs are unique and continuous in the radius r as well as in the parameter θ; confer Subsection 2.1.4, Proposition 2.1.9 and Theorem 2.1.11, respectively. This is visualized in Figure 4.16 and Figure 4.17, respectively. In Figure 4.17 we use the `R` function `points` (cf. R Development Core Team (2005)) to illustrate that the lower clipping bound and the asymptotic variance indeed have no jumps but attain values between the local extrema, too. But, these Lagrange multipliers and hence the standardized asymptotic bias b_r, the asymptotic variance $A_r - r^2 b_r^2$ and the maximum asymptotic MSE A_r are not necessarily smooth functions in r, respectively θ. Only the maximum asymptotic MSE A_r seems to be smooth in r. Moreover, in case $r = \infty$ the lower as well as the upper clipping bound are discontinuous for $\theta \in \mathbb{N}$; confer Figure 4.18.

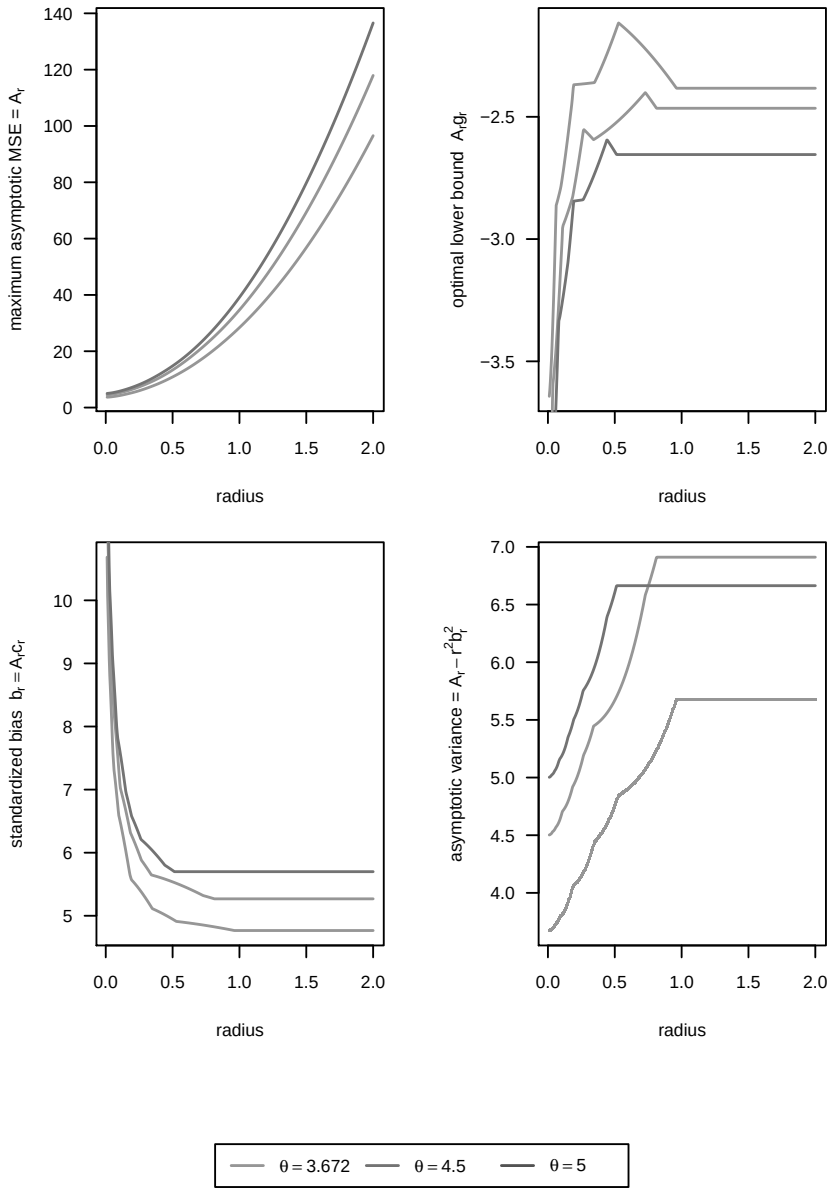

Figure 4.16: Continuity in the radius r of the Lagrange multipliers contained in the MSE optimal ICs for $r \in (0, 2.0]$ and $\theta = 3.672, 4.5, 5.0$ in case of total variation neighborhoods ($* = v$).

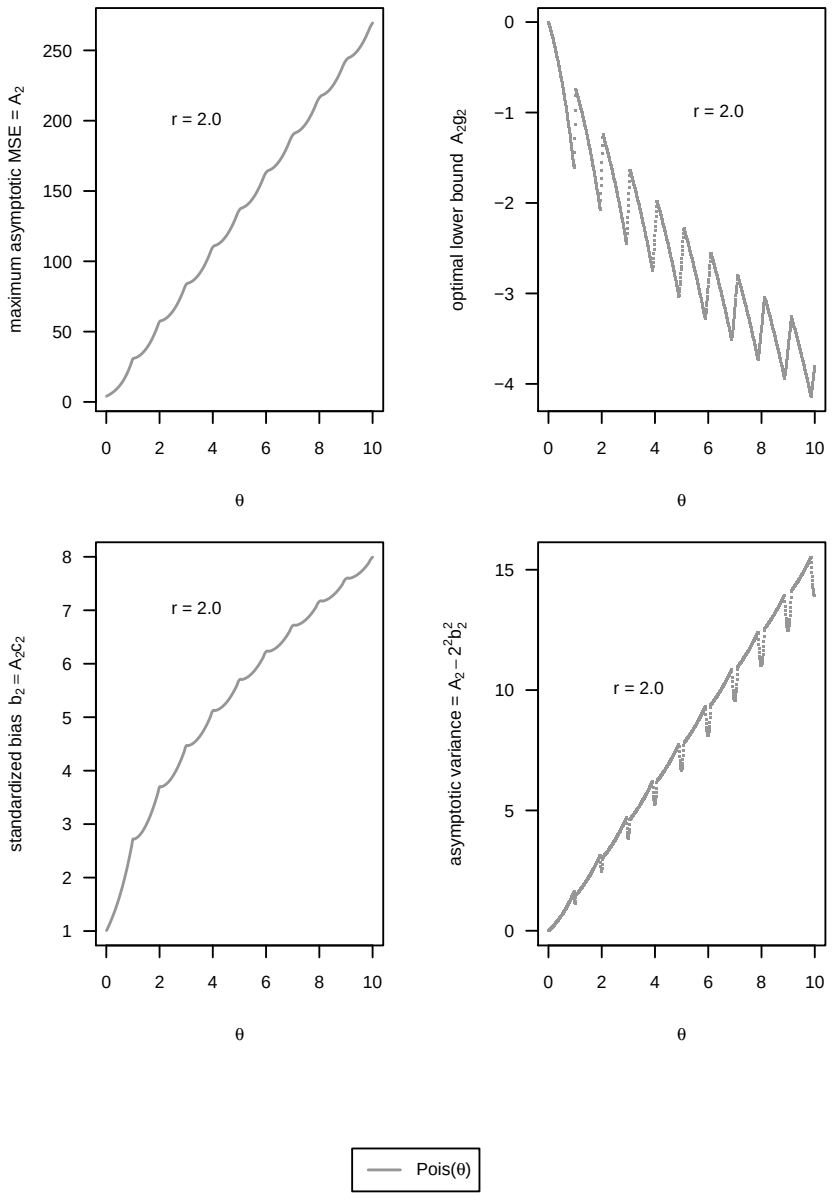

Figure 4.17: Continuity in the parameter θ of the Lagrange multipliers contained in the MSE optimal ICs for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$ with radius $r = 2.0$.

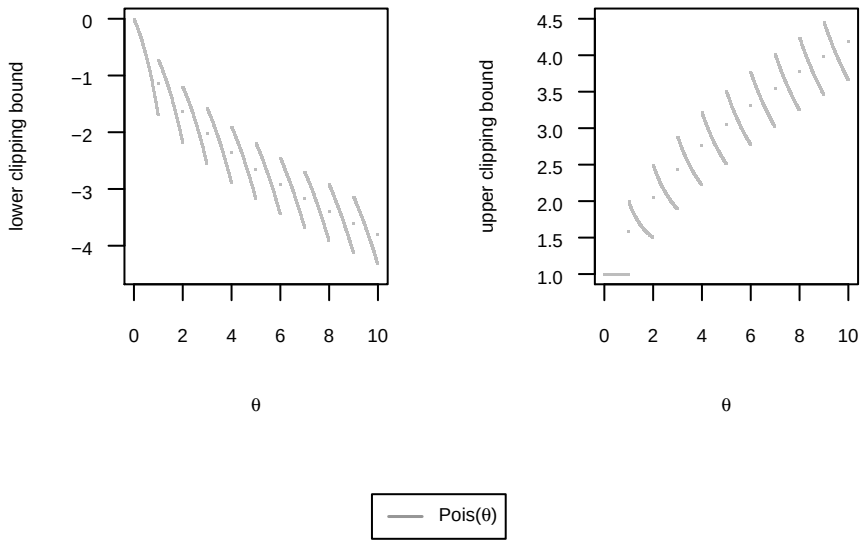

Figure 4.18: Discontinuity of the lower and upper clipping bound for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$ with radius $r = \infty$.

4.2.2.3 Normal Approximation

The following lemma states the normal approximation of the Lagrange multipliers contained in the optimally robust ICs in case of total variation neighborhoods where the optimally robust IC in case of one-dimensional normal location and total variation neighborhoods is specified in (3.2.23)–(3.2.25). The corresponding lower case solution given in (3.2.26) achieves minimum bias $\omega_v^{\min,\,1.\mathrm{loc}} = \sqrt{2\pi} = 2\omega_c^{\min,\,1.\mathrm{loc}}$.

Lemma 4.2.7 *Let $D = 1$. It holds,*

$$\lim_{\theta\to\infty} \theta^{-1} A_r = A_r^{1.\mathrm{loc}} \qquad \lim_{\theta\to\infty} \sqrt{\theta}\, g_r = g_r^{1.\mathrm{loc}} \qquad \lim_{\theta\to\infty} \sqrt{\theta}\, c_r = c_r^{1.\mathrm{loc}} \tag{4.2.22}$$

for all $r \in (0, \infty)$ and

$$\lim_{\theta\to\infty} \theta^{-1/2} \omega_v^{\min} = \omega_v^{\min,\,1.\mathrm{loc}} \tag{4.2.23}$$

Proof Consequence of Theorem 2.4.1; confer also the proof of Lemma 4.2.2. ////

Remark 4.2.8 **(a)** Like in the cases before, we again illustrate the convergence of the standardized lower clipping bounds $\theta^{-1/2}A_r g_r$, respectively of the standardized infinitesimal bias terms $\theta^{-1/2}\omega_v(\tilde{\eta}_{v,r})$; see Figure 4.19 and Figure 4.20, respectively. In case $r = \infty$, we again see that the discontinuity points of the lower clipping bound $A_r g_r$ coincide with those values of θ for which $m\theta \in \mathbb{N}$. At these non-uniqueness points the standardized infinitesimal bias terms attain local maxima in case $r = \infty$. These results are very similar to the binomial model; confer Remark 3.2.6 (a).

(b) Some examples for the convergence of the standardized minimax asymptotic MSE $\gamma_m^2 A_r$, respectively the MSE–inefficiencies are given in Figure 4.21 and Subsection 4.3.2, respectively.

(c) Figure 4.21 shows the MSE–inefficiency of the normal approximated IC. That is, we used the Lagrange multipliers $A_r^{1.\mathrm{loc}}\theta$, $g_r^{1.\mathrm{loc}}/\sqrt{\theta}$ and $c_r^{1.\mathrm{loc}}/\sqrt{\theta}$ instead of the optimal A_r, g_r and c_r; confer Lemma 4.2.7. To make sure that the resulting function is indeed an IC (with respect to the Poisson model), we additionally centered and standardized this function. The results are similar to the binomial model (cf. Subsubsection 3.2.2.3). We get very small MSE–inefficiencies in case of total variation neighborhoods and the approximation even seems to become better with increasing radius. In contrast, the MSE–inefficiencies in case of contamination neighborhoods may be quite large for radii $r > \bar{r}$ and the quality of the approximation decreases with increasing radius. Interestingly, for small radii r ($r < 0.5$) the approximation in case of contamination neighborhoods seem to be better than in case of total variation neighborhoods. ////

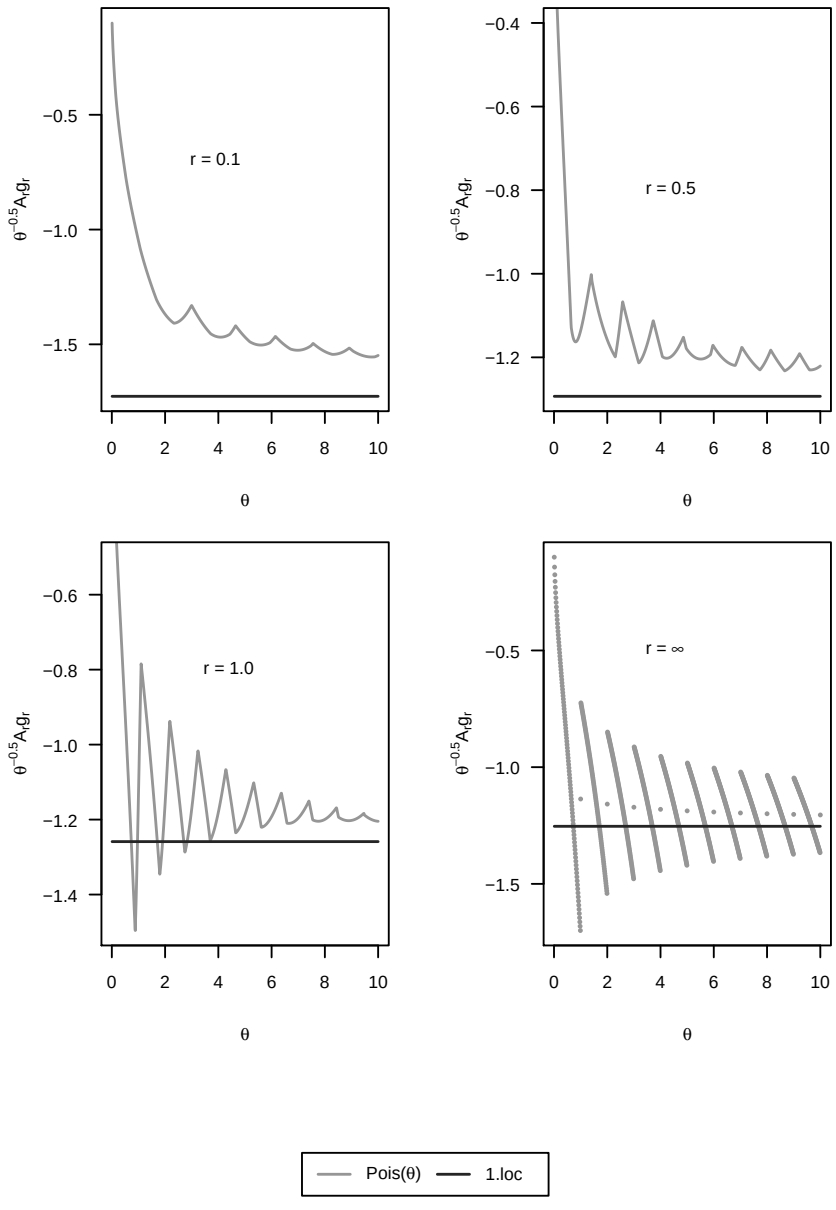

Figure 4.19: Normal approximation of the standardized lower clipping bound $\theta^{-1/2}A_r g_r$ for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.25, 0.5, \infty$.

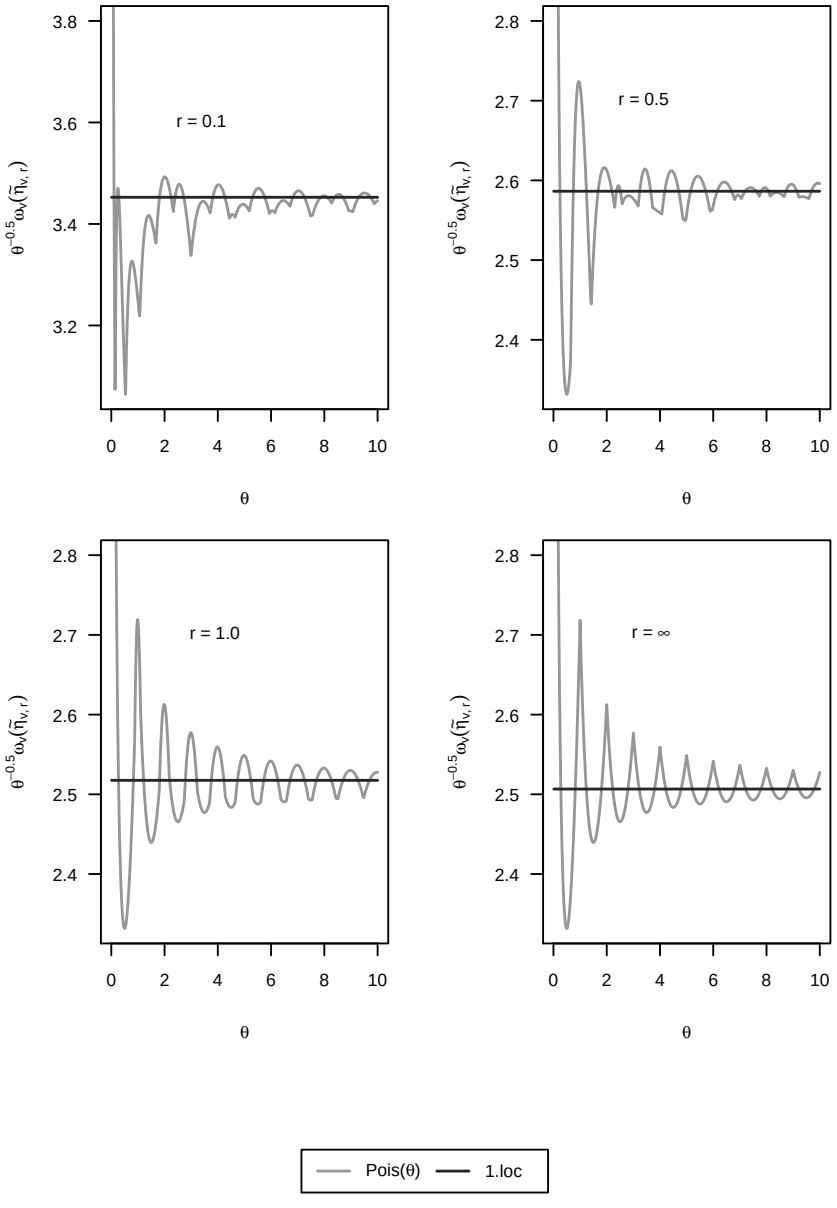

Figure 4.20: Normal approximation of the standardized infinitesimal bias terms $\theta^{-1/2}\omega_v(\tilde{\eta}_{v,r})$ for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

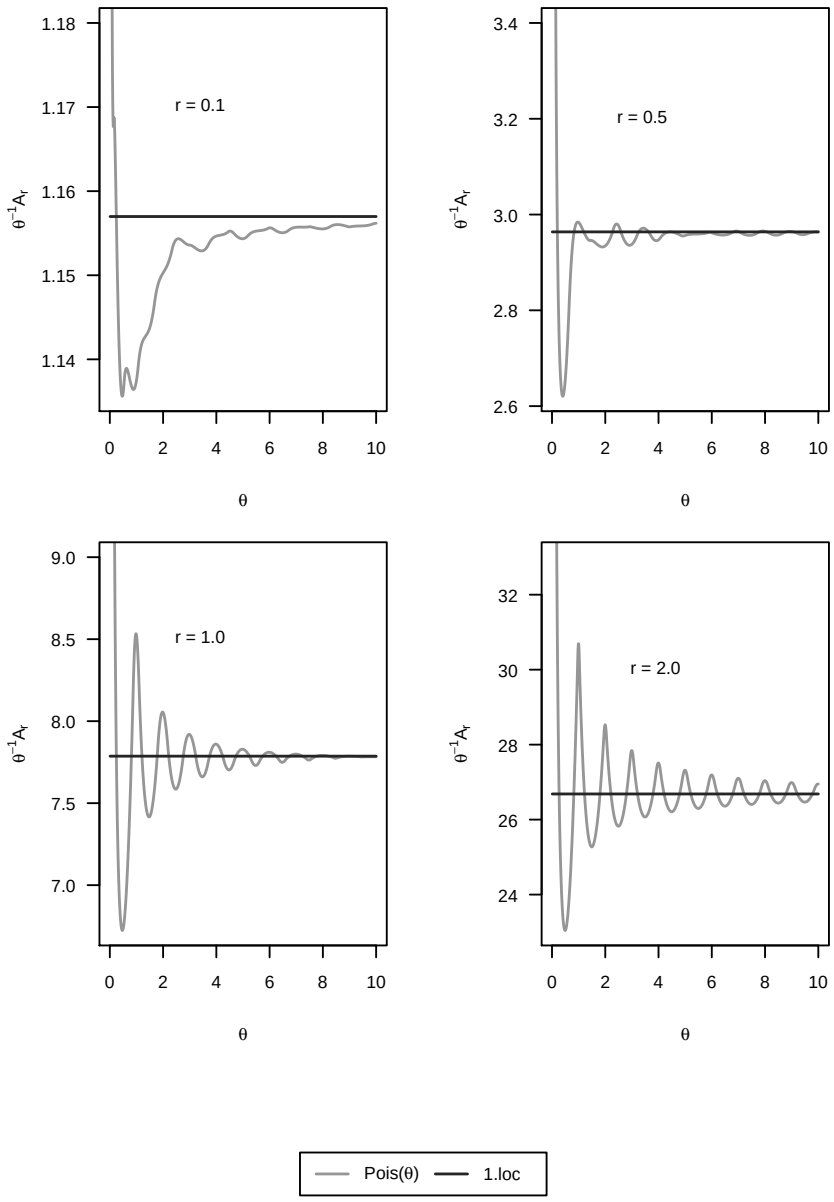

Figure 4.21: Normal approximation of the standardized maximum asymptotic MSE $\theta^{-1}A_r$ for $\theta \in (0, 10]$ in case of total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.5, 1.0, 2.0$.

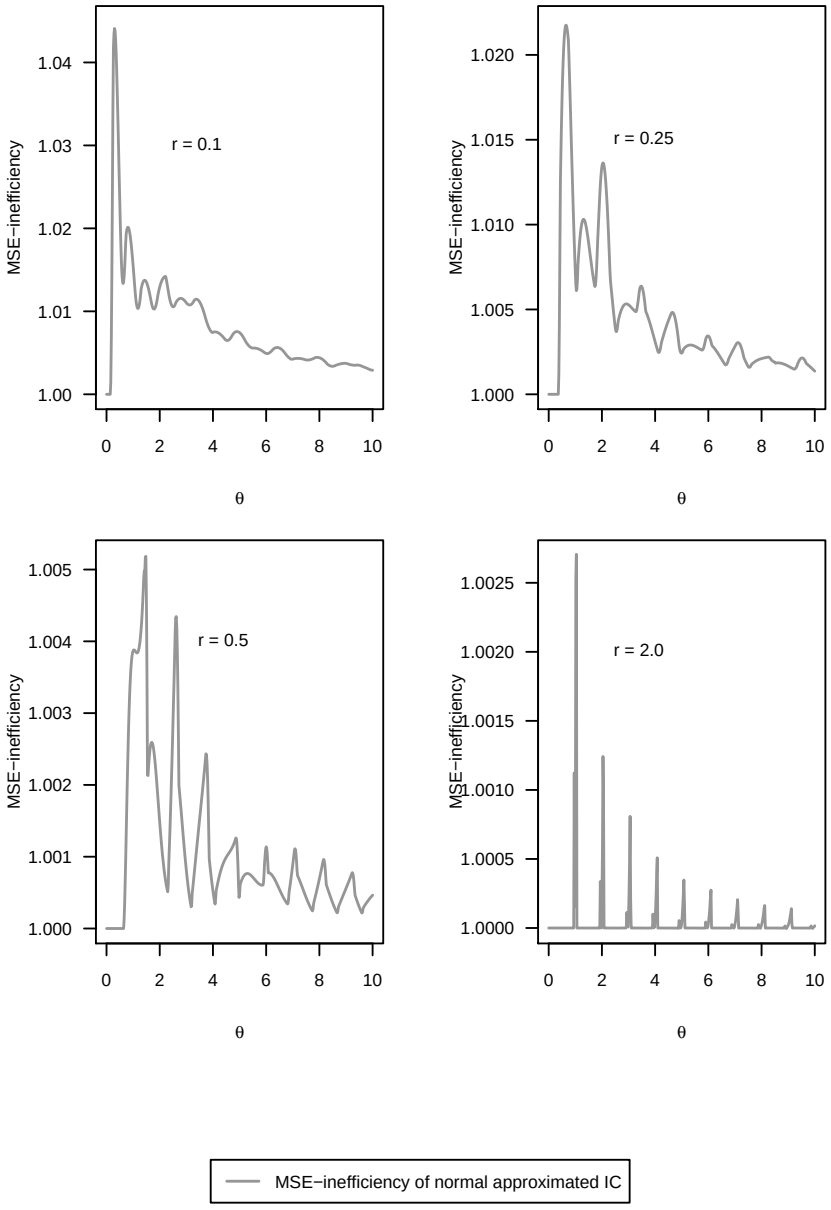

Figure 4.22: MSE–Inefficiency of the normal approximated IC for $\theta \in (0, 10]$ in case of total variation neighborhoods ($* = v$) with radius $r = 0.1, 0.25, 0.5, 2.0$.

4.2.2.4 Poisson Approximation

In case of total variation neighborhoods, the Lagrange multipliers contained in the optimally robust ICs in case of the Poisson model are also approximated by the corresponding Lagrange multipliers in case of the binomial model. This fact is stated in the following lemma. To distinguish between the two models, we again introduce the additional superscripts "bin" and "pois", respectively.

Lemma 4.2.9 *Let* $D = 1$. *Consider the binomial model* $(* = v)$ *with size* $m \in \mathbb{N}$ *and probability of success* $p_m \in (0,1)$ *and the Poisson model* $(* = v)$ *with parameter* $\theta \in (0,\infty)$ *where* $mp_m \to \theta \in (0,\infty)$ *as* $m \to \infty$. *Then,*

$$\lim_{m\to\infty} \gamma_m^2 A_r^{\text{bin}} = A_r^{\text{pois}} \qquad \lim_{m\to\infty} \gamma_m^{-1} g_r^{\text{bin}} = g_r^{\text{pois}} \qquad \lim_{m\to\infty} \gamma_m^{-1} c_r^{\text{bin}} = c_r^{\text{pois}} \quad (4.2.24)$$

for all $r \in (0,\infty)$ *where* $\gamma_m = \frac{m}{1-p_m}$ *and*

$$\lim_{m\to\infty} \gamma_m \omega_c^{\min,\,\text{bin}} = \omega_c^{\min,\,\text{pois}} \quad (4.2.25)$$

Proof Consequence of Theorem 2.4.1; confer also the proof of Lemma 4.2.4. ////

Remark 4.2.10 **(a)** The convergences of the standardized lower clipping bounds, the standardized infinitesimal bias terms and the standardized minimax asymptotic MSE are visualized in Figure 4.23, Figure 4.24 and Figure 4.25, respectively. For an example of the Poisson approximation of the MSE–inefficiencies we refer to Section 4.3.

(b) Figure 4.26 shows the MSE–inefficiency of the Poisson approximated IC. That is, we used the Lagrange multipliers $A_r^{\text{pois}}/\gamma_m^2$, $z_r^{\text{pois}}\gamma_m$ and $c_r^{\text{pois}}\gamma_m$ instead of the optimal A_r^{bin}, z_r^{bin} and c_r^{bin}; confer Lemma 4.2.9. To make sure that the resulting function is indeed an IC (with respect to the binomial model), we additionally centered and standardized this function. As in case of the normal approximation the quality of the Poisson approximation seem to get better for increasing radius. For the chosen parameter value $\theta = 4.5$ we even get a perfect approximation (MSE–inefficiency $\equiv 1$) for large radii. This is contrary to the results in case of contamination neighborhoods where the approximation gets worse with increasing radius; confer Subsubsection 4.2.1.4. ////

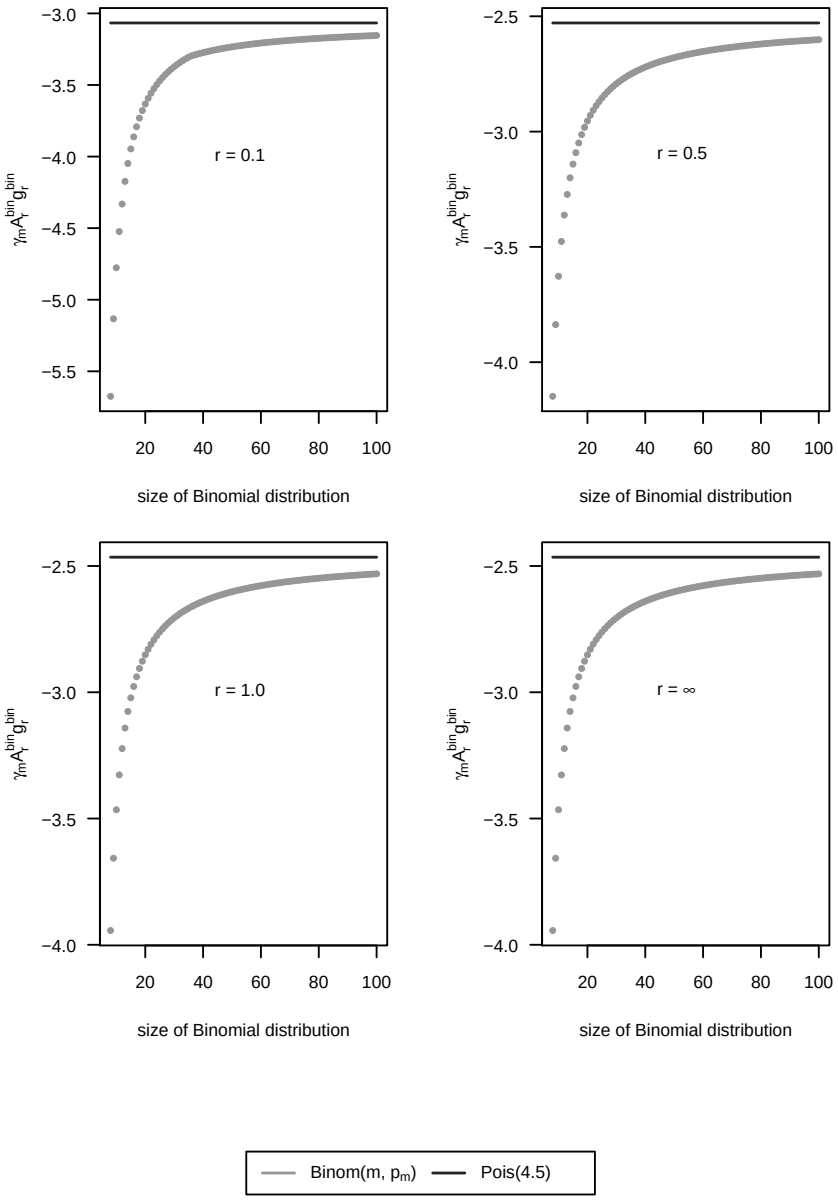

Figure 4.23: Poisson approximation of the standardized lower clipping bound $\gamma_m A_r^{\mathrm{bin}} g_r^{\mathrm{bin}}$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ for total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

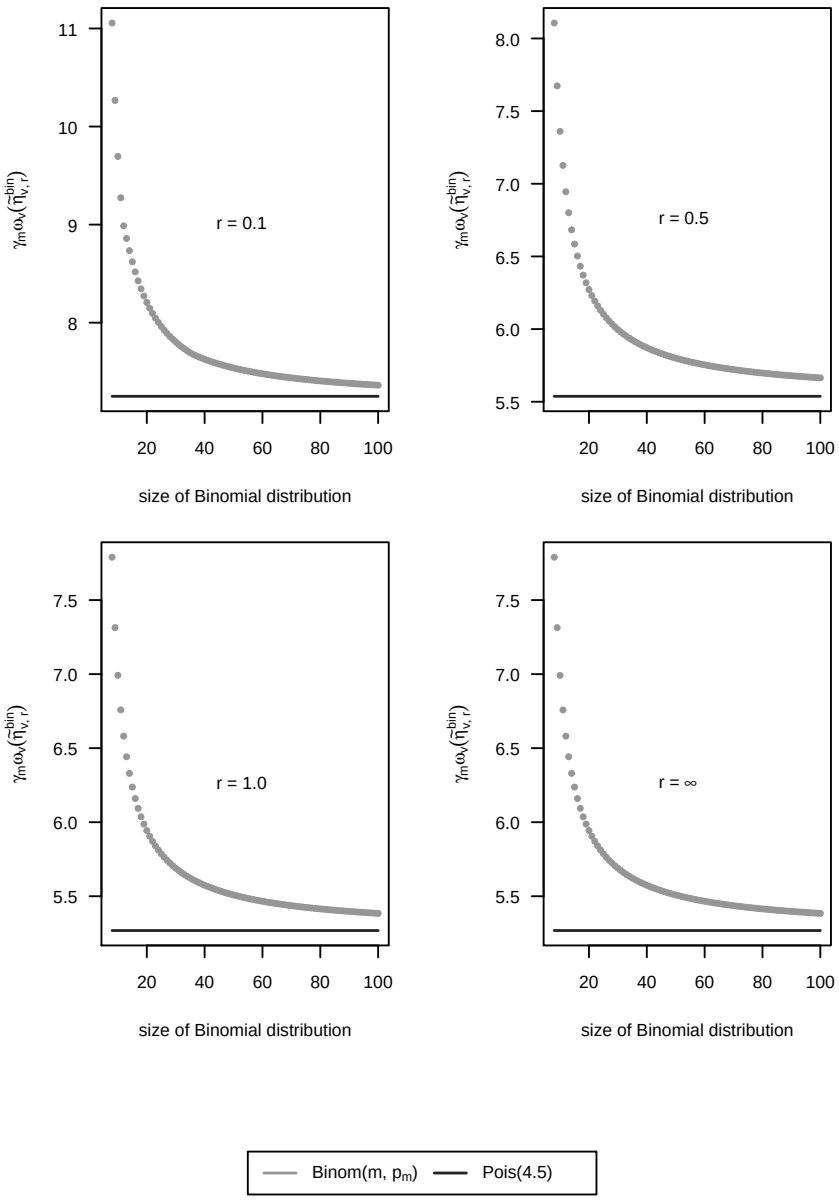

Figure 4.24: Poisson approximation of the standardized infinitesimal bias terms $\gamma_m \omega_v(\tilde{\eta}_{v,r}^{\mathrm{bin}})$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ for total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.5, 1.0, \infty$.

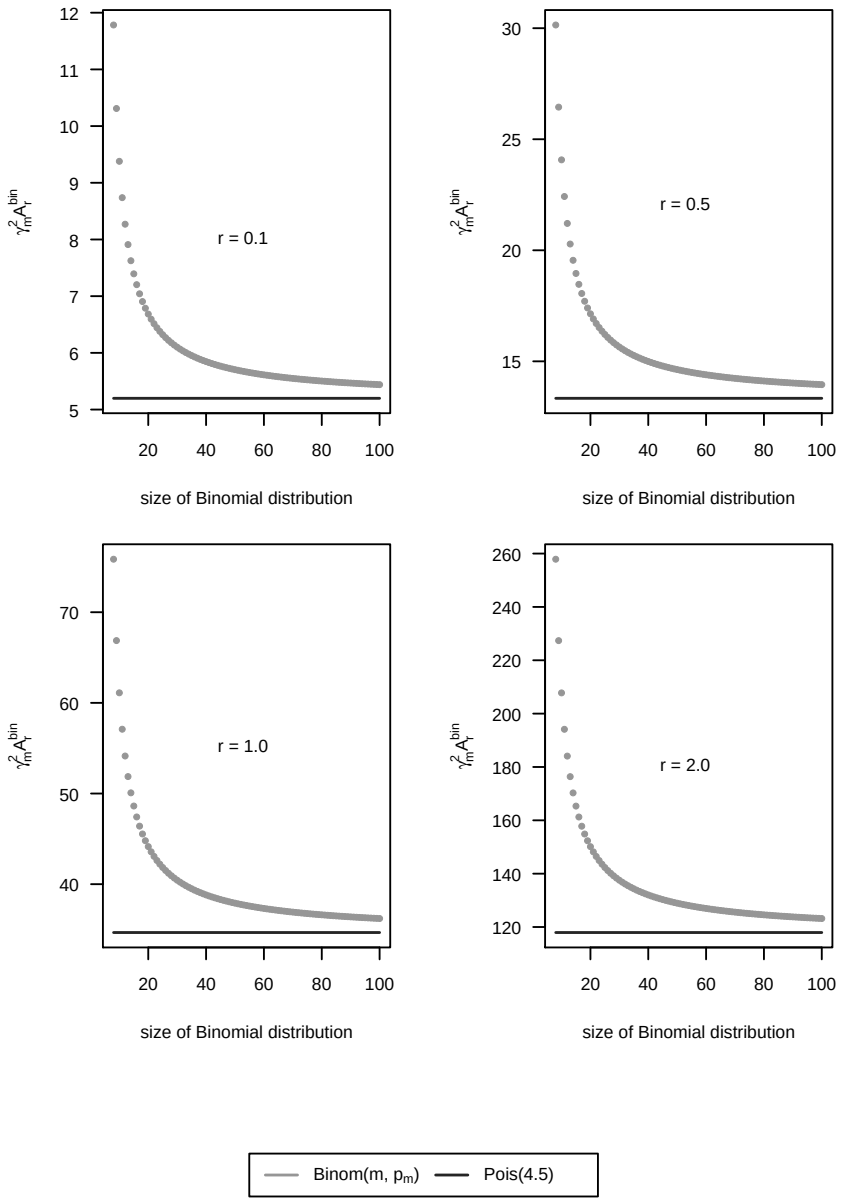

Figure 4.25: Poisson approximation of the standardized maximum asymptotic MSE $\gamma_m^2 A_r^{\mathrm{bin}}$ for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ for total variation neighborhoods $(* = v)$ with radius $r = 0.1, 0.5, 1.0, 2.0$.

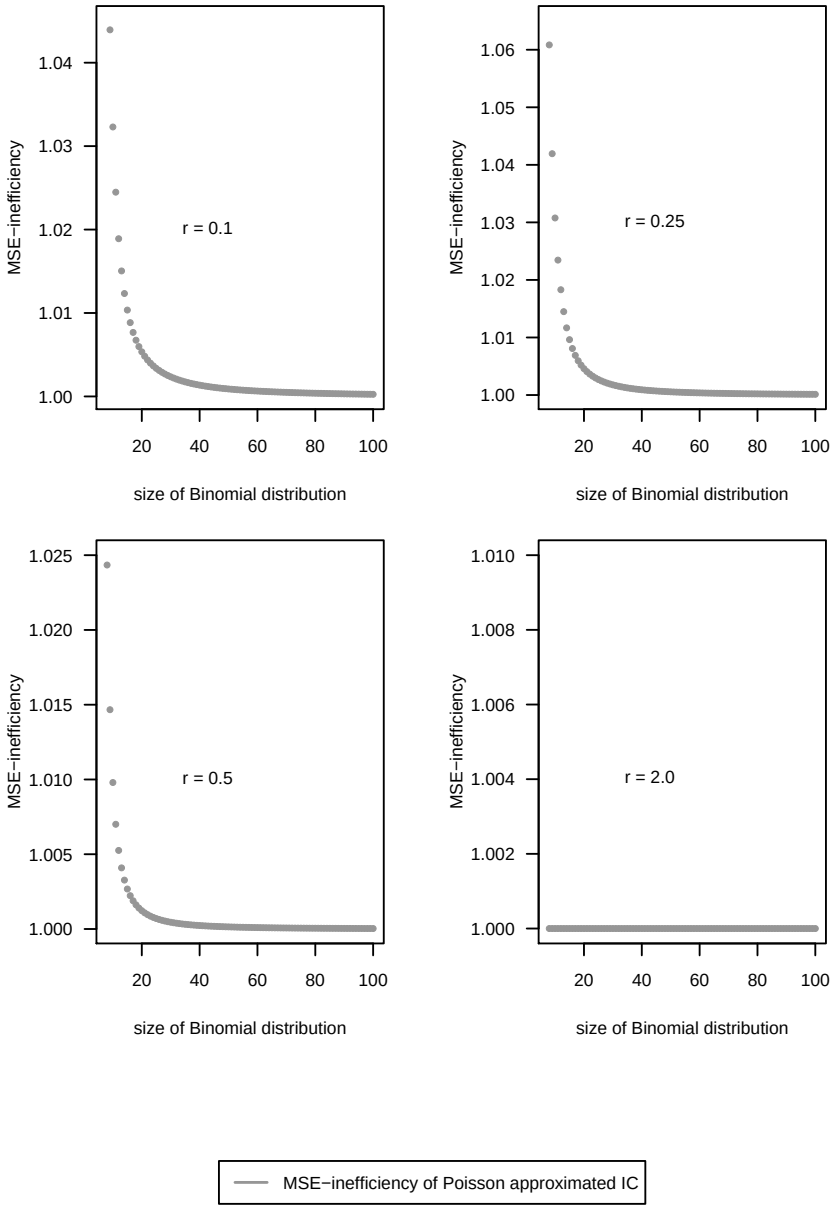

Figure 4.26: MSE–inefficiency of the Poisson approximated IC for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of total variation neighborhoods with radius $r = 0.1, 0.5, 1.0, 2.0$.

4.3 Least Favorable Radius

In this section we specify the least favorable radii r_0, r_2 and r_3 for different values of $\theta \in (0, \infty)$. The corresponding definitions are given in Section 2.2. Since the algorithm becomes less stable for $\theta \to 0$, because the inefficiency curves are very flat and the corresponding least favorable radii are very small, the smallest value of θ we consider is 0.01. As the inefficiency curves are also very flat near the least favorable radii, the computations lead to accurate numerical approximations of the inefficiencies but are less accurate concerning the least favorable radii. However, we get a good impression of the range of the least favorable radii and the corresponding inefficiencies. In addition, we included the corresponding inefficiencies and least favorable radii in case of one-dimensional normal location in view of the normal approximation mentioned in Remark 3.2.3 (b) and Remark 3.2.6 (b), respectively. It turns out, that the least favorable radii as well as the inefficiencies are already very close to the corresponding values of one-dimensional normal location in case $\theta > 10$. Hence, these values serve as good approximations and we restrict the presentation of the results to $\theta \leq 10$. We also illustrate the Poisson approximation of the MSE–inefficiencies; confer Figure 4.30 and Figure 4.34.

4.3.1 Contamination Neighborhoods

Similarly to the binomial case and contamination neighborhoods which is treated in Subsection 3.3.1, the behavior at values of θ for which the median of Λ is non-unique is different from the other values of θ. They lead to the largest inefficiencies and the largest least favorable radii. We determined these non-uniqueness points numerically and included these numerical approximations in the considered grid.
The resulting minimax subefficiencies are small, respectively very small. In all considered cases ($\theta \in (0, 10]$) they stay below about 30%, 16% and 10% in case of r_0, r_3 and r_2, respectively which is very similar to the binomial model (cf. Subsection 4.3.1). The corresponding least favorable radii are also small. In all cases they lie in the interval $(0, 0.92]$; confer Figures 4.27–4.29.

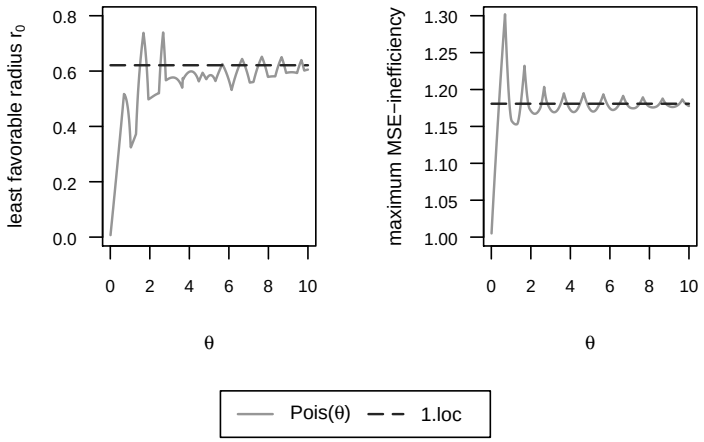

Figure 4.27: Least favorable radius r_0 and maximum MSE–inefficiency in case of contamination neighborhoods $(* = c)$ and $\theta \in (0, 10]$.

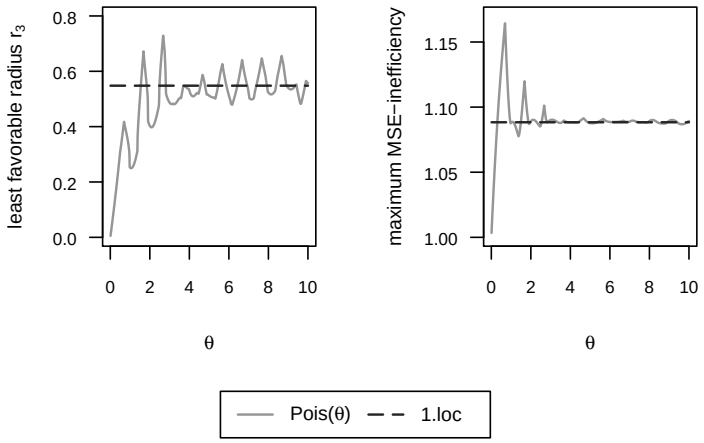

Figure 4.28: Least favorable radius r_3 and maximum MSE–inefficiency in case of contamination neighborhoods $(* = c)$ and $\theta \in (0, 10]$.

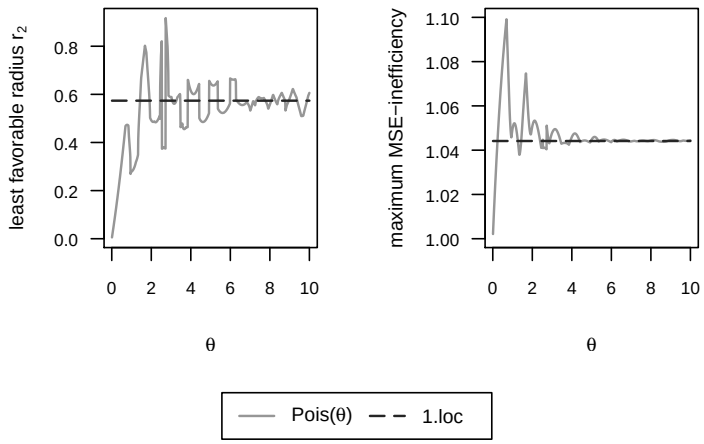

Figure 4.29: Least favorable radius r_2 and maximum MSE–inefficiency in case of contamination neighborhoods $(* = c)$ and $\theta \in (0, 10]$.

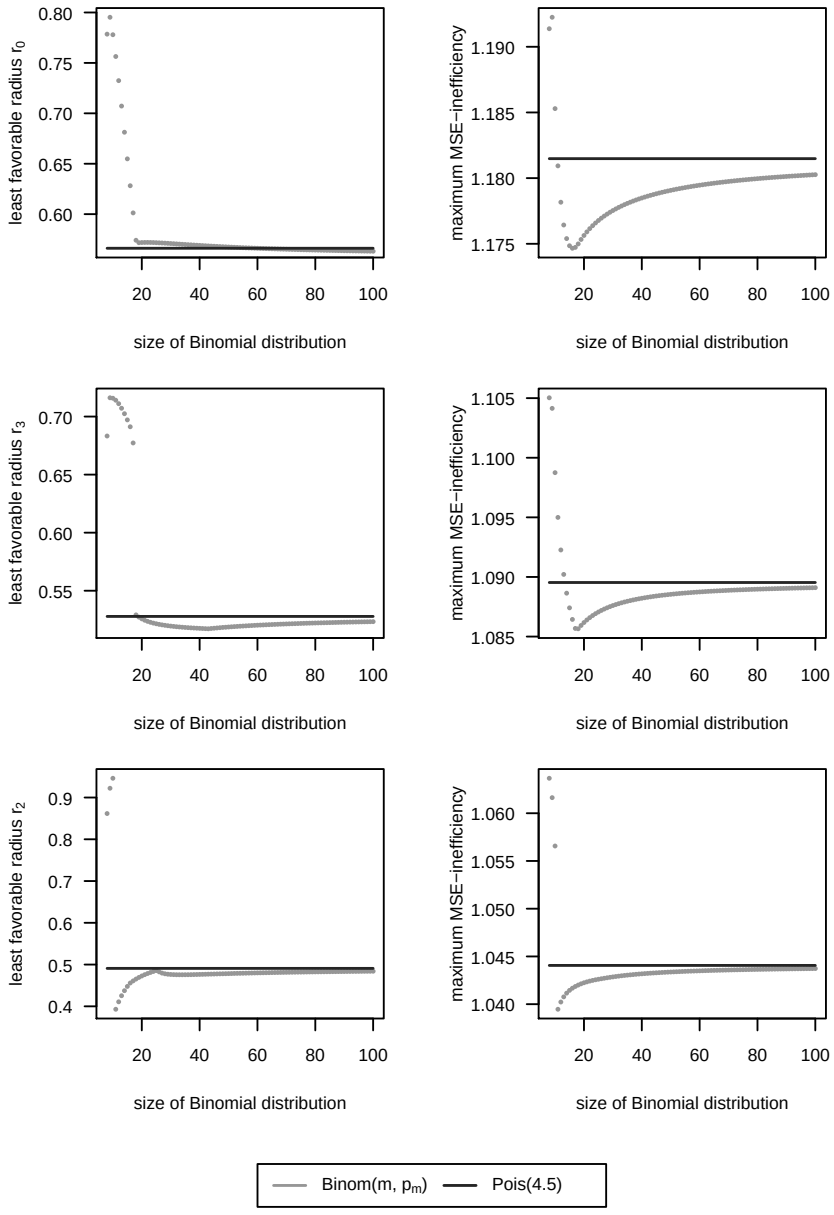

Figure 4.30: Poisson approximation of least favorable radii r_0, r_3, r_2 and maximum MSE–inefficiency for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of contamination neighborhoods $(* = c)$.

4.3.2 Total Variation Neighborhoods

Like in the binomial model, the minimax subefficiencies are even a little bit smaller than in case of contamination neighborhoods with respect to the whole range of θ values. They stay below about 22%, 13% and 8% in case of r_0, r_3 and r_2, respectively. However, there are values of θ where the subefficiency is larger than in case of contamination neighborhoods. In particular, the least favorable radii are larger. They lie in the interval $[0, 1.2]$. The corresponding results are plotted in Figures 4.31–4.33.

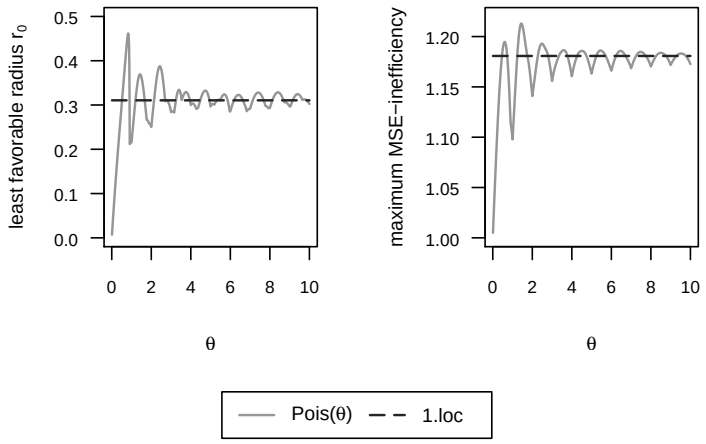

Figure 4.31: Least favorable radius r_0 and maximum MSE–inefficiency in case of total variation neighborhoods ($* = v$) and $\theta \in (0, 10]$.

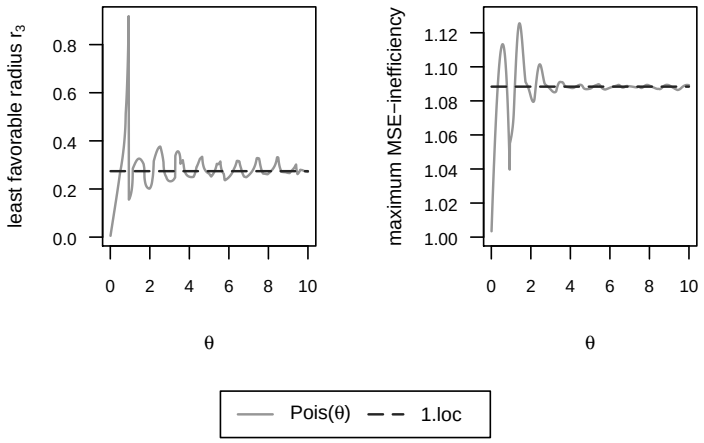

Figure 4.32: Least favorable radius r_3 and maximum MSE–inefficiency in case of total variation neighborhoods $(* = v)$ and $\theta \in (0, 10]$.

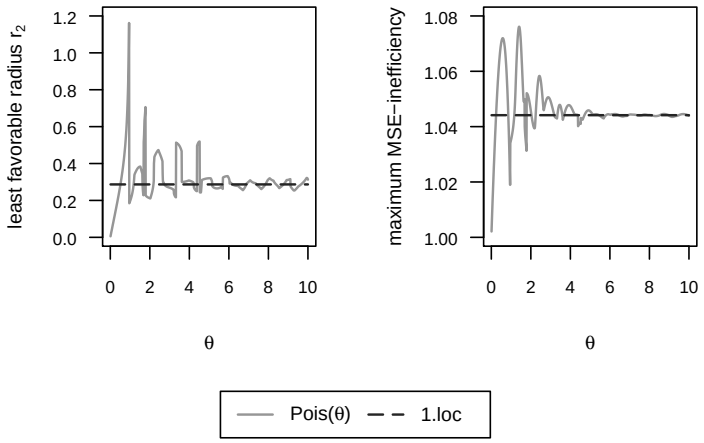

Figure 4.33: Least favorable radius r_2 and maximum MSE–inefficiency in case of total variation neighborhoods $(* = v)$ and $\theta \in (0, 10]$.

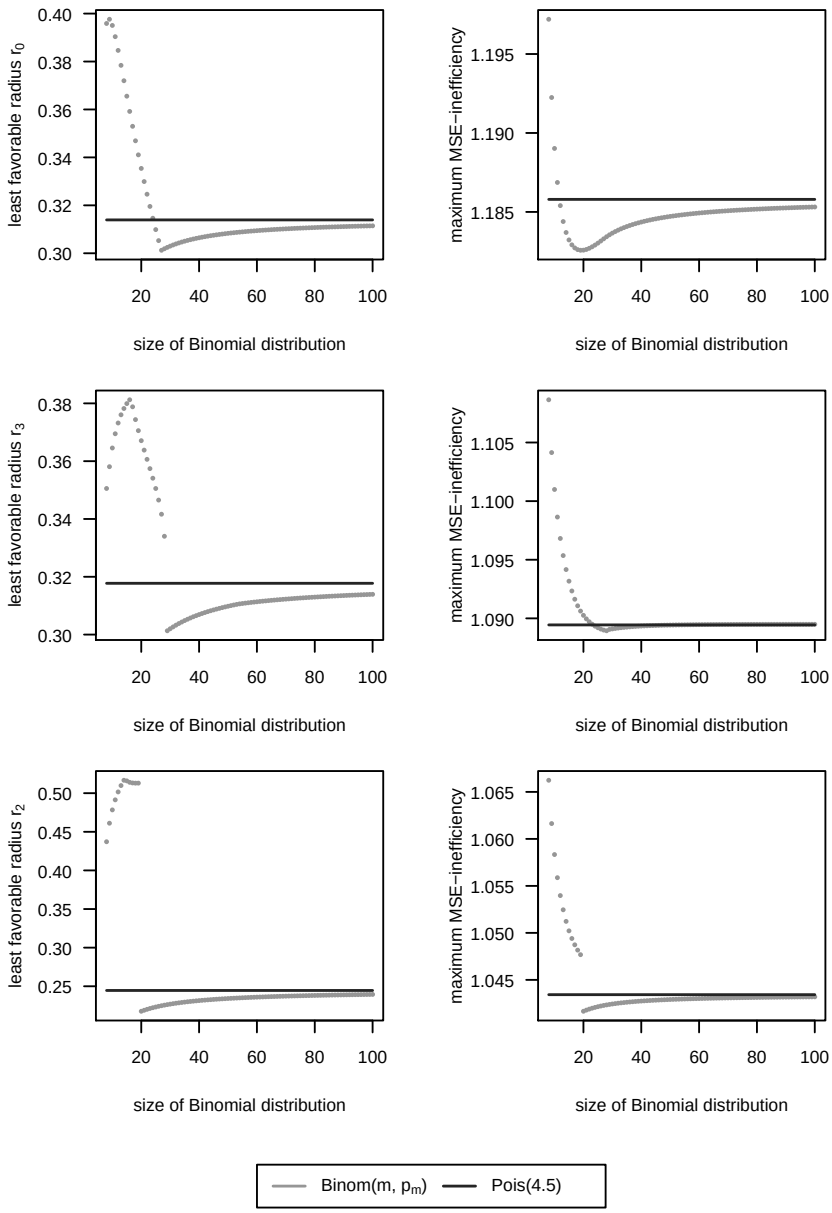

Figure 4.34: Poisson approximation of least favorable radii r_0, r_3, r_2 and maximum MSE–inefficiency for $m \in [8, 100]$, $p_m = \theta/m$ and $\theta = 4.5$ in case of total variation neighborhoods ($* = v$).

4.4 One-Step Construction

We can apply Lemma 2.3.6. That is, we can construct the optimally robust estimator as an one-step estimator at least in case of total variation neighborhoods. In case of contamination neighborhoods the centering constant z_r may be non-unique; confer Remark 4.2.1 (b). However, this only happens if the median of Λ is non-unique and $r \geq \bar{r}$. Furthermore, already for very small deviations ($\pm 1e-07$ or even $\pm 1e-08$) from the exact values of θ the numerical computations yield unique centering constants. Moreover, for $\mathrm{Pois}(\theta)$ there are only countable many values of $\theta \in (0,\infty)$ for which the median is non-unique, namely the solutions of the following equation in θ

$$\sum_{k=0}^{l} \frac{e^{-\theta}\theta^k}{k!} = 0.5 \qquad \text{for all } l \in \mathbb{N}_0 \tag{4.4.1}$$

Of course, it is equivalent to solve

$$\sum_{k=l+1}^{\infty} \frac{e^{-\theta}\theta^k}{k!} = 0.5 \qquad \text{for all } l \in \mathbb{N}_0 \tag{4.4.2}$$

where the left hand side is identical to the cumulative distribution function at θ of the Gamma distribution with scale parameter 1 and shape parameter $l+1$. That is, the solutions θ_l are unique for all $l \in \mathbb{N}_0$ as they coincide with the unique medians of the corresponding Gamma distributions. In this context Choi (1994) proves that $\theta_l - l \in (2/3, \log(2)]$ and Alm (2003) additionally shows that the sequence $(\theta_l - l)_{l\in\mathbb{N}_0}$ is decreasing from $\log(2)$ to $2/3$.
In view of the previous considerations, the non-uniqueness of the centering constant z_r is only a minor problem for the one-step construction of the corresponding optimally robust asymptotic estimators and we may neglect this non-uniqueness in practice.
As in the binomial model, we again use the minimum Kolmogorov(–Smirnov) distance estimator as our starting estimator. It has the necessary properties (strict and $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta) \supset \mathcal{U}_v(\theta) \supset \mathcal{U}_c(\theta)$) by Subsection 6.3.2 of Rieder (1994) and is well computable in this simple discrete model.

4.5 Implementation using `R`

The `R` package `ROptEst` described in Appendix D.3 provides the necessary `S4` classes and methods. In case of the Poisson model, we implemented the generating function `PoisFamily` which is called with some parameter `lambda=`θ ($\theta \in (0,\infty)$). The rest of the implementation including the calls to the methods for the computation of the optimal ICs, the lower case radii and the least favorable radii is completely analogous to the binomial model. This also holds for the robust estimation by means of the one-step construction. Therefore, we refer to Section 3.5 where the implementation of the binomial model is specified in detail.

Remark 4.5.1 (a) The computations are based on the slot `support` of $\mathrm{Pois}(\theta)$ which is determined using the `TruncQuantile` variable of the `distr` package whose default value is $1\mathrm{e}{-05}$. Hence, if we want to increase the precision of the computations, we have to decrease the value of this variable which can be done via the function `distroptions`; e.g.,

```
> distroptions("TruncQuantile", 1e-10)
```

(b) After installing our `R` bundle `RobASt` there is the `R` script `PoissonModel.R`, which contains some examples for the Poisson model, in the directory "`.../RHome/library/ROptEst/scripts/`" where `RHome` stands for the local home directory of `R`. ////

4.6 A Small Simulation Study

As in case of the binomial model (cf. Section 3.6), we conclude this chapter with the presentation of a small simulation study which demonstrates the use of our `R` package `ROptEst` (cf. Appendix D.3) and the need of robust estimation in case of the Poisson model.
We again consider three different situations. First, the data is generated by the ideal model $\mathrm{Pois}(4.5)$. We then replace 5% of the data by $\mathrm{I}_{\{12\}}$ in the following way

$$k \mapsto \big(\mathrm{Pois}(4.5)(\{k\}) - 0.05\big)_+ + 0.05\,\mathrm{I}_{\{12\}}(k) \qquad k \in \mathbb{Z} \tag{4.6.1}$$

i.e., mass is moved from the left hand side to 12 which is a possible but rare event ($\approx 0.16\%$). We regard this as a realistic situation, whereas in the third ("extreme") case we instead substitute these points by 100; i.e., we use

$$k \mapsto \big(\mathrm{Pois}(4.5)(\{k\}) - 0.05\big)_+ + 0.05\,\mathrm{I}_{\{100\}}(k) \qquad k \in \mathbb{Z} \tag{4.6.2}$$

We computed 1000 samples of size 100 and determined the mean (classical optimal) and the Kolmogorov(–Smirnov) minimum distance (ksMD) estimator. The ksMD estimator then serves as initial estimator for two robust one-step estimators. The first one-step estimator ($r = 0.5$) is based on the optimally robust IC for radius $r = 0.5$; i.e., the amount of contamination is known ($r/\sqrt{100} = 0.05$). The second one-step estimator ($r = r_0$) was calculated using the radius–minimax IC for the least favorable radius r_0; i.e., the radius is completely unknown. For a boxplot of the results see Figure 4.35.

Remark 4.6.1 Since the ideal model distribution is concentrated on $\mathbb{Z}$, moving mass from the left hand side to ∞ heuristically seems to have the largest effect on the estimation of θ – at least in case of the mean – and we approximate this by $\mathrm{I}_{\{100\}}$. However, as already encountered in Section 3.6, this does not hold for the ksMD estimator which now performs even better in the "extreme" than in the realistic situation. Hence, not surprisingly, this is also true for the robust estimators. ////

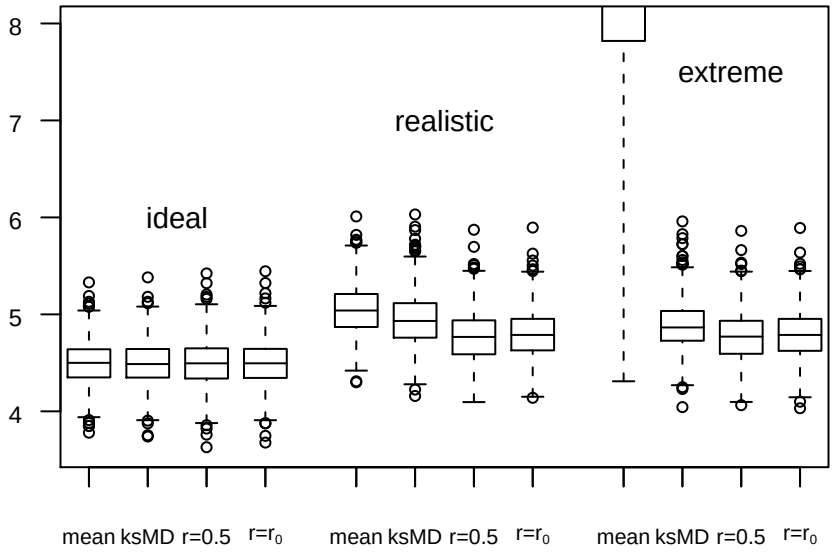

Figure 4.35: Boxplot for a small simulation study in case of total variation neighborhoods.

situation	mean	ksMD	$r = r_0$	$r = 0.5$
emp. ideal	4.745 ± 1.371	4.938 ± 1.428	5.256 ± 1.557	5.534 ± 1.637
as. ideal	4.500	—	5.334	5.670
realistic	36.810 ± 5.978	28.011 ± 5.782	14.613 ± 3.517	13.835 ± 3.512
extreme	2845.2 ± 443.6	21.253 ± 4.705	14.454 ± 3.487	13.770 ± 3.501
asymptotic	∞	—	13.563	13.336

Table 4.1: Empirical and asymptotic MSEs for a small simulation study in case of total variation neighborhoods.

As the numerical results in Table 4.1 show, where we additionally provide 95% confidence intervals based on the CLT, there are large differences in absolute values between the empirical MSEs of the considered estimators in the contaminated situations whereas the differences in the ideal model are only small.
This is also reflected by the corresponding MSE–inefficiencies. In the ideal case, the subefficiency of the ksMD and the two one-step estimator are about 5% (ksMD), 11% ($r = r_0$) and 17% ($r = 0.5$), respectively; confer Table 4.2. Whereas in

situation	mean	ksMD	$r = r_0$	$r = 0.5$
emp. ideal	1.000	1.041	1.108	1.166
as. ideal	1.000	—	1.185	1.260
realistic	2.661	2.025	1.056	1.000
extreme	206.6	1.543	1.050	1.000
asymptotic	∞	—	1.017	1.000

Table 4.2: MSE–inefficiencies for a small simulation study in case of total variation neighborhoods.

the contaminated samples the mean has a very large efficiency loss ($\gg 150\%$) and hence, can not be regarded as an appropriate estimator. Since, in addition, the computational effort in case of the robust estimators is not much larger than in case of the mean, we recommend to use the robust estimators in any situation. Moreover, we see that the empirical values and probably also the exact finite-sample values are rather larger than the asymptotic values. This agrees with the results obtained in Part V of this thesis and in Ruckdeschel and Kohl (2005) as well as with the results of the higher order studies of Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c) and Ruckdeschel (2005e).

Remark 4.6.2 **(a)** We do not give the asymptotic values for the ksMD estimator as we are not sure about its asymptotic distribution in this setup.

(b) In view of these results and the results obtained in Section 3.6, further studies might be of interest. For more details we refer to Remark 3.6.2. ////

Chapter 5

Exponential Scale and Gumbel Location Model

In Subsection 5.1.1 and Subsection 5.1.2 we briefly introduce the one-dimensional scale and the one-dimensional location model, respectively. We then show that there is a connection between certain scale and location models via the transformations $\pm \log|\cdot|$ which for instance holds for the exponential scale and the Gumbel location model. This connection entails a strong relationship between the corresponding optimal solutions; confer Subsection 5.1.3. In Section 5.2 we specify the optimally robust ICs for the exponential scale and the Gumbel location model. The numerical results for the least favorable radii and the corresponding inefficiencies for these two models are given in Section 5.3. We conclude this chapter with the verification of the sufficient conditions for the one-step construction of the corresponding optimally robust estimators (cf. Section 5.4) and describe how one can use our `R` package `ROptEst` in case of one-dimensional scale, respectively one-dimensional location models (cf. Section 5.5).

5.1 Introduction

5.1.1 One-Dimensional Scale Model

Consider the one-dimensional scale model

$$x = \sigma u \tag{5.1.1}$$

where $u \sim P$; i.e., the scale family is given by

$$\mathcal{P} = \{P_\sigma(dx) = P(dx/\sigma) \,|\, \sigma \in (0,\infty)\} \tag{5.1.2}$$

where P is some probability measure on $(\mathbb{R}, \mathbb{B})$. The Fisher information of scale is defined as

$$\mathcal{I}^{\text{sc}} := \sup_{\varphi \in \mathcal{D}_1} \frac{\left(\int u\varphi' \, dP\right)^2}{\int \varphi^2 \, dP} \qquad \varphi \neq 0 \quad \text{a.e. } P \tag{5.1.3}$$

where $\mathcal{D}_1$ is set of all functions $\varphi\colon \bar{\mathbb{R}} \to \mathbb{R}$ that are infinitely differentiable on $\mathbb{R}$, continuous on $\bar{\mathbb{R}}$ and whose derivative φ' has compact support. We identify $\bar{\mathbb{R}}$ with $[0,1]$ via the smooth isometry $u \mapsto \frac{e^u}{e^u+1}$.

Proposition 5.1.1 *The following two statements are equivalent:*

(a) $\mathcal{I}^{\mathrm{sc}} < \infty$

(b) *(i)* P *is absolutely continuous on* $\mathbb{R} \setminus \{0\}$ *with density* f

(ii) uf *is absolutely continuous*

(iii) $uf'/f \in L_2(P)$

In both cases we have

$$\mathcal{I}^{\mathrm{sc}} = \int \Lambda_f^{\mathrm{sc}}(u)^2 f(u)\, du \tag{5.1.4}$$

where

$$\Lambda_f^{\mathrm{sc}}(u) = \begin{cases} -uf'(u)/f(u) - 1 & \text{if } u \neq 0 \\ 0 & \text{if } u = 0 \end{cases} \tag{5.1.5}$$

Proof Ruckdeschel (2005d), Theorem 4.7. ////

The scale model (5.1.2) stays invariant under the group of scale transformations $g_\sigma(u) = \sigma u$ in the sense, that for all $\sigma \in (0,\infty)$, P_σ is the image measure $g_\sigma(P)$. We define the scale scores at $\sigma \in (0,\infty)$

$$\Lambda_\sigma(x) := \sigma^{-1}\Lambda_f^{\mathrm{sc}}(x/\sigma) = \sigma^{-1}\Lambda_1 \circ g_\sigma^{-1}(x) \tag{5.1.6}$$

where $g_\sigma^{-1} = g_{\sigma^{-1}}$. Thus, the Fisher information of scale is

$$\mathcal{I}_\sigma = \mathrm{E}_\sigma\left(\Lambda_\sigma\right)^2 = \sigma^{-2}\,\mathrm{E}_1\left(\Lambda_f^{\mathrm{sc}}\right)^2 = \sigma^{-2}\mathcal{I}^{\mathrm{sc}} \tag{5.1.7}$$

Proposition 5.1.2 *Assume* $\mathcal{I}^{\mathrm{sc}} < \infty$. *The scale model (5.1.2) is* L_2 *differentiable at* $\sigma \in (0,\infty)$ *with* L_2 *derivative* Λ_σ *and Fisher information* $\mathcal{I}_\sigma$ *given by (5.1.6) and (5.1.7), respectively.*

Proof Swensen (1980) Chapter 2, Section 3. ////

Remark 5.1.3 By (5.1.6) and (5.1.7), Theorem 2.4.1 applies to the one-dimensional scale model. Hence, we obtain the following relations between the Lagrange multipliers contained in the optimally robust ICs at σ and the Lagrange multipliers contained in the optimally robust ICs at $\sigma = 1$,

$$A_\sigma = \sigma^2 A_1 \qquad b_\sigma = \sigma b_1 \qquad c_\sigma = \sigma c_1 \tag{5.1.8}$$

and we can choose

$$a_\sigma = \sigma a_1 \tag{5.1.9}$$

Moreover,

$$\omega_{*,\sigma}^{\min} = \sigma^2 \omega_{*,1}^{\min} \qquad * = c, v \tag{5.1.10}$$

and

$$\operatorname{maxMSE}(\tilde{\eta}_{*,\sigma,r_0}, r) = \sigma^2 \operatorname{maxMSE}(\tilde{\eta}_{*,1,r_0}, r) \qquad * = c, v \tag{5.1.11}$$

respectively

$$\operatorname{relMSE}(\tilde{\eta}_{*,\sigma,r_0}, r) = \operatorname{relMSE}(\tilde{\eta}_{*,1,r_0}, r) \qquad * = c, v \tag{5.1.12}$$

for all $r_0, r \in [0, \infty]$. ////

5.1.2 One-Dimensional Location Model

Second, consider the one-dimensional location model

$$y = \mu + v \tag{5.1.13}$$

where $v \sim Q$; i.e., the location family is given by

$$\mathcal{Q} = \{Q_\mu(dy) = Q(dy - \mu) \,|\, \mu \in \mathbb{R}\} \tag{5.1.14}$$

where Q is some probability measure on $(\mathbb{R}, \mathbb{B})$. The Fisher information of location is defined as

$$\mathcal{I}^{\text{loc}} := \sup_{\varphi \in \mathcal{C}_c^\infty} \frac{\left(\int \varphi' \, dQ\right)^2}{\int \varphi^2 \, dQ} \qquad \varphi \neq 0 \quad \text{a.e. } Q \tag{5.1.15}$$

where $\mathcal{C}_c^\infty = \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$ is set of all infinitely differentiable functions $\varphi\colon \mathbb{R} \to \mathbb{R}$ having compact support.

Proposition 5.1.4 *The following two statements are equivalent:*

(i) $\mathcal{I}^{\text{loc}} < \infty$

(ii) Q *has an absolutely continuous density* h *with* $\int \Lambda_h^{\text{loc}}(v)^2 h(v) \, dv < \infty$ *where* $\Lambda_h^{\text{loc}}(v) = -(h'/h)(v)$.

In both cases we have $\mathcal{I}^{\text{loc}} = \int \Lambda_h^{\text{loc}}(v)^2 h(v) \, dv$.

Proof Huber (1981), Theorem 4.2. ////

Remark 5.1.5 Huber's proof remains unchanged if we use the smaller set $\mathcal{C}_c^\infty$ which is also dense in $L_2(Q)$ instead of the set $\mathcal{C}_c^1$ of all continuously differentiable functions $\psi\colon \mathbb{R} \to \mathbb{R}$ having compact support. In addition, Ruckdeschel (2005d) shows that the functional $A\varphi = -\int \varphi' dP$ $(\varphi \in \mathcal{C}_c^\infty)$ which Huber introduces in his proof is well-defined (consequence of Lemma 4.12 (ibid.)). ////

The location model stays invariant under the group of location transformations $g_\mu(v) = \mu + v$ in the sense, that for all $\mu \in \mathbb{R}$, Q_μ is the image measure $g_\mu(Q)$. We define the location scores at $\mu \in \mathbb{R}$

$$\Lambda_\mu(y) := \Lambda_h^{\mathrm{loc}}(y - \mu) = \Lambda_0 \circ g_{-\mu}(y) \tag{5.1.16}$$

Hence, the Fisher information of location is

$$\mathcal{I}_\mu = \mathrm{E}_\mu \left(\Lambda_\mu\right)^2 = \mathrm{E}_0 \left(\Lambda_h^{\mathrm{loc}}\right)^2 = \mathcal{I}^{\mathrm{loc}} \tag{5.1.17}$$

Proposition 5.1.6 *Assume $\mathcal{I}^{\mathrm{loc}} < \infty$. The location model (5.1.14) is L_2 differentiable at $\mu \in \mathbb{R}$ with L_2 derivative Λ_μ and Fisher information $\mathcal{I}_\mu$.*

Proof Rieder (1994), Proposition 2.4.1. ////

Remark 5.1.7 By (5.1.16) and (5.1.17), Theorem 2.4.1 also applies to the one-dimensional location model. Therefore, there are the following relations between the Lagrange multipliers contained in the optimally robust ICs at μ and the Lagrange multipliers contained in the optimally robust ICs at $\mu = 0$,

$$A_\mu = A_0 \qquad b_\mu = b_0 \qquad c_\mu = c_0 \tag{5.1.18}$$

and we can choose

$$a_\mu = a_0 \tag{5.1.19}$$

Moreover,

$$\omega_{*,\mu}^{\min} = \omega_{*,0}^{\min} \qquad * = c, v \tag{5.1.20}$$

and

$$\mathrm{maxMSE}(\tilde{\eta}_{*,\mu,r_0}, r) = \mathrm{maxMSE}(\tilde{\eta}_{*,0,r_0}, r) \qquad * = c, v \tag{5.1.21}$$

respectively

$$\mathrm{relMSE}(\tilde{\eta}_{*,\mu,r_0}, r) = \mathrm{relMSE}(\tilde{\eta}_{*,0,r_0}, r) \qquad * = c, v \tag{5.1.22}$$

for all $r_0, r \in [0, \infty]$. ////

5.1.3 Connection between One-Dimensional Scale and One-Dimensional Location

Let P be some absolutely continuous probability measure on $(\mathbb{R}, \mathbb{B})$ with finite Fisher information of scale; i.e., $\mathcal{I}^{\mathrm{sc}} < \infty$. Hence, P has a density f such that uf is absolutely continuous and $\mathcal{I}^{\mathrm{sc}} = \int \Lambda_f^{\mathrm{sc}}(u)^2 f(u)\, du$ with $\Lambda_f^{\mathrm{sc}}(u) = -u(f'/f) - 1$; confer Proposition 5.1.1. In addition, assume one of the following three conditions holds

(i) P is symmetric

(ii) P is concentrated on $(0,\infty)$

(iii) P is concentrated on $(-\infty,0)$

Following the idea of Section 5.6 in Huber (1981), we apply the transformation $\pm\log|\cdot|$ to the one-dimensional scale model (5.1.2). This leads us to the one-dimensional location model

$$y=\mu+v \tag{5.1.23}$$

where $v\sim Q$ and $Q(dv)=h(v)dv$ with

$$h(v)=\begin{cases}2e^{\pm v}f(e^{\pm v}) & \text{if (i)}\\ e^{\pm v}f(e^{\pm v}) & \text{if (ii)}\\ e^{\pm v}f(-e^{\pm v}) & \text{if (iii)}\end{cases} \tag{5.1.24}$$

The corresponding location scores read

$$\Lambda_h^{\mathrm{loc}}(v)=\begin{cases}\pm\Lambda_f^{\mathrm{sc}}(e^{\pm v}) & \text{if (i) or (ii)}\\ \pm\Lambda_f^{\mathrm{sc}}(-e^{\pm v}) & \text{if (iii)}\end{cases} \tag{5.1.25}$$

i.e., $\Lambda_f^{\mathrm{sc}}=\pm\Lambda_h^{\mathrm{loc}}\circ(\pm\log|\cdot|)$. By an application of the transformation formula and using the symmetry in case (i) (i.e., $f(-u)=f(u)$ and $\Lambda_f^{\mathrm{sc}}(-u)=\Lambda_f^{\mathrm{sc}}(u)$), we obtain

$$\int\Lambda_h^{\mathrm{loc}}(v)^2h(v)\,dv=\begin{cases}\int_{-\infty}^{\infty}\Lambda_f^{\mathrm{sc}}(u)^2f(u)\,du & \text{if (i)}\\ \int_0^{\infty}\Lambda_f^{\mathrm{sc}}(u)^2f(u)\,du & \text{if (ii)}\\ \int_{-\infty}^{0}\Lambda_f^{\mathrm{sc}}(u)^2f(u)\,du & \text{if (iii)}\end{cases}=\mathcal{I}^{\mathrm{sc}}<\infty \tag{5.1.26}$$

Thus, the location model (5.1.23) is L_2 differentiable with L_2 derivative

$$\Lambda_\mu(y):=\Lambda_h^{\mathrm{loc}}(y-\mu)=\begin{cases}\pm\Lambda_f^{\mathrm{sc}}(e^{\pm(y-\mu)}) & \text{if (i) or (ii)}\\ \pm\Lambda_f^{\mathrm{sc}}(-e^{\pm(y-\mu)}) & \text{if (iii)}\end{cases} \tag{5.1.27}$$

and Fisher information of location

$$\mathcal{I}_\mu=\mathrm{E}_\mu\left(\Lambda_\mu\right)^2=\mathrm{E}_0\left(\Lambda_h^{\mathrm{loc}}\right)^2=\mathcal{I}^{\mathrm{sc}} \tag{5.1.28}$$

Remark 5.1.8 By (5.1.27) and (5.1.28), Theorem 2.4.1 at once applies to the cases (ii) and (iii). In case (i), we additionally make use of the symmetry of P which entails $f(-u)=f(u)$ and $\Lambda_f^{\mathrm{sc}}(-u)=\Lambda_f^{\mathrm{sc}}(u)$. Thus, in all three cases the Lagrange multipliers $(A_\mu,\ a_\mu,\ b_\mu,\ c_\mu)$ contained in the optimally robust ICs in case of one-dimensional location can easily be obtained from the Lagrange multipliers $(A_\sigma,\ a_\sigma,\ b_\sigma,\ c_\sigma)$ included in the optimally robust IC in case of one-dimensional scale and $\sigma=1$. More precisely, we get

$$A_\mu=\sigma^{-2}A_\sigma\qquad b_\mu=\sigma^{-1}b_\sigma\qquad c_\mu=\sigma^{-1}\begin{cases}c_\sigma & \text{if } +\log|\cdot|\\ -(b_\sigma+c_\sigma) & \text{if } -\log|\cdot|\end{cases} \tag{5.1.29}$$

and we can choose

$$a_\mu = \pm\sigma^{-1} a_\sigma \tag{5.1.30}$$

Furthermore,

$$\omega^{\min}_{*,\mu} = \sigma^{-2}\omega^{\min}_{*,\sigma} \qquad * = c, v \tag{5.1.31}$$

and

$$\operatorname{maxMSE}(\tilde{\eta}^{\mathrm{loc}}_{*,\mu,r_0}, r) = \sigma^{-2}\operatorname{maxMSE}(\tilde{\eta}^{\mathrm{sc}}_{*,\sigma,r_0}, r) \qquad * = c, v \tag{5.1.32}$$

respectively

$$\operatorname{relMSE}(\tilde{\eta}^{\mathrm{loc}}_{*,\mu,r_0}, r) = \operatorname{relMSE}(\tilde{\eta}^{\mathrm{sc}}_{*,\sigma,r_0}, r) \qquad * = c, v \tag{5.1.33}$$

for all $r_0, r \in [0,\infty]$. Thus, in view of Section 2.4, these location and scale models (experiments) may be regarded as equivalent. ////

We now specify a few examples.

Example 5.1.9 **(a)** On page 119 Huber (1981) uses the transformation $\log|\cdot|$ to transform the normal scale model into a location model. This way he obtains a location family with density

$$h_\mu(y) = \sqrt{\tfrac{2}{\pi}}\,\exp\left\{y - \mu - \tfrac{1}{2}e^{2(y-\mu)}\right\} \tag{5.1.34}$$

Since h_μ seems to belong to no well known distribution family, we will not study, respectively use this connection further in this thesis.

(b) In allusion to the normal distribution $\mathcal{N}(\zeta, \alpha^2)$ the density of the (three-parameter) lognormal distribution is usually parameterized as

$$f_{\mu,\zeta,\alpha}(x) = \frac{1}{\sqrt{2\pi}\,\alpha(x-\mu)}\exp\left\{-\tfrac{1}{2}(\log(x-\mu)-\zeta)^2/\alpha^2\right\}\mathrm{I}(x \geq \mu) \tag{5.1.35}$$

with $\mu, \zeta \in \mathbb{R}$, $\alpha \in (0,\infty)$; confer Section 14.1 of Johnson et al. (1994). However, we prefer to work with

$$f_{\mu,\sigma,\alpha}(x) = \frac{1}{\sqrt{2\pi}\,\alpha(x-\mu)}\exp\left\{-\tfrac{1}{2}\left[\log((x-\mu)/\sigma)\right]^2/\alpha^2\right\}\mathrm{I}(x \geq \mu) \tag{5.1.36}$$

($\sigma = \exp(\zeta) \in (0,\infty)$) which in our opinion makes the location, scale and shape parameter of the lognormal distribution clearer. Hence, in case $\mu = 0$, $\alpha = 1$ known, the scale scores reads

$$\Lambda^{\mathrm{sc}}_f(x) = \log(x) \tag{5.1.37}$$

and the transformation $\log|\cdot|$ leads us to the well-known normal location family with density

$$h_\mu(y) = e^{y-\mu} f_{e^\mu}(e^y) = \frac{1}{\sqrt{2\pi}}\exp\left\{\tfrac{1}{2}(y-\mu)^2\right\} \tag{5.1.38}$$

and location scores

$$\Lambda^{\mathrm{loc}}_h(y) = \Lambda^{\mathrm{sc}}_f(e^y) = y \tag{5.1.39}$$

By symmetry of $\mathcal{L}(\Lambda_f^{\mathrm{sc}}) = \mathcal{N}(0,1)$, the optimally robust ICs for total variation neighborhoods $(* = v)$ and radius r coincide with the optimally robust ICs for contamination neighborhoods $(* = c)$ and radius $2r$. Therefore, the Lagrange multipliers included in the optimally robust ICs in case of the lognormal scale model and $* = c, v$ can easily be taken from the normal location model via the relations specified in Remark 5.1.8. Moreover, the least favorable radii and the corresponding MSE–inefficiencies for the lognormal scale model can be read off from Section 5.1 of Rieder et al. (2001); confer also Table 5.1. We use this connection in Subsection 9.3.3.2 where we consider the ARCH(1) model with lognormal innovations as one example.

relMSE $(\tilde{\eta}_{c,\infty}, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1.571	1.181	1.088	1.044	0.621	0.548	0.574

Table 5.1: Least favorable radius and maximum MSE–inefficiency for lognormal scale and normal location in case of contamination and total variation neighborhoods. [In case of total variation neighborhoods the least favorable radii are $\frac{1}{2}r_0$, $\frac{1}{2}r_3$ and $\frac{1}{2}r_2$, respectively.]

(c) Of course, the exponential model is a special case of the Gamma model treated in Chapter 6. However, the exponential model with density

$$f_\sigma(x) = \sigma^{-1} \exp\big(-x/\sigma\big)\, \mathrm{I}(x \geq 0) \tag{5.1.40}$$

can also be considered as a scale model with scale scores

$$\Lambda_f^{\mathrm{sc}}(x) = x - 1 \tag{5.1.41}$$

Moreover, via the transformation $-\log|\cdot|$ it is connected to the Gumbel location family which has density

$$h_\mu(y) = e^{-(y-\mu)} f_{e^\mu}(e^{-y}) = \exp(-(y-\mu) - e^{-(y-\mu)}) \tag{5.1.42}$$

and location scores

$$\Lambda_h^{\mathrm{loc}}(y) = -\Lambda_f^{\mathrm{sc}}(e^{-y}) = 1 - e^{-y} \tag{5.1.43}$$

Thus, we study the exponential model separately from the Gamma model and use it to demonstrate the effect of the derived connection between certain scale and location models; confer Section 5.2 and Section 5.3. Moreover, we also make use of this connection in case of the AR(1) and MA(1) model where we assume innovations that have a Gumbel distribution in one of the numerical examples; confer Subsubsection 9.3.3.1. ////

Remark 5.1.10 (a) In contrast to Huber (1981) (cf. Sections 5.6 and 5.7), who uses the connection mentioned in Example 5.1.9 (a) to find the distribution which minimizes Fisher information in case of the one-dimensional normal scale model

and contamination neighborhoods with fixed size ε, this is not needed for the determination of the MSE optimal ICs in the framework of shrinking neighborhoods. In our setup, we at once obtain the optimally robust IC for any L_2 differentiable scale family and there are no further limitations like symmetry or $\varepsilon \leq 0.04$.

(b) By considering lognormal scale, the approach of Section 5.6 in Huber (1981) passes through without limitations on the fixed size ε of the symmetric contamination neighborhoods. More precisely, the results of normal location apply; confer Example 5.1.9 (b).

(c) By Remark 5.1.8 and (5.1.25) we even obtain

$$\tilde{\eta}_1^{\mathrm{sc}}(\cdot) = \pm\tilde{\eta}_0^{\mathrm{loc}} \circ (\pm \log|\cdot|) \tag{5.1.44}$$

i.e., the MSE optimal solutions for $\sigma = 1$ and $\mu = 0$ transform analogously to the corresponding models; confer also Section 5.2.

(d) The Gumbel distribution is also known as extreme value distribution of type I and the connection between the exponential scale and the Gumbel location model is a special case of the corresponding connection between the Weibull and the Gumbel distribution, respectively between the extreme value distributions of type II and III and the Gumbel distribution; confer Section 22.2 of Johnson et al. (1995).

(e) Our algorithm for the determination of the optimally robust ICs in case of a one-dimensional parameter is based on the distribution of the L_2 derivative; confer Appendix D. Hence, the computations performed for such pairs of location and scale models are identical, respectively identical except for the sign. ////

5.2 Optimally Robust Influence Curves

In this section, the solutions to the asymptotic MSE problems (1.3.5) given in Theorem 1.3.11 are specified for the exponential scale and the Gumbel location model; confer Example 5.1.9 (c). Because of the relations between the Lagrange multipliers contained in the optimally robust ICs mentioned in Remark 5.1.3, Remark 5.1.7 and Remark 5.1.8, we can restrict our considerations to $\sigma = 1$, respectively $\mu = 0$ and we omit the subscript 1, respectively 0.

5.2.1 Contamination Neighborhoods

We consider the exponential scale and the Gumbel location model first with infinitesimal contamination neighborhoods (1.2.4) and given $D \in \mathbb{R} \setminus \{0\}$. The optimally robust ICs $\tilde{\eta}_{c,r}^{\mathrm{Exp}}$ and $\tilde{\eta}_{c,r}^{\mathrm{Gumbel}}$ given by Theorem 1.3.7 (a) and Theorem 1.3.11 (b) can be rewritten as

$$\tilde{\eta}_{c,r}^{\mathrm{Exp}}(x) = A_r(x - z_r) \min\left\{1, \frac{c_r}{|x - z_r|}\right\} \tag{5.2.1}$$

respectively

$$\tilde{\eta}_{c,r}^{\mathrm{Gumbel}}(y) = -A_r(e^{-y} - z_r) \min\left\{1, \frac{c_r}{|e^{-y} - z_r|}\right\} \tag{5.2.2}$$

Since $\mathcal{L}_{\mathrm{Gumbel}(0,1)}(e^{-y}) = \mathrm{Exp}\,(1)$, the following representations hold for both cases

$$0 = \int_0^\infty (x - z_r) \min\Big\{1,\ \frac{c_r}{|x - z_r|}\Big\} e^{-x}\, dx \tag{5.2.3}$$

$$D = A_r \int_0^\infty |x - z_r| \min\big\{|x - z_r|,\ c_r\big\} e^{-x}\, dx \tag{5.2.4}$$

and

$$r^2\, c_r = \int_0^\infty \big(|x - z_r| - c_r\big)_+ e^{-x}\, dx \tag{5.2.5}$$

In case $r = \infty$, we can read off from Remark 1.3.8

$$\tilde\eta^{\mathrm{Exp}}_{c,\infty}(x) = \omega_c^{\min} \operatorname{sign}(D) \operatorname{sign}(x - M) \tag{5.2.6}$$

respectively

$$\tilde\eta^{\mathrm{Gumbel}}_{c,\infty}(y) = -\omega_c^{\min} \operatorname{sign}(D) \operatorname{sign}(e^{-y} - M) \tag{5.2.7}$$

with $M = \mathrm{med}\big(\mathrm{Exp}\,(1)\big) = \log(2)$ and minimum bias

$$\omega_c^{\min} = |D| \Big(\int_0^\infty |x - M| e^{-x}\, dx \Big)^{-1} = |D| \log(2)^{-1} \tag{5.2.8}$$

For a plot of the optimally robust ICs for different values of r see Figure 5.1.

Remark 5.2.1 **(a)** Since ICs are defined with respect to the ideal model, the optimally robust IC $\tilde\eta^{\mathrm{Exp}}_{c,r}$ in case of the exponential scale model is defined on $[0, \infty)$. However, if we consider neighborhoods about the ideal model, the members of these neighborhoods may have any support $T \subset \mathbb{R}$. Hence, we may want to extend $\tilde\eta^{\mathrm{Exp}}_{c,r}$ to $\mathbb{R} \setminus [0, \infty)$. In view of the construction of the corresponding optimally robust estimator one has to choose an extension such that $|\tilde\eta^{\mathrm{Exp}}_{c,r}| \le b_r = A_r c_r$ on $\mathbb{R} \setminus [0, \infty)$. Otherwise the bias would increase if we pass over from the neighborhood submodel to full neighborhoods. Moreover, in view of Theorem 2.3.3 and Lemma 2.3.6 we extend $\tilde\eta_{c,r}$ to $(-\infty, 0)$ simply by regarding Λ_f^{sc} as function on $\mathbb{R}$.

(b) In this simple one-dimensional setup, the computation of z_r and c_r is independent of D which only occurs in equation (5.2.4). Hence, we can compute A_r, z_r and c_r for $D = 1$ and then replace A_r by DA_r. Note, that $D = D_\sigma$, respectively $D = D_\mu$; i.e., D may depend on σ, respectively μ. Nevertheless, the relations between the Lagrange multipliers contained in the optimally robust ICs mentioned in Remark 5.1.3, Remark 5.1.7 and Remark 5.1.8 remain valid in case $D_\sigma \ne 1$, respectively $D_\mu \ne 1$. We only have to modify the relations for the standardizing constants A which now read

$$A_{\sigma,r} = D_\sigma \sigma^2 A_{1,r} \qquad \text{resp.} \qquad A_{\mu,r} = D_\mu A_{0,r} \tag{5.2.9}$$

(c) The lower case solution $\tilde\eta^{\mathrm{Exp}}_{c,\infty}$ corresponds to the standardized median introduced by Gather and Schultze (1999). ////

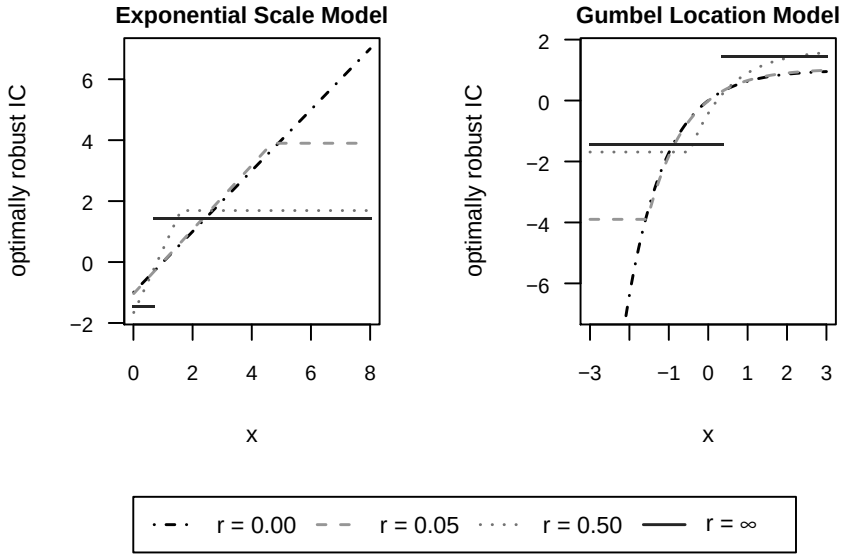

Figure 5.1: Optimally Robust ICs for $\mathrm{Exp}(1)$ and $\mathrm{Gumbel}(0,1)$ in case of contamination neighborhoods with radius $r = 0, 0.05, 0.50, \infty$.

5.2.2 Total Variation Neighborhoods

Second, we consider the exponential scale and the Gumbel location model with infinitesimal total variation neighborhoods (1.2.5) and given $D \in \mathbb{R}\setminus\{0\}$. The optimally robust IC $\tilde{\eta}_{v,r}^{\mathrm{Exp}}$, respectively $\tilde{\eta}_{v,r}^{\mathrm{Gumbel}}$ supplied by Theorem 1.3.9 (a) and Theorem 1.3.11 (c) can be rewritten as

$$\tilde{\eta}_{v,r}^{\mathrm{Exp}}(x) = A_r\big[g_r \vee x \wedge (g_r + c_r)\big] \tag{5.2.10}$$

respectively

$$\tilde{\eta}_{v,r}^{\mathrm{Gumbel}}(y) = -A_r\big[g_r \vee e^{-y} \wedge (g_r + c_r)\big] \tag{5.2.11}$$

and the following representations hold

$$0 = \int_0^\infty (g_r - x)_+ e^{-x}\,dx - \int_0^\infty (x - (g_r + c_r))_+ e^{-x}\,dx \tag{5.2.12}$$

$$D = A_r \int_0^\infty x\big[g_r \vee x \wedge (g_r + c_r)\big] e^{-x}\,dx \tag{5.2.13}$$

and

$$r^2\, c_r = \int_0^\infty (x - (g_r + c_r))_+ e^{-x}\,dx \tag{5.2.14}$$

In case $r = \infty$, we can read off from Theorem 1.3.9 (b)

$$\tilde{\eta}_{v,\infty}^{\mathrm{Exp}}(x) = \omega_v^{\min} \operatorname{sign}(D)\big[P(x \le 1) - \mathrm{I}(x \le 1)\big] \tag{5.2.15}$$

respectively

$$\tilde{\eta}_{v,\infty}^{\mathrm{Gumbel}}(y) = \omega_v^{\min} \operatorname{sign}(D)\big[\mathrm{I}(e^{-y} \le 1) - P(e^{-y} \le 1)\big] \tag{5.2.16}$$

where $P(e^{-y} \le 1) = P(x \le 1) = 1 - e^{-1}$ and the minimum bias is

$$\omega_v^{\min} = |D|\Big(\int_1^\infty x e^{-x}\, dx\Big)^{-1} = |D|\, e^1 \tag{5.2.17}$$

For a plot of the optimally robust ICs for different values of r see Figure 5.2.

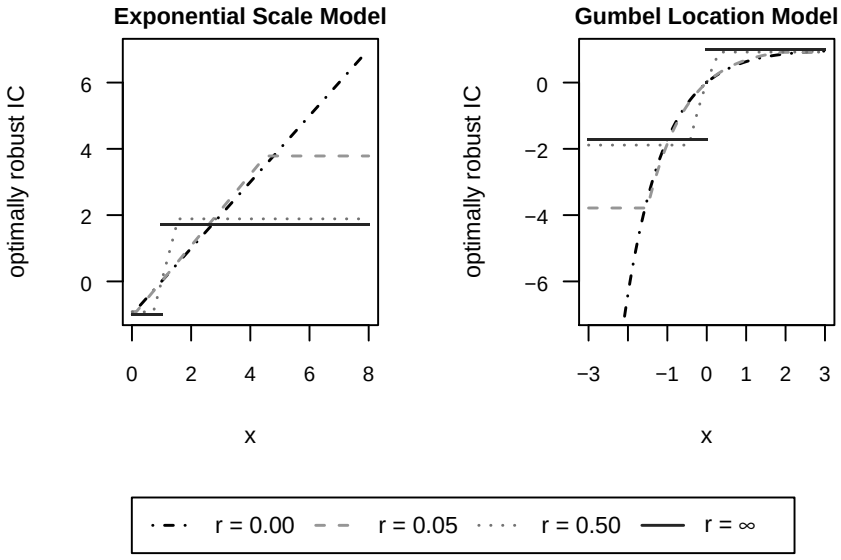

Figure 5.2: Optimally Robust ICs for $\mathrm{Exp}\,(1)$ and $\mathrm{Gumbel}\,(0,1)$ in case of total variation neighborhoods with radius $r = 0, 0.05, 0.50, \infty$.

Remark 5.2.2 **(a)** The optimally robust IC $\tilde{\eta}_{c,r}$ for the exponential scale model can be extended to $\mathbb{R}$ analogously to the case of contamination neighborhoods; confer Remark 5.2.1 (a).

(b) With the modification given in (5.2.9), the relations between the Lagrange multipliers contained in the optimally robust ICs mentioned in Remark 5.1.3, Remark 5.1.7 and Remark 5.1.8 remain valid in case $D_\sigma \neq 1$, respectively $D_\mu \neq 1$;

confer also Remark 5.2.1 (b). We again only have to modify the relations for the standardizing constants A which now read

$$A_{\sigma,r} = D_\sigma \sigma^2 A_{1,r} \qquad \text{resp.} \qquad A_{\mu,r} = D_\mu A_{0,r} \tag{5.2.18}$$

////

5.3 Least Favorable Radius

In this section we specify the least favorable radii r_0, r_2 and r_3 for the exponential scale and the Gumbel location model. Since the MSE–inefficiencies are identical in these two models (cf. Remark 5.1.8), the least favorable radii and the corresponding maximum MSE–inefficiencies coincide.

5.3.1 Contamination Neighborhoods

The least favorable radii and the corresponding MSE–inefficiencies for the exponential scale and the Gumbel location model are given in Table 5.2. In case the radius is completely unknown, we do not lose more than about 38% efficiency with our radius–minimax estimator. Moreover, if we have little knowledge about the radius we can reduce this loss to about 17% or even 8%.

$\text{relMSE}\,(\tilde{\eta}_{c,\infty}, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
2.081	1.381	1.167	1.082	0.495	0.418	0.436

Table 5.2: Least favorable radius and maximum MSE–inefficiency in case of contamination neighborhoods.

5.3.2 Total Variation Neighborhoods

In most models considered so far, the MSE–inefficiencies are smaller in case of total variation neighborhoods. This also holds in case of the exponential scale and the Gumbel location model; confer Table 5.3. In any case, we do not lose more than about 22% efficiency with our radius–minimax estimator and only little knowledge about the radius reduces this to about 10% or even 5%.

$\text{relMSE}\,(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1.718	1.222	1.103	1.051	0.285	0.257	0.267

Table 5.3: Least favorable radius and maximum MSE–inefficiency in case of total variation neighborhoods.

5.4 One-Step Construction

If the considered scale, respectively location family is an exponential family of full rank and the assumptions of Lemma 2.3.6 hold, we can construct the optimally robust estimator as an one-step estimator. This is true for the exponential scale family where $\zeta(\sigma) = -\sigma^{-1}$, $\beta(\sigma) = \log\sigma$, $T(x) = x$ and $h(x) = 1$ as well as for the Gumbel location family where $\zeta(\mu) = -e^{\mu}$, $\beta(\mu) = -\mu$, $T(y) = e^{-y}$ and $h(y) = e^{-y}$. In addition, the median of Λ_σ, respectively Λ_μ is always unique. Thus, we can use the one-step construction with a suitable (i.e., strict and $\sqrt{n}$ consistent on $\mathcal{U}_v(\theta) \supset \mathcal{U}_c(\theta)$) initial estimator in case of total variation as well as contamination neighborhoods. One candidate is again the Kolmogorov(–Smirnov) minimum distance estimator which has these properties (cf. Subsection 6.3.2 of Rieder (1994)) and is well computable in these simple one-dimensional models. In addition, we can use the median and the MAD; confer Subsection 2.3.4. Consequently, the standardized median introduced by Gather and Schultze (1999) may serve as starting estimator in the exponential scale model.

Remark 5.4.1 **(a)** The one-step construction can also be applied in case of normal location and normal scale which are exponential families of full rank with sufficiently smooth parameter functions α and β; confer Beispiel 1.151 of Witting (1985). In case of lognormal scale with parameter $\mu = 0$ and $\alpha = 1$, we obtain $\zeta(\sigma) = \log\sigma$, $\beta(\sigma) = 1/2(\log\sigma)^2$, $T(x) = \log x$ and $h(x) = \frac{1}{\sqrt{2\pi}\,x}\exp\{-1/2(\log x)^2\}\,\mathrm{I}(x \geq 0)$; i.e., again the assumptions of Lemma 2.3.6 hold.

(b) We conjecture that also the estimators proposed by Rousseeuw and Croux (1993) could be possible starting estimators. However, we do not know anything about their $\sqrt{n}$ consistency on $\mathcal{U}_c(\theta)$ and $\mathcal{U}_v(\theta)$, respectively. ////

5.5 Implementation using R

For a detailed description of the R package ROptEst we refer to Appendix D.3. To make some common models easier to handle, we implemented the generating functions given in Table 5.4. Objects can be instantiated completely analogous to

class	parameter
ExpScaleFamily	$\sigma = 1/$rate
GumbelLocationFamily	$\mu =$ loc
NormLocationFamily	$\mu =$ mean
NormScaleFamily	$\sigma =$ sd
LnormScaleFamily	$\sigma = \exp($meanlog$)$

Table 5.4: Implemented scale and location models. [The column "parameter" shows the connection between μ, respectively σ and the parametrization in R.]

the binomial model and also the rest of the implementation including the calls to the methods for the computation of the optimal ICs and for the least favorable radii is completely analogous to the binomial model; confer Section 3.5. However, the numerical computations seem to be faster and more stable in case of the exponential scale model. To make use of this, first compute the optimally robust IC (`IC1`) for the exponential scale model ($\sigma = 1$) and radius $r \in (0, \infty]$ and then convert `IC1` into the optimally robust IC for the Gumbel location model (`loc` $= \mu$) via the connections given in Remark 5.1.8. This can be done via

```
> cent(IC1) <- -cent(IC1)
> CallL2Fam(IC1) <- call("GumbelLocationFamily", loc = μ)
```

in case of contamination neighborhoods, respectively via

```
> c0 <- clipLo(IC1)
> clipLo(IC1) <- -clipUp(IC1)
> clipUp(IC1) <- -c0
> CallL2Fam(IC1) <- call("GumbelLocationFamily", loc = μ)
```

in case of total variation neighborhoods.

Remark 5.5.1 (a) In case of absolutely continuous distributions, the computation of the Kolmogorov(–Smirnov) minimum distance estimator via `ksEstimator` is based on the `R` function `ks.test` (cf. R Development Core Team (2005)) which performs a Kolmogorov–Smirnov test. We determine the Kolmogorov(–Smirnov) minimum distance estimator by minimizing the corresponding test statistic.

(b) At least in some one-dimensional location models one can construct the optimal robust estimator by solving the corresponding M equation; confer Theorem 6.2.4 of Rieder (1994). This M estimator can be computed using the method `locMEstimator` which for some sample `X` and some IC `IC1` is called via

```
> Mest <- locMEstimator(x = X, IC = IC1)
```

(c) After the installation of our `R` bundle `RobASt` one can find the `R` scripts `ExponentialScaleModel.R`, and `GumbelLocationModel.R`, which contain some examples for the exponential scale and the Gumbel location model, in the directory "`.../RHome/library/ROptEst/scripts/`" where `RHome` stands for the local home directory of `R`. In addition, we provide the scripts `LognormalAndNormalModel.R` and `NormalScaleModel.R` for some examples in case of the lognormal scale, the normal location and the normal scale model. ////

Chapter 6

Gamma Model

In the initial section (cf. Section 6.1) we introduce the Gamma model where the scale and the shape parameter have to be estimated. We then specify the optimally robust ICs (cf. Section 6.2) and describe the effect of different parameterizations. Some numerical results for the least favorable radii and the corresponding inefficiencies are given in Section 6.3. Subsequently, we verify that the optimally robust estimators can be constructed as one-step estimators; confer Section 6.4. Finally, we describe how one can use our `R` package `ROptEst` in case of the Gamma model; confer Section 6.5.

6.1 Introduction

The last non-standard robust model we investigate is the Gamma model

$$\mathcal{P} = \left\{ P_\theta = \text{Gamma}(\theta) \,\middle|\, \theta = (\sigma, \alpha)^\tau,\ \sigma, \alpha \in (0, \infty) \right\} \tag{6.1.1}$$

which is absolutely continuous with density

$$f_\theta(y) = \frac{1}{\sigma \Gamma(\alpha)} \, (y/\sigma)^{\alpha - 1} \exp(-y/\sigma) \, \mathrm{I}(y \geq 0) \tag{6.1.2}$$

Remark 6.1.1 The Gamma model (6.1.1) forms a two parameter exponential family with respect to the Lebesgue measure on $\mathbb{R}$, since we can rewrite

$$f_\theta(y) = y^{-1} \exp\left\{ -\sigma^{-1} y + \alpha \log y - [\alpha \log \sigma + \log \Gamma(\alpha)] \right\} \tag{6.1.3}$$

confer also Beispiel 1.156 (a) of Witting (1985) and Example 5.14 of Lehmann and Casella (1998). With the notation of Lemma 2.3.6 we obtain, $\zeta(\theta) = (-\sigma^{-1}, \alpha)$, $\beta(\theta) = \alpha \log \sigma + \log \Gamma(\alpha)$, $T(y) = (y, \log y)$ and $h(y) = y^{-1}$ which leads to

$$\mathcal{J}_\zeta = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathrm{E}_\theta \, T = \begin{pmatrix} \alpha\sigma \\ \mathrm{di}\Gamma(\alpha) + \log \sigma \end{pmatrix} \tag{6.1.4}$$

where $\mathrm{di}\Gamma(\alpha) = \frac{\partial}{\partial\alpha}\log\Gamma(\alpha)$, $\mathrm{E}_\theta\, y = \alpha\sigma$ and

$$\mathrm{E}_\theta \log(y/\sigma) = \frac{\frac{\partial}{\partial\alpha}\Gamma(\alpha)}{\Gamma(\alpha)} = \mathrm{di}\Gamma(\alpha) \tag{6.1.5}$$

Furthermore,

$$\mathrm{Cov}_\theta\, T = \begin{pmatrix} \mathrm{Var}_\theta\, y & \sigma d \\ \sigma d & \mathrm{Var}_\theta \log(y/\sigma) \end{pmatrix} = \begin{pmatrix} \alpha\sigma^2 & \sigma \\ \sigma & \mathrm{tri}\Gamma(\alpha) \end{pmatrix} \tag{6.1.6}$$

since $\mathrm{Var}_\theta\, y = \alpha\sigma^2$, $d = \mathrm{E}_\theta\big(y/\sigma - \alpha\big)\big(\log(y/\sigma) - \mathrm{di}\Gamma(\alpha)\big) = 1$ and

$$\mathrm{Var}_\theta \log(y/\sigma) = \mathrm{E}_\theta\big(\log(y/\sigma)\big)^2 - \big(\mathrm{E}_\theta \log(y/\sigma)\big)^2 = \frac{\frac{\partial^2}{\partial\alpha^2}\Gamma(\alpha)}{\Gamma(\alpha)} - \mathrm{di}\Gamma(\alpha)^2 \tag{6.1.7}$$

$$= \frac{\partial^2}{\partial\alpha^2}\log\Gamma(\alpha) = \mathrm{tri}\Gamma(\alpha) \tag{6.1.8}$$

As a consequence of Satz 1.153 in Witting (1985), the Gamma model is a strictly 2-parametric exponential family and therefore $\mathrm{Cov}_\theta\, T \succ 0$ for all $\theta \in \Theta$ such that $\zeta(\theta)$ lies in the interior of the natural parameter space Z_*; confer Satz 1.164 (b) (ibid.). For more details on the Gamma (Γ), digamma ($\mathrm{di}\Gamma$) and trigamma ($\mathrm{tri}\Gamma$) functions we refer to Artin (1964) and Chapter 6 of Abramowitz and Stegun (1984), respectively. ////

Lemma 6.1.2 *The Gamma model (6.1.1) is L_2 differentiable at $\theta = (\sigma, \alpha)^\tau \in (0,\infty)\times(0,\infty)$ with L_2 derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by*

$$\Lambda_\theta(y) = \begin{pmatrix} \sigma^{-1}(y/\sigma - \alpha) \\ \log(y/\sigma) - \mathrm{di}\Gamma(\alpha) \end{pmatrix} \qquad \mathcal{I}_\theta = \begin{pmatrix} \alpha/\sigma^2 & 1/\sigma \\ 1/\sigma & \mathrm{tri}\Gamma(\alpha) \end{pmatrix} \tag{6.1.9}$$

Proof A consequence of Lemma 2.3.6 (a) in connection with Remark 6.1.1. ////

Remark 6.1.3 Gamma distributions are for instance used to model personal-income or the length of stay in hospitals; confer Victoria-Feser and Ronchetti (1994) and Marazzi et al. (1998), respectively. ////

In Section 4.4 Hampel et al. (1986) consider the robust estimation of the parameter α where the parameter $\nu = \log(\sigma)$ is regarded as nuisance. As they show in Example 1 of Subsection 4.3d (ibid.), the re-parametrization of the Gamma model via $\vartheta = (\nu, \alpha)^\tau = (\log(\sigma), \alpha)^\tau$ leads to score function

$$\Lambda_\vartheta = \begin{pmatrix} (e^{-\nu}y - \alpha) \\ \log(e^{-\nu}y) - \mathrm{di}\Gamma(\alpha) \end{pmatrix} \tag{6.1.10}$$

and Fisher information

$$\mathcal{I}_\vartheta = \begin{pmatrix} \alpha & 1 \\ 1 & \mathrm{tri}\Gamma(\alpha) \end{pmatrix} \tag{6.1.11}$$

where $\mathcal{L}_\vartheta(\Lambda_\vartheta) = \mathcal{L}_{\vartheta_0}(\Lambda_{\vartheta_0})$ with $\vartheta_0 = (0, \alpha)^\tau$ and $\mathcal{I}_\vartheta$ is independent of ν. Hence, with this parametrization the Gamma model has a certain invariance structure.

Remark 6.1.4 **(a)** Using the parametrization ϑ, the Gamma model reads

$$\mathcal{Q} = \{Q_\vartheta = \text{Gamma}\,(e^\nu, \alpha)\,|\,\vartheta = (\nu,\alpha)^\tau \in \mathbb{R}^2\} \tag{6.1.12}$$

With the bijective and differentiable transformation $g(\nu,\alpha) = (e^\nu,\alpha)^\tau = (\sigma,\alpha)^\tau$ which has Jacobian matrix

$$D_\theta = dg(\nu,\alpha) = \begin{pmatrix} e^\nu & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma & 0 \\ 0 & 1 \end{pmatrix} \tag{6.1.13}$$

of full rank, we get $Q_\vartheta = P_{g(\vartheta)}$. Now, fix some parameter value ϑ and define $\theta := g(\vartheta)$; i.e., $Q_\vartheta = P_\theta$. Then,

$$\sqrt{dQ_{\vartheta+s}} = \sqrt{dP_{g(\vartheta+s)}} = \sqrt{dP_{g(\vartheta)+D_\theta s+\mathrm{o}(|s|)}} = \sqrt{dP_{\theta+D_\theta s+\mathrm{o}(|s|)}} \tag{6.1.14}$$

and the L_2 differentiability of $\mathcal{P}$ at θ yields

$$\sqrt{dP_{\theta+D_\theta s+\mathrm{o}(|s|)}} = \big[1+\tfrac{1}{2}\big(D_\theta s+\mathrm{o}(|s|)\big)^\tau \Lambda_\theta\big]\sqrt{dP_\theta} + \mathrm{o}(|Ds|) \tag{6.1.15}$$

$$= \big(1+\tfrac{1}{2}s^\tau D_\theta^\tau \Lambda_\theta\big)\sqrt{dP_\theta} + \mathrm{o}(|s|) \tag{6.1.16}$$

confer also Rieder (1994), p 119. Hence, $\mathcal{Q}$ is L_2 differentiable in ϑ with L_2 derivative $D_\theta^\tau \Lambda_\theta = \Lambda_\vartheta$ and Fisher information $D_\theta^\tau \mathcal{I}_\theta D_\theta = \mathcal{I}_\vartheta$. Such parameter transformations are in more detail considered in Subsection 5.5.4 of Rieder (1994).

(b) Since $\mathcal{L}_\vartheta(\Lambda_\vartheta) = \mathcal{L}_{\vartheta_0}(\Lambda_{\vartheta_0})$ and $\mathcal{I}_\vartheta = \mathcal{I}_{\vartheta_0}$, we are in the setup of Theorem 2.4.1. That is, the re-parametrization of Hampel et al. (1986) leads to a simplification in our setup, too. For more details we refer to Section 6.2. ////

Marazzi and Ruffieux (1996) who discuss the implementation of the M estimators for the Gamma model proposed by Hampel et al. (1986) also use the parametrization ϑ. In addition, they consider the parameter transformation $h(\nu,\alpha) = (\log(\alpha)+\nu,\alpha)^\tau = (\kappa,\alpha)^\tau$ as their main interest is the estimation of the mean of the Gamma distribution which is $\alpha\sigma = e^\kappa$. Such differentiable transformations with Jacobian matrix of full rank can easily be taken into consideration in case of the optimal solutions presented in Section 6.2. One just has to specify the corresponding Jacobian matrix for the transformation which in this example is

$$D_\kappa = dh(\nu,\alpha) = \begin{pmatrix} 1 & -1/\alpha \\ 0 & 1 \end{pmatrix} \tag{6.1.17}$$

More details are given in the subsequent section.

6.2 Optimally Robust Influence Curves

We consider the Gamma model (6.1.12) only with infinitesimal contamination neighborhoods (1.2.4) as the exact solution in case of total variation neighborhoods and $p > 1$ is not available. We fix $\vartheta = \vartheta_0 = (0,\alpha)^\tau$ and drop the fixed

parameter ϑ_0 as an index. The unique MSE optimal IC $\tilde{\eta}_r$ for radius $r \in (0,\infty)$ provided by Theorem 1.3.7 (a) and Theorem 1.3.11 (b) is

$$\tilde{\eta}_r(y) = A_r(\Lambda(y) - z_r) \min\left\{1, \frac{b_r}{|A_r(\Lambda(y) - z_r)|}\right\} \mathrm{I}(y > 0) \tag{6.2.1}$$

where

$$0 = \mathrm{E}\, A_r(\Lambda - z_r) \min\left\{1, \frac{b_r}{|A_r(\Lambda - z_r)|}\right\} \tag{6.2.2}$$

$$\mathbb{I}_2 = A_r\, \mathrm{E}(\Lambda - z_r)(\Lambda - z_r)^\tau \min\left\{1, \frac{b_r}{|A_r(\Lambda - z_r)|}\right\} \tag{6.2.3}$$

and

$$r^2\, b_r = \mathrm{E}\left(|A_r(\Lambda - z_r)| - b_r\right)_+ \tag{6.2.4}$$

For $r = \infty$ we obtain by Theorem 1.3.7 (b)

$$\omega_c^{\min} = \max\left\{\frac{\operatorname{tr} A_\infty}{\mathrm{E}\,|A_\infty(\Lambda - z_\infty)|} \,\middle|\, z_\infty \in \mathbb{R}^2,\ A_\infty \in \mathbb{R}^{2\times 2} \setminus \{0\}\right\} \tag{6.2.5}$$

and

$$\tilde{\eta}_\infty(y) = \omega_c^{\min} \frac{A_\infty(\Lambda(y) - z_\infty)}{|A_\infty(\Lambda(y) - z_\infty)|} \tag{6.2.6}$$

achieves the minimum bias $\omega_c^{\min}$.
For a plot of the optimally robust ICs in case $\nu = 0$, $\alpha = 2$ and for different values of r see Figure 6.1.

Remark 6.2.1 By $\mathcal{L}_\vartheta(\Lambda_\vartheta) = \mathcal{L}_{\vartheta_0}(\Lambda_{\vartheta_0})$ and $\mathcal{I}_\vartheta = \mathcal{I}_{\vartheta_0}$ we can apply Theorem 2.4.1 and obtain the following connections between the Lagrange multipliers at ϑ_0 and ϑ

$$A_{\vartheta_0,r} = A_{\vartheta,r} \qquad z_{\vartheta_0,r} = z_{\vartheta,r} \qquad b_{\vartheta_0,r} = b_{\vartheta,r} \tag{6.2.7}$$

////

Now, we want to estimate a differentiable transformation $\tau : \mathbb{R}^2 \to \mathbb{R}^p$ $(p \le 2)$ of ϑ whose Jacobian matrix $D = d\tau(\vartheta) \in \mathbb{R}^{p\times k}$ has full rank p. For instance, $\tau = g$, respectively $\tau = h$ as specified in Section 6.1. In this case we have to replace $\mathbb{I}_2$ in (6.2.3) by the Jacobian matrix D and $\operatorname{tr} A_\infty$ by $\operatorname{tr} A_\infty D^\tau$ in (6.2.5), respectively. Unlike to the one-dimensional models considered in Chapters 3–5, where we introduce $c_r = b_r/|A_r|$ and are then able to compute z_r and c_r independent of A_r. This does not work in general in dimension $k > 1$. As a consequence, the Lagrange multipliers for the estimation of $\tau(\vartheta)$ cannot easily be read off from the corresponding Lagrange multipliers for the estimation of ϑ in general. It holds, $D\tilde{\eta}_{\vartheta,r} \in \Psi_2^D(\vartheta)$ but as a rule $\operatorname{maxMSE}(D\tilde{\eta}_{\vartheta,r}, r) > \operatorname{maxMSE}(\tilde{\eta}_{\tau(\vartheta),r}, r)$; confer also Rieder (1994), p 198 (Full and Partial Solutions). The MSE–inefficiency of $D_\theta\tilde{\eta}_{\vartheta,r}$ with respect to $\tilde{\eta}_{\theta,r}$ for the estimation of $\theta = (2,2)^\tau$ is given in Table 6.1. As we see, it may become quite large.

r	maxMSE($D_\theta\tilde{\eta}_{\vartheta,r}, r$)	maxMSE($\tilde{\eta}_{\theta,r}, r$)	MSE–Ineff.
0.00	15.799	15.799	1.000
0.10	19.220	18.697	1.028
0.50	45.068	38.859	1.160
1.00	101.365	76.709	1.321
2.00	350.540	203.305	1.724
∞	∞	∞	2.333

Table 6.1: MSE–inefficiencies of $D_\theta\tilde{\eta}_{\vartheta,r}$ with respect to $\tilde{\eta}_{\theta,r}$ for $\theta = (2,2)^\tau$ in case of contamination neighborhoods $(* = c)$.

Remark 6.2.2 (a) The Lagrange multipliers for the estimation of $\tau(\vartheta)$ can be read off from the corresponding Lagrange multipliers for the estimation of ϑ if $D = \gamma G$ where $\gamma \in (0, \infty)$ and G is an orthogonal matrix. This corresponds to

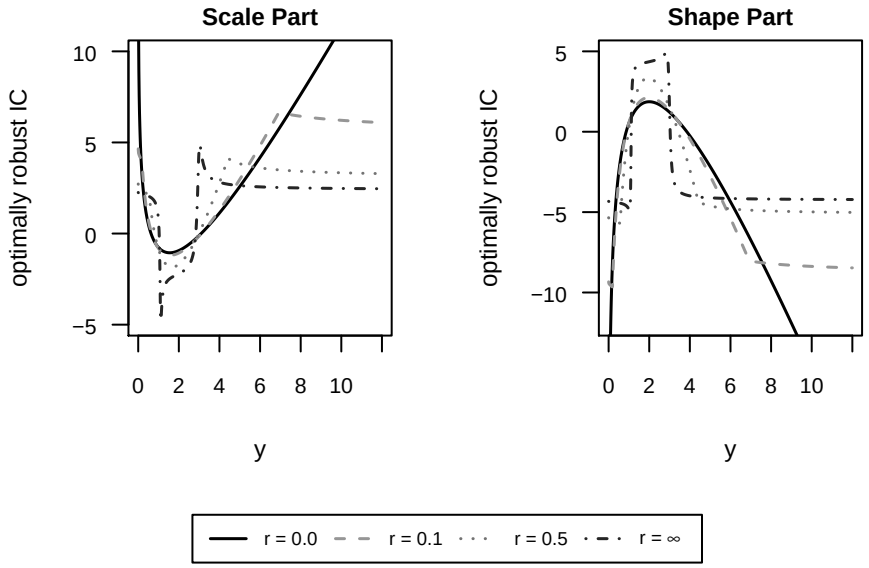

Figure 6.1: Optimally Robust ICs for Gamma$(1,2)$ in case of contamination neighborhoods $(* = c)$ with radius $r = 0, 0.1, 0.5, \infty$.

the setup of Theorem 2.4.1; confer also Subsection 5.5.4 of Rieder (1994) (p 214).

(b) In the setup of Marazzi and Ruffieux (1996) (cf. Section 4.4) the standardizing matrices for bijective and differentiable parameter transformations can always be obtained via the corresponding Jacobian matrices. However, the considered M estimators possess only a weaker optimality compared to the MSE optimal AL estimators. ////

6.3 Least Favorable Radius

For the Gamma model we only specify the least favorable radii r_0 and the corresponding MSE–inefficiencies since the computational effort for the determination of the least favorable radii r_2 and r_3 is too large; confer also Remark 6.5.1. As we see in Table 6.2, the efficiency loss of the radius–minimax estimator in all considered cases is around 50% where the corresponding radii r_0 are about 0.5. With the experience of Rieder et al. (2001) and this thesis, we expect a MSE–inefficiency of around 20% for the least favorable radii r_3 and about 10% for the least favorable radii r_2 in case of the examples considered in Table 6.2.

σ	α	relMSE $(\eta_\infty, 0)$	$\rho = 0$	r_0
1.0	1.0	2.490	1.486	0.520
	2.0	2.611	1.495	0.510
	3.0	2.677	1.501	0.504
2.0	2.0	2.585	1.494	0.512
3.0	2.0	2.618	1.497	0.509

Table 6.2: Least Favorable Radius r_0 and corresponding MSE–inefficiency in case of contamination neighborhoods ($* = c$).

6.4 One-Step Construction

As in the non-classical models before, Lemma 2.3.6 applies. That is, we can construct the optimally robust estimator as an one-step estimator with a suitable (i.e., strict and $\sqrt{n}$ consistent on $\mathcal{U}_c(\theta)$) initial estimator in the Gamma model. Our candidate is once again the Kolmogorov(–Smirnov) minimum distance estimator which has these properties (cf. Subsection 6.3.2 of Rieder (1994)) and which is still well computable in this slightly more complicated model.

6.5 Implementation using `R`

For more details about the `R` package `ROptEst` which provides the necessary `S4` classes and methods (cf. Chambers (1998)) we refer to Appendix D.3. In case of the Gamma model, we implemented the generating function `GammaFamily` which

is called with some parameters `scale` $= \sigma$ and `shape` $= \alpha$ ($\sigma, \alpha \in (0, \infty)$). The rest of the implementation including the calls to the methods for the computation of the optimal ICs and for the least favorable radii is completely analogous to the binomial model. This also holds for the robust estimation by means of the one-step construction. Therefore, we refer to Section 3.5 where the implementation of the binomial model is specified in detail.

Remark 6.5.1 (a) Like in the binomial model, it is possible to compute the radius–minimax IC and the least favorable radii. However, the computational effort is very large. The computation of the radius minimax IC for some given radius interval lasts between about 15 and 30 minutes depending on the given interval using `R 2.0.1` (cf. R Development Core Team (2005)) on an AMD Athlon with 2.5 GHz and 512 MB RAM. Consequentially, the determination of the least favorable radii r_2 and r_3 takes several hours (approx. 6).

(b) After installing our `R` bundle `RobASt` one can find the `R` script `GammaModel.R`, which contains some examples for the Gamma model, in the directory "`.../RHome/library/ROptEst/scripts/`" where `RHome` stands for the local home directory of `R`. ////

Part III

Robust Regression and Scale

The treatment of robust linear regression with unknown error scale in the theories by Huber (1981) and Hampel et al. (1986) is only tentative in several respects, and this is true already for the simpler model of joint location and scale. Thus, a more systematic investigation is called for.

Huber's (1981) Treatment

Neither Huber's minimax theory for the asymptotic variance of location M estimates (cf. Huber (1964) and Huber (1981), respectively), based on minimum Fisher information and symmetric error law even under contamination, nor Huber's finite-sample minimax interval estimation (cf. Huber (1968)), based on robust testing and least favorable pairs – both approaches employing neighborhoods of fixed size – have been extended from location to joint location and scale.

Already for scale alone, in the symmetrically contaminated model about a centered normal, the minimax variance approach remains incomplete since the saddle-point for the (relative) asymptotic variance of scale M estimates could be verified only for neighborhood size at most 4%; confer Section 5.7 of Huber (1981). Besides, the least favorable distributions minimizing Fisher information for contamination size $\geq 20.5\%$ have "pathological" densities with a singularity at zero; confer Section 5.6 of Huber (1981).

Assuming the scale of the errors known, however, the minimax asymptotic variance theory does have an immediate extension from location to linear regression, if regressor and error remain stochastically independent under contamination, if the error distribution may be perturbed as in the location case, while the regressor distribution stays fixed (ideal), and if the regression M estimates of Section 7.3 in Huber (1981), modifying residuals irrespectively of the regressor, are employed.

Huber's finite-sample minimax interval estimate of location (cf. Huber (1968)) has been extended to simple regression (one-dimensional, through the origin) by Rieder (1989), again assuming known error scale. This is in detail considered in Part V of this thesis.

For his definition of joint location and scale estimates, Huber (1981) (cf. Section 6.4, ibid.) starts with a pair of M equations employing generalized location scores in the two MLE equations. He generalizes this by dropping the connection of location and scale score functions in the two equations. This distinction will make the difference between M and AL estimates in this part of the thesis.

Apart from the determination of breakdown points in Section 6.6 (ibid.), Huber (1981) does not pursue any quantitative, let alone optimal, robustness of his joint location and scale estimates. His examples simply consist in combinations of estimates that are separately optimal: A minimax location estimate with a minimax (restricted, as mentioned) scale estimate in Example 4.1 (ibid.), extending Proposal 2 of Huber (1964), and, in Example 4.2 (ibid.), the median with the median absolute deviation from the median (MAD), where the first is distinguished by its minimax bias in the pure location case by Section 4.2 of Huber (1981). Even in the pure scale problem, the corresponding property of the MAD, besides its maximum breakdown point 50%, is left open. In Section 6.5 (ibid.) and elsewhere, Huber (1981) considers scale subordinate to location and invokes symmetry of the error distribution in order to make the influence curve of the location estimate not

depended on the scale estimate, except from its limiting value, assuming any $\sqrt{n}$ -consistent AL estimate of scale.

Hampel's et al. (1986) Treatment

On the one hand, the model is in fact covered by the local and asymptotic, infinitesimal robustness theories of Hampel et al. (1986) and Rieder (1994), respectively, as these related two approaches can deal with a general (finite-dimensional, smooth) parameter. But as yet, the results for location (regression) and scale have not been spelled out very explicitly in these setups.

Hampel et al. (1986) determine the ICs of Gateaux-differentiable Fisher consistent functionals of location and scale under the assumption of a symmetric error distribution which is a case of robust adaptivity as defined in Part IV of this thesis and give a few examples (cf. Section 4.2.d, pp 232–237, ibid.). The model of location and scale is invoked next to introduce models with partitioned parameter (cf. Section 4.4.a, p 253, ibid.). Again assuming symmetric errors, the estimation of scale is considered subordinate to location, which may be true from a practical viewpoint. With reference to intuition and the simulation study by Andrews et al. (1972), a most robust scale estimate (w.r.t. minimum bias, resp. maximum breakdown point?) is recommended to go with a robust location estimate.

Then, in Section 4.4.b, pp 253–256, Hampel et al. (1986) minimize each diagonal element (or block) of the asymptotic covariance subject to separate bounds in supnorm on the corresponding components of ICs (4.4.b Theorem 1, p 255, ibid.). No subordination of components is made, and no criterion is given for the choice of the sensitivity bounds. As for location and scale, the authors mention experience to choose the bounds on the location and scale components minimally, respectively near the minimum (4.3.d Remark 4, 4.4.b Remark 4, ibid.).

In Rieder (1994), robust location (regression) and scale is not mentioned explicitly at all. But it should be noted that both 4.4 Theorem 1 of Hampel et al. (1986) and 4.3 Theorem 1 (invoked for 4.4 Theorem 1) are generalized in several respects by Theorem 5.5.1, Remark 5.5.4 and the paragraph on one-at-a-time optimality (bottom of p 197) in Rieder (1994). Moreover, the sensitivity bound may be determined according to the mean square error criterion as a function of the (starting) radius $r \in (0, \infty)$ (of the $r/\sqrt{n}$-neighborhoods, which shrink with sample size n); confer Rieder (1994), Section 5.5.2, Theorem 5.5.7.

Therefore, some efforts remain to make infinitesimal robustness more explicit for this model.

Our Treatment

In Chapter 7 we first specify the ideal model; confer Subsection 7.1.1. In addition to the general class of AL estimators given in Subsubsection 7.1.3.1, we introduce several narrower classes of M estimators; confer Subsubsection 7.1.3.2. The general regression and scale M estimators are motivated by equations (4.3) and (4.4) of Subsection 6.4 in Huber (1981) where he considers location and scale. As a special case we present particular M estimates invented by Bednarski and Müller (2001) which we therefore call BM estimators. These estimators are restricted to regressor distributions of finite support as in the context of experimental design, which they consider.

Next, we consider the topic of equivariance for these classes of estimators; confer Subsection 7.1.4. Contrary to previous literature, equivariance is obtained not by a restriction on estimators but as a consequence of the optimization problems.

Throughout Chapter 7 we assume unconditional ($t = 0$), respectively average conditional ($t = \alpha = 1$) contamination neighborhoods ($* = c$).

In Section 7.2 we derive the optimally robust ICs for the simultaneous estimation of regression and scale. First, we consider the general AL estimators; confer Subsection 7.2.1. The optimal solutions are obtained by specializing the convex optimization results of Chapters 5 and 7 of Rieder (1994). But new aspects arise. For instance, the regression coordinate of the optimal IC, due to additional scale, is redescending. In a next step, we further specialize and simplify the solutions by assuming elliptical, respectively spherical symmetry of the ideal regressor distribution.

Second, in Subsection 7.2.2, we consider general M estimators. For this subclass the optimization problem of the simultaneous estimation of regression and scale is not convex. Thus, we solve a restricted problem and need an additional outer optimization to obtain the MSE solution; confer Subsubsections 7.2.2.1 and 7.2.2.2. The outer problem can only be solved via numerical optimization. Again, we give specializations to elliptically and spherically symmetric ideal regressor distributions.

In Section 7.3 we then derive the optimally robust ICs for the separate estimation of regression and scale. That is, today we want to estimate regression where scale is regarded as nuisance whereas tomorrow we are interested in the estimation of scale and the regression parameter is only nuisance. We include this problem since BM estimators may be motivated as estimators for the separate estimation of regression and scale.

First, we specify the AL solutions for the separate estimation; confer Subsection 7.3.1. Due to robust adaptivity, which is verified in Subsection 9.2.1, the solution coincides with the solutions to the separate problems.

Second, we derive the M solutions for the separate estimation of regression and scale; confer Subsection 7.3.2. In analogy to Bednarski and Müller (2001) we consider only M estimators with conditionally centered ICs and assume ideal error distribution $F = \mathcal{N}(0,1)$. As in case of simultaneous estimation, the optimization problem is not convex and we solve a restricted problem. That is, an additional outer optimization is needed to obtain the MSE solution.

Third, we treat BM estimators; confer Subsection 7.3.3. For the derivation of optimal BM solutions we decided to give our own proofs since Bednarski and Müller (2001) need an interior point assumption to derive the optimal solutions and give only incomplete Lagrange arguments. In case of BM estimators the problems of simultaneous and separate estimation of regression and scale coincide (contrary to AL and M estimators).

From optimal ICs optimal estimators have to be constructed. We consider this problem for the optimal AL estimators in case of linear regression with unconditional contamination neighborhoods ($* = c,\ t = 0$); confer Section 7.4. By virtue of Theorem 2.3.3 it is possible, depending on a $\sqrt{n}$-consistent initial estimator, to construct the optimally robust estimator by means of the one-step method. However, our result is limited to ideal error distribution $F = \mathcal{N}(0,1)$ and bounded

regressors; confer Subsections 7.4.1 and 7.4.2. Moreover, we consider only the optimal AL estimators for simultaneous estimation of regression and scale.

Numerical evaluations are already needed for the outer optimization loop for determining M solutions. We calculate both AL and M solutions numerically in some simple examples and determine the efficiency loss of M relative to AL; confer Section 7.5. For the sake of this introduction we select a few efficiency comparisons.

First, we consider simultaneous estimation of regression and scale; confer Subsection 7.5.1. In case of unconditional contamination neighborhoods ($* = c,\ t = 0$), the subefficiency of the general M estimators is small ($< 10\%$) in all situations considered. However, in case of average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$) the efficiency loss of the general M estimators relative to the optimal AL estimators, which we call ALc estimators, may become quite large. The maximum subefficiency in the examples considered is about 300%. In case of the BM estimators this efficiency loss further increases and reaches up to about 425% with respect to the optimal ALc estimators.

Second, we give some results for the separate estimation of regression and scale; confer Subsection 7.5.2. Again, the optimal AL estimators, which we call ALs estimators in this context, perform much better than Ms and BM estimators. The maximum efficiency loss of these estimators are about 315% and 360%, respectively. In all examples considered the subefficiency of BM relative to Ms is moderate and stays below 15%.

These numerical comparisons can be done with our `R` packages `ROptRegTS` and `RobRex` which are part of our `R` bundle `RobASt`; confer Appendix D. A description of these packages can be found in Subsections 7.6.1 and 7.6.2, respectively. Package `ROptRegTS` is an extension of our package `ROptEst` to regression-type models as defined in Appendix A. It makes use of object orientation by means of `S4` classes and methods; confer Chambers (1998). In addition, package `RobRex` includes `R` functions which are suited for the computation of the optimal ICs for all estimators considered throughout this chapter.

A more lucid comparison between these estimators is possible in case of location and scale since no regressor distribution has to be chosen. In addition, we can take into account several well-known robust estimators which have been proposed in literature for location and scale.

In Chapter 8 we compare 18 different estimators for the estimation of normal location and scale.

First, we present the normal location and scale model with infinitesimal contamination neighborhoods of (starting) radius $r \in (0, \infty)$; confer Section 8.1. The corresponding optimal MSE solutions in case of AL, M and BM estimators are then given in Sections 8.2–8.4. Subsequently, we specify other well known robust location and scale estimators which are based on proposals of Huber (cf. Subsection 8.5.1), Hampel (cf. Subsection 8.5.2), Andrews (cf. Subsection 8.5.3), Tukey (cf. Subsection 8.5.4) and Yohai (cf. Subsection 8.6), respectively. All these estimators are asymptotically linear.

In Section 8.1 we further introduce the notions of absolute and relative information in terms of the norm of the IC. These notions are used to compare the various

estimators with respect to the amount of absolute and relative information which they associate with a given observation. In contrast to the classically optimal IC, which is unbounded in case of normal location and scale (i.e., its absolute information is unbounded), we expect the IC of some robust estimator to be bounded (i.e., its absolute information is bounded). Indeed, this is true for all estimators considered in this chapter. However, in case of the relative information the situation looks different. As it turns out, the relative information of the IC of the optimal AL estimators is very similar to the relative information of the classically optimal IC. Moreover, the deviations in case of the general M estimators are only slightly larger. But, in case of the remaining estimators the deviations are clearly larger and the relative information in some cases even looks very different.

We supplement these qualitative comparisons with further numerical evaluations. In Section 8.7 we compare the (numerical) minimax asymptotic MSE of the considered estimators and determine the efficiency loss relative to the optimal AL estimators. For the sake of this introduction we only mention a few efficiency comparisons: Among the remaining estimators the general M estimators perform best and lose only a few permille efficiency with respect to the optimal AL estimators. The Proposal 2 of Huber (1964) and the estimator proposed in Example 6.4.1 of Huber (1981) perform well for small starting radii ($r \leq 0.5$) and have subefficiencies up to about 21% and 12%, respectively in the limiting case $r \to \infty$. Moreover, the combination of Huber's ψ-function (cf. Huber (1964)), respectively Hampel's three part redescending ψ-function (cf. Subsections 2C3, 3C3 of Andrews et al. (1972)) with the MAD as proposed in Andrews et al. (1972) performs quite well for large (starting) radii r. Their efficiency losses in case $r \geq 1$ vary between about 10% and 12%.

Given some optimal IC we need to construct the corresponding estimator. At least the optimal AL estimators can be obtained by means of the one-step method. This is verified in Section 8.8 by an application of Lemma 2.3.6.

For the numerical determination of the MSE-optimal ICs we provide our `R` package `RobLox`. This package contains `R` functions for the computation of the MSE-optimal ICs for all estimators considered throughout this chapter; confer Section 8.9. Moreover, in case of the AL estimators the corresponding optimal ICs can also be determined via our `R` package `ROptEst`. Both `R` packages are part of our `R` bundle `RobASt`; confer Appendix D.

Chapter 7

Regression and Scale

In Section 7.1, we first specify the linear regression model with unknown scale of the error distribution (cf. Subsection 7.1.1). Then, we introduce different classes of M estimators and take a closer look at the equivariance of these estimators as well as of the more general AL estimators; confer Subsection 7.1.3. The optimal solutions for these different robust estimators are in detail derived in Section 7.2 (simultaneous estimation) and Section 7.3 (separate estimation). In Section 7.4 we investigate the one-step method for the optimal AL estimators in case of unconditional contamination neighborhoods ($* = c, t = 0$). Subsequently, we give some numerical results for the optimal AL and M estimators (simultaneous and separate estimation) introduced in this chapter; confer Section 7.5. Finally, we briefly present our `R` packages `ROptRegTS` (cf. Subsection 7.6.1) and `RobRex` (cf. Subsection 7.6.2) which are part of our `R` bundle `RobASt` (cf. Appendix D).

7.1 Introduction

7.1.1 Ideal Model

We consider the linear regression model with regression parameter β and unknown scale parameter σ of the form

$$y_i = x_i^\tau \beta + \sigma u_i \qquad (i = 1, \ldots, n) \tag{7.1.1}$$

where $x_1, \ldots, x_n$ are i.i.d. realizations of the regressor x distributed according to some probability K on $\mathbb{B}^k$, and $u_1, \ldots, u_n$ are i.i.d. copies of the error $u \sim F$. It is assumed that $x \sim K$ and $u \sim F$ are stochastically independent. Furthermore we make the following assumptions on F and K in the ideal case,

(F1) F is symmetric.

(F2) The Fisher information of location for F is finite; i.e., F has an absolutely continuous density f and $\mathcal{I}_F^{\mathrm{loc}} = \int (\Lambda_f^{\mathrm{loc}})^2 f \, d\lambda < \infty$ where $\Lambda_f^{\mathrm{loc}} := -f'/f$; confer Proposition 5.1.4.

(F3) The Fisher information of scale for F is finite; i.e., $u \mapsto uf(u) =: uf$ is absolute continuous and $\mathcal{I}_F^{\mathrm{sc}} = \int (\Lambda_f^{\mathrm{sc}})^2 f \, d\lambda < \infty$ where $\Lambda_f^{\mathrm{sc}} := u\Lambda_f^{\mathrm{loc}} - 1$; confer Proposition 5.1.1.

(K) $\mathcal{K} := \int xx^\tau K(dx) \in \mathbb{R}^{k\times k}$ has $\mathrm{rk}(\mathcal{K}) = k$.

Remark 7.1.1 **(a)** Assumption (F1) leads to a clear simplification of the solutions as some of the constraints will be fulfilled automatically; confer Section 7.2.

(b) Conditions (F2), (F3) and (K) guarantee the L_2 differentiability of the Regression and Scale model (7.1.1); confer Proposition 7.1.2.

(c) A possible generalization of the regression, respectively regression and scale model is to consider converging regressor distributions K_n, $n \in \mathbb{N}$. As a first step, the LAN property extends under the assumptions that K_n converge weakly together with their second moments. But in that context, one has to adapt the definitions of IC and AL estimators, and has to derive the corresponding infinitesimal bias terms. ////

Thus, the ideal model distributions with parameter $\theta = (\beta^\tau, \sigma)^\tau \in \mathbb{R}^k \times (0,\infty)$ read

$$P_\theta(dx, dy) = \frac{1}{\sigma} f\Big(\frac{y - x^\tau\beta}{\sigma}\Big)\lambda(dy)K(dx) \tag{7.1.2}$$

This model stays invariant under the following transformations,

$$g_\theta(x, y) = (x, x^\tau\beta + \sigma y) \tag{7.1.3}$$

in the sense that, for all $\theta \in \mathbb{R}^k \times (0,\infty)$, P_θ is the image measure

$$P_\theta = g_\theta(P_{\theta_0}) \tag{7.1.4}$$

of P_{θ_0} under the transformation g_θ where $\theta_0 := (0,1)^\tau$. We define the scores at $\theta \in \mathbb{R}^k \times (0,\infty)$,

$$\Lambda_\theta(x, y) := \frac{1}{\sigma}\begin{pmatrix} x\Lambda_f^{\mathrm{loc}}(u) \\ \Lambda_f^{\mathrm{sc}}(u) \end{pmatrix} = \frac{1}{\sigma}\Lambda_{\theta_0} \circ g_\theta^{-1}(x, y) \tag{7.1.5}$$

where $u = \sigma^{-1}(y - x^\tau\beta)$. Thus we have for the corresponding Fisher information

$$\mathcal{I}_\theta = \mathrm{E}_\theta \, \Lambda_\theta\Lambda_\theta^\tau = \sigma^{-2}\,\mathrm{E}_{\theta_0}\,\Lambda_{\theta_0}\Lambda_{\theta_0}^\tau = \sigma^{-2}\mathcal{I}_{\theta_0} \tag{7.1.6}$$

for all $\theta \in \mathbb{R}^k \times (0,\infty)$ and by the symmetry of the ideal error distribution F we get

$$\mathcal{I}_{\theta_0} = \begin{pmatrix} \mathcal{I}_F^{\mathrm{loc}}\mathcal{K} & 0 \\ 0 & \mathcal{I}_F^{\mathrm{sc}} \end{pmatrix} \tag{7.1.7}$$

Here E_θ and E_{θ_0} are expectations taken under P_θ and P_{θ_0}, respectively.

Proposition 7.1.2 *Assume (F2), (F3) and (K). Then, the linear regression and scale model (7.1.2) is L_2-differentiable at every $\theta \in \mathbb{R}^k \times (0,\infty)$ with L_2-derivative Λ_θ given by (7.1.5) and Fisher information $\mathcal{I}_\theta$ given by (7.1.6).*

PROOF The scale model is L_2-differentiable by Swensen (1980), Section 3 of Chapter 2 and even multivariate by Ruckdeschel (2001), Theorem B.3.4. Moreover, the linear regression model is L_2-differentiable by Rieder (1994), Theorem 2.4.6. Putting these arguments together we get the L_2-differentiability of the linear regression and scale model (7.1.2). ////

Remark 7.1.3 There are quite a lot of references for robust regression; for example Chapter 7 of Huber (1981) or Chapter 6 of Hampel et al. (1986) where the computational aspects are in detail treated in Marazzi (1993). Our results are closely related to the results given in Bickel (1984) and especially to the results derived in Chapter 7 of Rieder (1994). ////

7.1.2 Infinitesimal Neighborhoods

We consider the neighborhoods which in Rieder (1994) are abbreviated by the subscripts $*=c,\ t=0$ and $*=c,\ t=\alpha=1$, respectively. That is, unconditional (errors-in-variables) and average conditional (error-free-variables) contamination neighborhoods. The definition of the first ones is given in Section 1.2 (c.f. (1.2.4)). The second ones, which are also called error-free-variables neighborhoods, for given (starting) radius $r \in [0,\infty]$ and with any contamination curve $\varepsilon\colon (\mathbb{R}^k, \mathbb{B}^k) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$ such that $\mathrm{E}\,\varepsilon \le 1$, consist of all probabilities $Q(dx, dy) = Q(dy \,|\, x)Q(dx)$ on $\mathbb{B}^{k+1}$ such that

$$Q(dx) = K(dx) \tag{7.1.8}$$

and

$$Q(dy \,|\, x) = (1 - r\varepsilon(x))P_\theta(dy \,|\, x) + r\varepsilon(x)M(dy \,|\, x) \tag{7.1.9}$$

for some Markov kernel $M(dy \,|\, x)$ from $\mathbb{R}^k$ to $\mathbb{B}$; confer also Bickel (1984), p 1350.

Remark 7.1.4 This study may be extended to conditional contamination neighborhoods of type $t=\varepsilon, \alpha=2,\infty$ as well as to total variation neighborhoods of type $t=\varepsilon, \alpha=1$; confer Subsection 7.2.2 of Rieder (1994). ////

7.1.3 Estimators

Following Section 6.4 of Huber (1981) the simultaneous M estimate of location and scale is any pair of estimators (T_n, S_n) determined by two equations of the form

$$\sum_{i=1}^{n} \psi\Big(\frac{y_i - T_n}{S_n}\Big) = 0 \tag{7.1.10}$$

and

$$\sum_{i=1}^{n} \Big[\frac{y_i - T_n}{S_n}\psi\Big(\frac{y_i - T_n}{S_n}\Big) - 1\Big] = 0 \tag{7.1.11}$$

where ψ is the generalization of the location scores Λ_f^{loc}. However without argumentation Huber gives up this special form and generalizes (7.1.11) to

$$\sum_{i=1}^{n} \chi\Big(\frac{y_i - T_n}{S_n}\Big) = 0 \tag{7.1.12}$$

where χ is an arbitrary function independent of ψ. By p 136 of Huber (1981) the corresponding influence curves $\mathrm{IC}_{\mathrm{loc}}$ and $\mathrm{IC}_{\mathrm{sc}}$ satisfy the system of equations

$$\mathrm{IC}_{\mathrm{loc}}\,\mathrm{E}_\theta\,\psi' + \mathrm{IC}_{\mathrm{sc}}\,\mathrm{E}_\theta\,u\psi' = \sigma\psi \tag{7.1.13}$$

$$\mathrm{IC}_{\mathrm{loc}}\,\mathrm{E}_\theta\,\chi' + \mathrm{IC}_{\mathrm{sc}}\,\mathrm{E}_\theta\,u\chi' = \sigma\chi \tag{7.1.14}$$

where E_θ is expectation taken under $P_\theta(dy) = \sigma^{-1} f\big(\sigma^{-1}(y-\beta)\big)\lambda(dy)$. Huber (1981) assumes F is symmetric, ψ is odd, and χ is even to obtain

$$\mathrm{IC}_{\mathrm{loc}} = \sigma\frac{\psi}{\mathrm{E}_\theta\,\psi'} \qquad \text{and} \qquad \mathrm{IC}_{\mathrm{sc}} = \sigma\frac{\chi}{\mathrm{E}_\theta\,u\chi'} \tag{7.1.15}$$

Without symmetry assumptions using only

$$\mathrm{E}_\theta\,\psi' \neq 0 \qquad \text{and} \qquad \mathrm{E}_\theta\,u\chi'\,\mathrm{E}_\theta\,\psi' \neq \mathrm{E}_\theta\,u\psi'\,\mathrm{E}_\theta\,\chi' \tag{7.1.16}$$

one gets

$$\mathrm{IC}_{\mathrm{loc}} = \sigma\Big[\frac{\psi}{\mathrm{E}_\theta\,\psi'} - \frac{\mathrm{E}_\theta\,u\psi'(\chi\,\mathrm{E}_\theta\,\psi' - \psi\,\mathrm{E}_\theta\,\chi')}{(\mathrm{E}_\theta\,\psi')^2\,\mathrm{E}_\theta\,u\chi' - \mathrm{E}_\theta\,\psi'\,\mathrm{E}_\theta\,\chi'\,\mathrm{E}_\theta\,u\psi'}\Big] \tag{7.1.17}$$

and

$$\mathrm{IC}_{\mathrm{sc}} = \sigma\frac{\chi\,\mathrm{E}_\theta\,\psi' - \psi\,\mathrm{E}_\theta\,\chi'}{\mathrm{E}_\theta\,u\chi'\,\mathrm{E}_\theta\,\psi' - \mathrm{E}_\theta\,u\psi'\,\mathrm{E}_\theta\,\chi'} \tag{7.1.18}$$

By generalizing the M equation (7.1.11) to (7.1.12), Huber (1981) in fact introduced and considered the more general AL estimators as the following two propositions show. We first verify $\eta = (\mathrm{IC}_{\mathrm{loc}}, \mathrm{IC}_{\mathrm{sc}})^\tau \in \Psi_2(\theta)$; i.e., η is an IC in sense of Definition 1.1.1.

Proposition 7.1.5 *Assume*

$$\lim_{u\to\pm\infty} \psi(u)f(u) = 0 = \lim_{u\to\pm\infty} \chi(u)f(u) \tag{7.1.19}$$

and

$$\lim_{u\to\pm\infty} u\psi(u)f(u) = 0 = \lim_{u\to\pm\infty} u\chi(u)f(u) \tag{7.1.20}$$

Then, $\eta = (\mathrm{IC}_{\mathrm{loc}}, \mathrm{IC}_{\mathrm{sc}})^\tau \in \Psi_2(\theta)$.

Proof Due to $\mathrm{E}_\theta\,\psi = 0$ and $\mathrm{E}_\theta\,\chi = 0$ we immediately get $\mathrm{E}_\theta\,\mathrm{IC}_{\mathrm{loc}} = 0$ and $\mathrm{E}_\theta\,\mathrm{IC}_{\mathrm{sc}} = 0$. Hence, Fisher consistency remains to be shown. By assumption (7.1.19), using integration by parts, we get

$$\mathrm{E}_\theta\,\psi\Lambda_f^{\mathrm{loc}} = -\int \psi f'\,d\lambda = -\psi f\Big|_{-\infty}^{+\infty} + \int \psi' f\,d\lambda = \mathrm{E}_\theta\,\psi' \tag{7.1.21}$$

where λ is the Lebesgue measure on $(\mathbb{R}, \mathbb{B})$. Analogously, we obtain

$$\mathrm{E}_\theta\, \chi \Lambda_f^{\mathrm{loc}} = \mathrm{E}_\theta\, \chi' \tag{7.1.22}$$

Moreover, we get by assumption (7.1.20), $\mathrm{E}_\theta\, \psi = 0$, and integration by parts

$$\mathrm{E}_\theta\, u\psi' = \int \psi'(uf)\, d\lambda = u\psi f\Big|_{-\infty}^{+\infty} - \int \psi(f + uf')\, d\lambda = \mathrm{E}_\theta\, u\psi\Lambda_f^{\mathrm{loc}} \tag{7.1.23}$$

and analogously

$$\mathrm{E}_\theta\, u\chi' = \mathrm{E}_\theta\, u\chi\Lambda_f^{\mathrm{loc}} \tag{7.1.24}$$

Towards Fisher consistency we first show

$$\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{loc}} \frac{1}{\sigma} \Lambda_f^{\mathrm{loc}} = 1 \tag{7.1.25}$$

which by inserting $\mathrm{IC}_{\mathrm{loc}}$ is equivalent to

$$\mathrm{E}_\theta \Big[\frac{\psi}{\mathrm{E}_\theta\, \psi'} - \frac{\mathrm{E}_\theta\, u\psi'(\chi\, \mathrm{E}_\theta\, \psi' - \psi\, \mathrm{E}_\theta\, \chi')}{(\mathrm{E}_\theta\, \psi')^2\, \mathrm{E}_\theta\, u\chi' - \mathrm{E}_\theta\, \psi'\, \mathrm{E}_\theta\, \chi'\, \mathrm{E}_\theta\, u\psi'} \Big] \Lambda_f^{\mathrm{loc}} = 1 \tag{7.1.26}$$

Using (7.1.21) and (7.1.22) this can be rewritten as

$$\frac{\mathrm{E}_\theta\, u\psi'(\mathrm{E}_\theta\, \chi\Lambda_f^{\mathrm{loc}}\, \mathrm{E}_\theta\, \psi' - \mathrm{E}_\theta\, \psi\Lambda_f^{\mathrm{loc}}\, \mathrm{E}_\theta\, \chi')}{(\mathrm{E}_\theta\, \psi')^2\, \mathrm{E}_\theta\, u\chi' - \mathrm{E}_\theta\, \psi'\, \mathrm{E}_\theta\, \chi'\, \mathrm{E}_\theta\, u\psi'} = 0 \tag{7.1.27}$$

Thus, (7.1.23) and (7.1.24) entail equation (7.1.25). Second, we verify

$$\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{sc}} \frac{1}{\sigma} (u\Lambda_f^{\mathrm{loc}} - 1) = 1 \tag{7.1.28}$$

which by $\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{sc}} = 0$ is equivalent to

$$\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{sc}} \frac{1}{\sigma} u\Lambda_f^{\mathrm{loc}} = 1 \tag{7.1.29}$$

Inserting $\mathrm{IC}_{\mathrm{sc}}$ yields

$$\mathrm{E}_\theta\, \frac{\chi\, \mathrm{E}_\theta\, \psi' - \psi\, \mathrm{E}_\theta\, \chi'}{\mathrm{E}_\theta\, u\chi'\, \mathrm{E}_\theta\, \psi' - \mathrm{E}_\theta\, u\psi'\, \mathrm{E}_\theta\, \chi'} u\Lambda_f^{\mathrm{loc}} = 1 \tag{7.1.30}$$

which follows by (7.1.23) and (7.1.24). Next, we show

$$\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{loc}} \frac{1}{\sigma} (u\Lambda_f^{\mathrm{loc}} - 1) = 0 \tag{7.1.31}$$

Due to $\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{loc}} = 0$ this is equivalent to

$$\mathrm{E}_\theta\, \mathrm{IC}_{\mathrm{loc}} \frac{1}{\sigma} u\Lambda_f^{\mathrm{loc}} = 0 \tag{7.1.32}$$

That is, we obtain

$$\mathrm{E}_\theta\left[\frac{\psi}{\mathrm{E}_\theta\,\psi'}-\frac{\mathrm{E}_\theta\,u\psi'(\chi\,\mathrm{E}_\theta\,\psi'-\psi\,\mathrm{E}_\theta\,\chi')}{(\mathrm{E}_\theta\,\psi')^2\,\mathrm{E}_\theta\,u\chi'-\mathrm{E}_\theta\,\psi'\,\mathrm{E}_\theta\,\chi'\,\mathrm{E}_\theta\,u\psi'}\right]u\Lambda_f^{\mathrm{loc}}=0 \tag{7.1.33}$$

$$\iff \mathrm{E}_\theta\,\psi u\Lambda_f^{\mathrm{loc}}=\frac{\mathrm{E}_\theta\,\psi'\,\mathrm{E}_\theta\,u\psi'(\mathrm{E}_\theta\,\chi u\Lambda_f^{\mathrm{loc}}\,\mathrm{E}_\theta\,\psi'-\mathrm{E}_\theta\,\psi u\Lambda_f^{\mathrm{loc}}\,\mathrm{E}_\theta\,\chi')}{(\mathrm{E}_\theta\,\psi')^2\,\mathrm{E}_\theta\,u\chi'-\mathrm{E}_\theta\,\psi'\,\mathrm{E}_\theta\,\chi'\,\mathrm{E}_\theta\,u\psi'} \tag{7.1.34}$$

$$\iff 1=\frac{\mathrm{E}_\theta\,\psi'(\mathrm{E}_\theta\,\chi u\Lambda_f^{\mathrm{loc}}\,\mathrm{E}_\theta\,\psi'-\mathrm{E}_\theta\,\psi u\Lambda_f^{\mathrm{loc}}\,\mathrm{E}_\theta\,\chi')}{(\mathrm{E}_\theta\,\psi')^2\,\mathrm{E}_\theta\,u\chi'-\mathrm{E}_\theta\,\psi'\,\mathrm{E}_\theta\,\chi'\,\mathrm{E}_\theta\,u\psi'} \tag{7.1.35}$$

which is implied by (7.1.23) and (7.1.24). Finally, we verify

$$\mathrm{E}_\theta\,\mathrm{IC}_{\mathrm{sc}}\Lambda_f^{\mathrm{loc}}=0 \tag{7.1.36}$$

Inserting $\mathrm{IC}_{\mathrm{sc}}$ we get

$$\mathrm{E}_\theta\,\frac{\chi\,\mathrm{E}_\theta\,\psi'-\psi\,\mathrm{E}_\theta\,\chi'}{\mathrm{E}_\theta\,u\chi'\,\mathrm{E}_\theta\,\psi'-\mathrm{E}_\theta\,u\psi'\,\mathrm{E}_\theta\,\chi'}\Lambda_f^{\mathrm{loc}}=0 \tag{7.1.37}$$

which is a consequence of (7.1.21) and (7.1.22). ////

Now, we change our point of view.

Proposition 7.1.6 *Let $\eta=(\eta_{\mathrm{loc}},\eta_{\mathrm{sc}})^\tau\in\Psi_2(\theta)$ and identify η_{loc} and η_{sc} with ψ and χ, respectively. Moreover, assume (7.1.16), (7.1.19) and (7.1.20) for η_{loc} and η_{sc}, respectively. Then,*

$$\mathrm{IC}_{\mathrm{loc}}=\eta_{\mathrm{loc}}\qquad\textit{and}\qquad\mathrm{IC}_{\mathrm{sc}}=\eta_{\mathrm{sc}} \tag{7.1.38}$$

Proof We get by (7.1.20) and (7.1.23)

$$\mathrm{E}_\theta\,u\eta'_{\mathrm{loc}}=\mathrm{E}_\theta\,\eta_{\mathrm{loc}}u\Lambda_f^{\mathrm{loc}} \tag{7.1.39}$$

where

$$\mathrm{E}_\theta\,\eta_{\mathrm{loc}}u\Lambda_f^{\mathrm{loc}}=\mathrm{E}_\theta\,\eta_{\mathrm{loc}}(u\Lambda_f^{\mathrm{loc}}-1)=0 \tag{7.1.40}$$

Therefore, $\mathrm{IC}_{\mathrm{loc}}$ simplifies to

$$\mathrm{IC}_{\mathrm{loc}}=\sigma\frac{\eta_{\mathrm{loc}}}{\mathrm{E}_\theta\,\eta'_{\mathrm{loc}}} \tag{7.1.41}$$

In addition, we obtain by (7.1.19) and (7.1.21)

$$\mathrm{E}_\theta\,\eta'_{\mathrm{loc}}=\mathrm{E}_\theta\,\eta_{\mathrm{loc}}\Lambda_f^{\mathrm{loc}} \tag{7.1.42}$$

where $\mathrm{E}_\theta\,\eta_{\mathrm{loc}}\Lambda_f^{\mathrm{loc}}=\sigma$. That is, $\mathrm{IC}_{\mathrm{loc}}=\eta_{\mathrm{loc}}$. Moreover, it holds

$$\mathrm{E}_\theta\,\eta'_{\mathrm{sc}}=\mathrm{E}_\theta\,\eta_{\mathrm{sc}}\Lambda_f^{\mathrm{loc}} \tag{7.1.43}$$

by (7.1.19) and (7.1.21) where $\mathrm{E}_\theta\, \eta_{\mathrm{sc}} \Lambda_f^{\mathrm{loc}} = 0$. Thus, $\mathrm{IC}_{\mathrm{sc}}$ simplifies to

$$\mathrm{IC}_{\mathrm{sc}} = \sigma \frac{\eta_{\mathrm{sc}}}{\mathrm{E}_\theta\, u\eta'_{\mathrm{sc}}} \tag{7.1.44}$$

with

$$\mathrm{E}_\theta\, u\eta'_{\mathrm{sc}} = \mathrm{E}_\theta\, \eta_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}} - 1) \tag{7.1.45}$$

as a consequence of (7.1.20) and (7.1.23). Since $\mathrm{E}_\theta\, \eta_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}} - 1) = \sigma$, we obtain $\mathrm{IC}_{\mathrm{sc}} = \eta_{\mathrm{sc}}$. ////

Remark 7.1.7 Consequently, under assumptions (7.1.16), (7.1.19) and (7.1.20) the M estimates for simultaneous estimation of location and scale defined by equations (7.1.10) and (7.1.12) correspond to the general AL estimators. In case of regression and scale these are specified in Subsubsection 7.1.3.1. In addition, we consider M estimators for regression and scale which we define in analogy to equations (7.1.10) and (7.1.11); confer Subsubsection 7.1.3.1. ////

7.1.3.1 AL Estimators

General AL Estimators

We consider general AL estimators as introduced in Section 1.1.

AL Estimators with Conditionally Centered ICs

In addition, we treat AL estimators with conditionally centered ICs. More precisely, we consider the set $\Psi_{2\bullet}(\theta)$ of all (square integrable) conditionally centered ICs at P_θ; i.e.,

$$\Psi_{2\bullet}(\theta) = \left\{\psi_\theta \in L_2^{k+1}(P_\theta) \mid \mathrm{E}_\bullet\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_{k+1}\right\} \tag{7.1.46}$$

where for fixed x, $\mathrm{E}_\bullet$ denotes expectation with respect to $F(du)$.

7.1.3.2 M Estimators

Generalizing equations (7.1.10) and (7.1.11) to regression and scale, M estimators have ICs of the form

$$\rho_\theta(x,y) = \sigma \begin{pmatrix} A_\psi x \psi(x,u) \\ \gamma_\psi(u\psi(x,u) - 1) \end{pmatrix} = \sigma \rho_{\theta_0}(x,u) \tag{7.1.47}$$

with

$$A_\psi^{-1} = \mathrm{E}\, xx^\tau \Lambda_f^{\mathrm{loc}}(u)\psi(x,u) \tag{7.1.48}$$

$$\gamma_\psi^{-1} = \mathrm{E}(u\psi(x,u) - 1)\Lambda_f^{\mathrm{sc}}(u) \tag{7.1.49}$$

where $u = \sigma^{-1}(y - x^\tau\beta)$, $A_\psi \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma_\psi \in \mathbb{R} \setminus \{0\}$. Moreover, A_ψ and γ_ψ are the same for all θ. For the multiplication of ψ by σ compare also equations (4.11) and (4.12) in Section 6.4 of Huber (1981) where $\sigma = S(F)$.

General M Estimators

To make sure, that $\rho_\theta \in L_2^{k+1}(P_\theta)$ with $\mathrm{E}_\theta\, \rho_\theta = 0$ and $\mathrm{E}_\theta\, \rho_\theta \Lambda_{\theta_0} = \mathbb{I}_{k+1}$, that is, ρ_θ is an IC in sense of Definition 1.1.1 (a), we get the following general conditions

(M1) $\mathrm{E}\, x\psi(x,u) = 0$

(M2) $\mathrm{E}\, u\psi(x,u) = 1$

(M3) $\mathrm{E}\, xx^\tau \Lambda_f^{\mathrm{loc}}(u)\psi(x,u) = A_\psi^{-1}$

(M4) $\mathrm{E}\, u^2 \Lambda_f^{\mathrm{loc}}(u)\psi(x,u) = 1 + \gamma_\psi^{-1}$

(M5) $\mathrm{E}\, xu\Lambda_f^{\mathrm{loc}}(u)\psi(x,u) = 0$

and define

$$\Psi_{\mathrm{M}}(\theta) := \left\{\rho \in L_2^{k+1}(P_{\theta_0}) \,\middle|\, \psi \text{ fulfills (M1)–(M5)}\right\} \tag{7.1.50}$$

Remark 7.1.8 We also consider M estimators with $A = \mathcal{K}^{-1}$ fixed which we therefore call MK estimators. In case of location and scale, these MK estimators coincide with the general M estimators. For more details we refer to Remark 7.2.12 (a). ////

M Estimators with Conditionally Centered ICs

To make sure, that $\rho_\theta \in L_2^{k+1}(P_\theta)$ is a conditionally centered IC (i.e., $\mathrm{E}_\bullet\, \rho_\theta = 0$), we get the conditions (M3)–(M5) for ψ together with

(Mc1) $\mathrm{E}_\bullet\, x\psi(x,u) = 0$ for all x

(Mc2) $\mathrm{E}_\bullet\, u\psi(x,u) = 1$ for all x

where for fixed x, $\mathrm{E}_\bullet$ denotes expectation with respect to $F(du)$ and we define

$$\Psi_{\mathrm{Mc}}(\theta) := \left\{\rho \in L_2^{k+1}(P_{\theta_0}) \,\middle|\, \psi \text{ fulfills (Mc1),(Mc2),(M3)–(M5)}\right\} \tag{7.1.51}$$

We call the corresponding estimator Mc estimator.

BM Estimators

Bednarski and Müller (2001) introduce a special class of sectionwise M estimators, which we call BM estimators. However, they only consider ideal error distribution $F = \mathcal{N}(0,1)$ and ideal regressor distribution K with finite support and they instead of (M1)–(M5) impose the following stronger conditions on ψ

(Mc15) $\psi(x,.)$ is odd in u for all x

(Mc2) $\mathrm{E}_\bullet\, u\psi(x,u) = 1$ for all x

(Mc4) $\mathrm{E}_\bullet\, u^3\psi(x,u) = 1 + \gamma_{\psi,x}^{-1}$ for all x

We define

$$\Psi_{\mathrm{BM}}(\theta) := \left\{\rho \in L_2^{k+1}(P_{\theta_0}) \,\middle|\, \psi \text{ fulfills (Mc15),(Mc2),(Mc4)}\right\} \tag{7.1.52}$$

Remark 7.1.9 **(a)** The dependence of γ_ψ on x in case of the BM estimators (i.e., $\gamma_\psi = \gamma_{\psi,x}$) seems to have been overlooked in Bednarski and Müller (2001). For more details we refer to Remark 7.3.8 (a).

(b) If the ideal error distribution F is symmetric and if (Mc15) holds, then the conditions (M1), (Mc1) and (M5) are automatically fulfilled.

(c) In case of regression and scale and ideal error distribution $F = \mathcal{N}(0,1)$ we get by (Mc2) that $A_\psi^{-1} = \mathcal{K}$ for all $\psi \in L_2(P_{\theta_0})$; i.e., (Mc2)=(M3).

(d) In case of location and scale and ideal error distribution $F = \mathcal{N}(0,1)$ we get by (M2)=(Mc2) that $A_\psi^{-1} = 1$ for all $\psi \in L_2(P_{\theta_0})$; i.e., (M2)=(Mc2)=(M3). Moreover, (M4)=(Mc4). ////

Sectionwise M Estimators

We generalize the BM estimators by using (M4) instead of (Mc4); i.e., we are no longer restricted to regressor distributions K with finite support. Correspondingly, we introduce the class of sectionwise M estimators

$$\Psi_{\mathrm{Ms}}(\theta) := \left\{\rho \in L_2^{k+1}(P_{\theta_0}) \,\middle|\, \psi \text{ fulfills (Mc15),(Mc2),(M4)}\right\} \tag{7.1.53}$$

which we call Ms estimators.

7.1.4 Equivariance

While in classical manner equivariance is formulated in terms of estimators as in Definition 7.1.12 below, we begin with equivariance of ICs. Let $\Psi(\theta)$ form an arbitrary class of ICs for the regression and scale model, such that the bijection

$$\chi_\theta(x,y) = \sigma\chi_{\theta_0} \circ g_\theta^{-1}(x,y) = \sigma\chi_{\theta_0}(x,u) \tag{7.1.54}$$

holds for set of ICs at $\theta \in \mathbb{R}^k \times (0,\infty)$ and the set of ICs at $\theta_0 = (0,1)^\tau$, where $u = \sigma^{-1}(y - x^\tau\beta)$ and g_θ is defined in (7.1.3).

Remark 7.1.10 **(a)** The setup is analogous to Subsection 7.2.3 of Rieder (1994) where translation equivariance of the linear regression model is considered.

(b) The class of ICs $\Psi(\theta)$ includes the classes $\Psi_2(\theta)$, $\Psi_{2\bullet}(\theta)$, $\Psi_{\mathrm{M}}(\theta)$, $\Psi_{\mathrm{Mc}}(\theta)$, $\Psi_{\mathrm{BM}}(\theta)$ and $\Psi_{\mathrm{Ms}}(\theta)$, respectively. For these sets of ICs the equivariance (7.1.54) is a consequence of (7.1.5) and (7.1.6). ////

By (7.1.54) we get

$$\mathrm{E}_\theta\,|\chi_\theta|^2 = \sigma^2\,\mathrm{E}_{\theta_0}\,|\chi_{\theta_0}|^2 \tag{7.1.55}$$

respectively

$$\mathrm{maxMSE}_{\,\theta}(\chi_\theta, r) = \sigma^2 \mathrm{maxMSE}_{\,\theta_0}(\chi_{\theta_0}, r) \tag{7.1.56}$$

for all $\theta \in \mathbb{R}^k \times (0,\infty)$. As a consequence, the correspondence (7.1.54) also holds between the solutions at θ and at θ_0 of the corresponding Hampel type problems (1.3.7), (A.2.1), respectively the corresponding MSE problems (1.3.5), (A.2.5).

Remark 7.1.11 **(a)** Since the solutions to these optimization problems are unique, we need to solve them only for the special parameter $\theta_0 = (0,1)^\tau$. Then, starting from the unique solution $\tilde{\chi}_{\theta_0}$ at θ_0 and defining $\tilde{\chi}_\theta$ via (7.1.54) we obtain the unique solution $\tilde{\chi}_\theta$ at the general parameter $\theta \in \mathbb{R}^k \times (0,\infty)$. Thus, we can restrict our considerations to the special parameter $\theta_0 = (0,1)^\tau$ in Section 7.2 without loss of generality.

(b) This connection between the optimal solutions at θ_0 and θ entails a certain connection for the corresponding Lagrange multipliers contained in the optimal solutions; confer Section 7.2. This is a consequence of (7.1.5) and (7.1.6); confer also Theorem 2.4.1 ////

Assume an asymptotically linear estimator $S_n = S_n((x_1^\tau, y_1)^\tau, \ldots, (x_n^\tau, y_n)^\tau)$ with IC χ_θ at θ. Then, the following asymptotic expansion holds

$$\sqrt{n}(S_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi_\theta(x_i, y_i) + o_{P_\theta^n}(n^0) \tag{7.1.57}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma \chi_{\theta_0}\left(x_i, \frac{y_i - x_i^\tau \beta}{\sigma}\right) + o_{P_\theta^n}(n^0) \tag{7.1.58}$$

which implies

$$S_n(\ldots, (x_i^\tau, y_i)^\tau, \ldots) = \theta + \sigma\big[S_n(\ldots, (x_i^\tau, u_i)^\tau, \ldots) - \theta_0\big] + o_{P_\theta^n}(n^{-1/2}) \tag{7.1.59}$$

for all x_i, u_i, where $u_i = \sigma^{-1}(y_i - x_i^\tau \beta)$. Thus, we may call S_n asymptotically equivariant. With (7.1.3), respectively

$$G_\theta(\ldots, (x_i^\tau, u_i)^\tau, \ldots) := (\ldots, g_\theta(x_i, u_i)^\tau, \ldots)^\tau \qquad \forall x_i, u_i \tag{7.1.60}$$

we define (exact) equivariance for any fixed sample size $n \in \mathbb{N}$ as follows:

Definition 7.1.12 *An asymptotically linear estimator S_n is called* equivariant *in the regression and scale model, if*

$$S_n \circ G_\theta(\ldots, (x_i^\tau, u_i)^\tau, \ldots) = \theta + \sigma\big[S_n(\ldots, (x_i^\tau, u_i)^\tau, \ldots) - \theta_0\big] \tag{7.1.61}$$

for all x_i, u_i, $\theta = (\beta^\tau, \sigma)^\tau \in \mathbb{R}^k \times (0,\infty)$ and $\theta_0 = (0,1)^\tau$.

Remark 7.1.13 Definition 7.1.12 is equivalent to

$$S_n(\ldots, (x_i^\tau, \sigma u_i)^\tau, \ldots) = \sigma S_n(\ldots, (x_i^\tau, u_i)^\tau, \ldots) \tag{7.1.62}$$

$$S_n(\ldots, (x_i^\tau, x_i^\tau \beta + u_i)^\tau, \ldots) = (\beta^\tau, 0)^\tau + S_n(\ldots, (x_i^\tau, u_i)^\tau, \ldots) \tag{7.1.63}$$

for all x_i, u_i, $\sigma \in (0,\infty)$, $\beta \in \mathbb{R}^k$. ////

Now, consider M equations and let us assume existence and uniqueness of the solution S_n^{M}. That is, we have to determine S_n^{M} such that

$$0 = \sum_{i=1}^{n} \chi_{S_n^{\mathrm{M}}}(x_i, y_i) = \sum_{i=1}^{n} S_{n,\mathrm{sc}}^{\mathrm{M}} \chi_{\theta_0}\left(x_i, \frac{y_i - x_i^\tau S_{n,\mathrm{rg}}^{\mathrm{M}}}{S_{n,\mathrm{sc}}^{\mathrm{M}}}\right) \tag{7.1.64}$$

or equivalently,

$$0 = \sum_{i=1}^{n} \chi_{\theta_0}\left(x_i, \frac{y_i - x_i^\tau S_{n,\mathrm{rg}}^{\mathrm{M}}}{S_{n,\mathrm{sc}}^{\mathrm{M}}}\right) \tag{7.1.65}$$

where $S_{n,\mathrm{rg}}^{\mathrm{M}}$ and $S_{n,\mathrm{sc}}^{\mathrm{M}}$ are the regression, respectively scale part of S_n^{M}. We substitute y_i by $x_i^\tau\beta + \sigma y_i$ and define

$$T_n^{\mathrm{M}} := S_n^{\mathrm{M}}(\ldots, (x_i^\tau, x_i^\tau\beta + \sigma y_i)^\tau, \ldots) = S_n^{\mathrm{M}} \circ G_\theta(\ldots, (x_i^\tau, y_i)^\tau, \ldots) \tag{7.1.66}$$

for all x_i, u_i. Using (7.1.54), we get

$$0 = \sum_{i=1}^{n} \chi_{T_n^{\mathrm{M}}}(x_i, x_i^\tau\beta + \sigma y_i) = \sum_{i=1}^{n} \chi_{\theta_0}\left(x_i, \frac{x_i^\tau\beta + \sigma y_i - x_i^\tau T_{n,\mathrm{rg}}^{\mathrm{M}}}{T_{n,\mathrm{sc}}^{\mathrm{M}}}\right) \tag{7.1.67}$$

$$= \sum_{i=1}^{n} \chi_{\theta_0}\left(x_i, \frac{\sigma y_i - x_i^\tau(T_{n,\mathrm{rg}}^{\mathrm{M}} - \beta)}{T_{n,\mathrm{sc}}^{\mathrm{M}}}\right) \tag{7.1.68}$$

which we have to compare with (7.1.65). By existence and uniqueness of the solution, we obtain

$$T_{n,\mathrm{sc}}^{\mathrm{M}} = \sigma S_{n,\mathrm{sc}}^{\mathrm{M}} \qquad T_{n,\mathrm{rg}}^{\mathrm{M}} = \sigma S_{n,\mathrm{rg}}^{\mathrm{M}} + \beta \tag{7.1.69}$$

Thus, if the construction of some asymptotically optimal estimator with IC χ_θ by means of M equations leads to a unique solution, the resulting M estimate S_n^{M} is equivariant in sense of Definition 7.1.12 not by restriction but as a consequence of the considered optimization problems. And this holds for the AL estimators specified in Subsubsection 7.1.3.1 as well as for the various classes of M estimators introduced in Subsubsection 7.1.3.2.

7.2 Simultaneous Estimation

We consider the simultaneous estimation of regression and scale. To lighten the notation, we drop the fixed parameter θ.

7.2.1 AL Estimators

7.2.1.1 Unconditional Contamination Neighborhoods

In case of unconditional contamination neighborhoods (1.2.4) ($* = c$, $t = 0$), the solution to the Hampel type problem (1.3.7), respectively to the MSE problem (1.3.5) can be read off from Theorem 1.3.7 and Theorem 1.3.11 (b), respectively. Under the additional assumption (F1), which is the symmetry of the ideal error distribution F, the solutions further simplify.

Proposition 7.2.1 *Assume (F1).*

(a) *In case* $\omega_c^{\min} < b \le \omega_{c,0}(\eta_h)$, *there exist some* $A_{\rm rg} \in \mathbb{R}^{k\times k}$ *and* $A_{\rm sc}, z_{\rm sc} \in \mathbb{R}$ *such that the unique solution to Problem (1.3.7) is of the form*

$$\tilde\eta(x,u) = \begin{pmatrix}\tilde\eta_{\rm rg}(x,u)\\ \tilde\eta_{\rm sc}(x,u)\end{pmatrix} = \begin{pmatrix} A_{\rm rg} x \Lambda_f^{\rm loc}(u) \\ A_{\rm sc}(u\Lambda_f^{\rm loc}(u) - z_{\rm sc})\end{pmatrix} w(x,u) \tag{7.2.1}$$

with

$$w(x,u) = \min\left\{1, \frac{b}{\left[|A_{\rm rg}x|^2 \Lambda_f^{\rm loc}(u)^2 + A_{\rm sc}^2 (u\Lambda_f^{\rm loc}(u) - z_{\rm sc})^2\right]^{1/2}}\right\} \tag{7.2.2}$$

Conversely, if some $\tilde\eta \in \Psi_2$ *is of form (7.2.1) for any* $b \in (0,\infty)$, $A_{\rm rg} \in \mathbb{R}^{k\times k}$ *and* $A_{\rm sc}, z_{\rm sc} \in \mathbb{R}$, *then* $\tilde\eta$ *is the solution, and the following representations hold,*

$$A_{\rm rg} = \left[\mathrm{E}\, x x^\tau (\Lambda_f^{\rm loc})^2 w\right]^{-1} \quad (\succ 0) \tag{7.2.3}$$

$$A_{\rm sc} = \left[\mathrm{E}(u\Lambda_f^{\rm loc} - z_{\rm sc})^2 w\right]^{-1} \quad (>0) \qquad z_{\rm sc} = \mathrm{E}\, u\Lambda_f^{\rm loc} w \left[\mathrm{E}\, w\right]^{-1} \tag{7.2.4}$$

(b) *The solution to the MSE problem (1.3.5) coincides with the solution (7.2.1), where bound* $b \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *are related via*

$$r^2 b = \mathrm{E}\left(\left[|A_{\rm rg}x|^2 (\Lambda_f^{\rm loc})^2 + A_{\rm sc}^2 (u\Lambda_f^{\rm loc} - z_{\rm sc})^2\right]^{1/2} - b\right)_+ \tag{7.2.5}$$

(c) *Assume*

$$P\left(\begin{pmatrix} A_1 x \Lambda_f^{\rm loc} \\ A_2(u\Lambda_f^{\rm loc} - z)\end{pmatrix} = 0\right) = 0 \tag{7.2.6}$$

for all $A_1 \in \mathbb{R}^{k\times k}$, $A_2, z \in \mathbb{R}$ *where* A_1, A_2 *not both zero. Then, it holds that*

$$\omega_{c,0}^{\min} = \max\left\{\frac{\operatorname{tr} A_{\rm rg} + A_{\rm sc}}{\mathrm{E}\, Y} \,\middle|\, A_{\rm rg} \in \mathbb{R}^{k\times k}\backslash\{0\},\ A_{\rm sc} \in (0,\infty), z_{\rm sc} \in \mathbb{R}\right\} \tag{7.2.7}$$

with

$$Y(x,u) = \left[|A_{\rm rg}x|^2 \Lambda_f^{\rm loc}(u)^2 + A_{\rm sc}^2 (u\Lambda_f^{\rm loc}(u) - z_{\rm sc})^2\right]^{1/2} \tag{7.2.8}$$

There exist $\bar A_{\rm rg} \in \mathbb{R}^{k\times k}\backslash\{0\}$ $(\bar A_{\rm rg} \succ 0)$, $\bar A_{\rm sc} \in (0,\infty)$, $\bar z_{\rm sc} \in \mathbb{R}$ *and* $\bar\eta \in \Psi_2$ *achieving* $\omega_{c,0}^{\min}$, *respectively. And then necessarily*

$$\bar\eta(x,u) = \begin{pmatrix}\bar\eta_{\rm rg}(x,u)\\ \bar\eta_{\rm sc}(x,u)\end{pmatrix} = \omega_{c,0}^{\min} \begin{pmatrix} \bar A_{\rm rg} x \Lambda_f^{\rm loc}(u) \\ \bar A_{\rm sc}(u\Lambda_f^{\rm loc}(u) - \bar z_{\rm sc})\end{pmatrix} \bar Y(x,u)^{-1} \tag{7.2.9}$$

where

$$\bar Y(x,u) := \left[|\bar A_{\rm rg}x|^2 \Lambda_f^{\rm loc}(u)^2 + \bar A_{\rm sc}^2 (u\Lambda_f^{\rm loc}(u) - \bar z_{\rm sc})^2\right]^{1/2} \tag{7.2.10}$$

Proof Assume (F1).

(a) We split $\eta \in L_2^{k+1}(P)$ into the two parts $\eta_{\rm rg} \in L_2^k(P)$ and $\eta_{\rm sc} \in L_2(P)$ and modify the Hampel type problem (1.3.7) using only the side conditions

$$\mathrm{E}\,\eta_{\rm rg} x^\tau \Lambda_f^{\rm loc} = \mathbb{I}_k \qquad \mathrm{E}\,\eta_{\rm sc} = 0 \qquad \mathrm{E}\,\eta_{\rm sc} u \Lambda_f^{\rm loc} = 1 \tag{7.2.11}$$

besides

$$|\eta_{\rm rg}|^2 + \eta_{\rm sc}^2 \le b^2 \qquad \text{a.e. } P \tag{7.2.12}$$

By analogous Lagrange arguments as in the proof of Theorem 5.5.1 (a) of Rieder (1994), the solution $\tilde{\eta}$ to this weaker problem is of form (7.2.1) and the representations (7.2.3) and (7.2.4) hold.
Since

$$\int x\Lambda_f^{\mathrm{loc}}(u)w(x,u)\,F(du) = 0 \qquad \text{for all } x \in \mathbb{R}^k \tag{7.2.13}$$

by symmetry of F, we get $\mathrm{E}\,\tilde{\eta} = 0$, if z_{sc} is determined such that (7.2.4) holds. Moreover, we have

$$\int xu\Lambda_f^{\mathrm{loc}}(u)^2 w(x,u)\,F(du) = 0 \qquad \text{for all } x \in \mathbb{R}^k \tag{7.2.14}$$

by symmetry of F and therefore we get $\mathrm{E}\,\tilde{\eta}\Lambda^\tau = \mathbb{I}_{k+1}$, if A_{rg} and A_{sc} are calculated via (7.2.3) and (7.2.4). Thus, this special $\tilde{\eta}$, which is obviously of form (1.3.16), fulfills the constraints of the original problem (1.3.7) and therefore by Theorem 1.3.7 (a) is the solution to that problem, too. The uniqueness of the solution $\tilde{\eta}$ is a consequence of the convexity of Ψ_2 and $\omega_{c,0}$ (cf. Lemma 1.3.4) and the strict convexity of the norm.

(b) With the arguments provided in the proof of Theorem 5.5.7 in Rieder (1994) the solution is of form (7.2.1) and the MSE equation (1.3.45) can be rewritten as (7.2.5).

(c) As in part (a) of this proof, we split $\eta \in L_2^{k+1}(P)$ into the two parts $\eta_{\mathrm{rg}} \in L_2^k(P)$ and $\eta_{\mathrm{sc}} \in L_2(P)$ and use only the constraints (7.2.11). By arguments given on page 198 of Rieder (1994), the minimum bias is always attained. Hence, if $\bar{\eta}$ achieves the minimum bias, then it solves the convex, well-posed problem

$$\omega_{c,0}(\eta) = \min! \qquad \eta \in L_2^{k+1}(P) \tag{7.2.15}$$

under the (weaker) side conditions (7.2.11). Using the same arguments as in proof of Theorem 5.5.1 (b) in Rieder (1994), the form

$$\bar{\eta}(x,u) = \omega_{c,0}^{\min}\begin{pmatrix} \bar{A}_{\mathrm{rg}}x\Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}) \end{pmatrix} \bar{Y}(x,u)^{-1}\,\mathrm{I}\left\{\begin{pmatrix} \bar{A}_{\mathrm{rg}}x\Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}) \end{pmatrix} \neq 0\right\} \tag{7.2.16}$$

is obtained and $\bar{A}_{\mathrm{rg}}$, $\bar{A}_{\mathrm{sc}}, \bar{z}_{\mathrm{sc}}$ achieve the maximum (7.2.7) where $\bar{A}_{\mathrm{rg}} \succeq 0$, $\bar{A}_{\mathrm{sc}} \geq 0$ and $\bar{A}_{\mathrm{rg}}, \bar{A}_{\mathrm{sc}}$ not both zero. Thus, using the additional assumption (7.2.6), we obtain the form (7.2.9) of the solution. Hence, we get $\bar{A}_{\mathrm{rg}} \succ 0$ and $\bar{A}_{\mathrm{sc}} > 0$ by the constraints $\mathrm{E}\,\eta_{\mathrm{rg}}x^\tau\Lambda_f^{\mathrm{loc}} = \mathbb{I}_k$ and $\mathrm{E}\,\eta_{\mathrm{sc}}(x,u)u\Lambda_f^{\mathrm{loc}}(u) = 1$, respectively. Moreover, as in part (a) of this proof, the symmetry of the ideal error distribution F yields

$$\mathrm{E}\,\bar{\eta}_{\mathrm{rg}} = 0 \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{rg}}\Lambda_f^{\mathrm{sc}} = 0 \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{sc}}x^\tau\Lambda_f^{\mathrm{loc}} = 0 \tag{7.2.17}$$

i.e., $\bar{\eta} \in \Psi_2$ and therefore is the solution to problem (7.2.15) among all $\eta \in \Psi_2$. ////

Remark 7.2.2 (a) The regression part $\tilde{\eta}_{\mathrm{rg}}$ of the solution $\tilde{\eta}$ is a redescending function in u given x.

(b) Similarly to Theorem 2.4.1, we obtain the following connection between the Lagrange multipliers contained in the solution at θ_0 and θ

$$A_{\mathrm{rg},\theta} = \sigma^2 A_{\mathrm{rg},\theta_0} \qquad A_{\mathrm{sc},\theta} = \sigma^2 A_{\mathrm{sc},\theta_0} \qquad z_{\mathrm{sc},\theta} = \sigma^{-1} z_{\mathrm{sc},\theta_0} \qquad b_\theta = \sigma b_{\theta_0} \quad (7.2.18)$$

Hence, $\tilde{\eta}_\theta$ is the solution to the corresponding Hampel type problem (1.3.7) with bound $b_\theta = \sigma b_{\theta_0}$. ////

Elliptically Symmetric Regressor Distribution

We now give a second specialization of the solution (1.3.16) by additionally assuming, that the ideal regressor distribution K is elliptically symmetric. More precisely, we assume $X = C\tilde{X}$ where C is a regular $k \times k$-matrix and $\mathcal{L}(\tilde{X})$ is spherically symmetric, which means $\mathcal{L}(G\tilde{X}) = \mathcal{L}(\tilde{X})$ for all orthogonal $k \times k$-matrices G; confer Serfling (2004). By the singular value decomposition we get $X = G^\tau D G_1 \tilde{X}$ with G, G_1 orthogonal $k \times k$-matrices and D a diagonal $k \times k$-matrix. Hence, the spherical symmetry of $\mathcal{L}(\tilde{X})$ yields

$$\mathcal{L}(X) = \mathcal{L}(G^\tau D\tilde{X}) = \mathcal{L}(G^\tau V) \qquad V := D\tilde{X} \quad (7.2.19)$$

Proposition 7.2.3 *Assume (F1) and (7.2.19).*

(a) *In case $\omega_c^{\min} < b \le \omega_{c,0}(\eta_h)$, the unique solution to Problem (1.3.7) is of form (7.2.1), where $A_{\mathrm{rg}} = G^\tau A_{\mathrm{rg}}^G G$ with $A_{\mathrm{rg}}^G = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$, $z_{\mathrm{sc}} = z_{\mathrm{sc}}^G$, $A_{\mathrm{sc}} = A_{\mathrm{sc}}^G$ and the following representations hold*

$$\alpha_j = \left[\mathrm{E}\, v_j^2 \left(\Lambda_f^{\mathrm{loc}}\right)^2 w_G\right]^{-1} \qquad j = 1, \ldots k \quad (7.2.20)$$

$$z_{\mathrm{sc}}^G = \mathrm{E}\,\Lambda_f^{\mathrm{sc}} w_G \left[\mathrm{E}\, w_G\right]^{-1} \qquad A_{\mathrm{sc}}^G = \left[\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 w_G\right]^{-1} \quad (7.2.21)$$

Here

$$w_G(v,u) = \min\left\{1, \frac{b}{\left[|A_{\mathrm{rg}}^G v|^2 (\Lambda_f^{\mathrm{loc}}(u))^2 + (A_{\mathrm{sc}}^G)^2 (\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}}^G)^2\right]^{1/2}}\right\} \quad (7.2.22)$$

and E *is expectation taken under $\tilde{P}(du, dv) := f(u)\lambda(du)\,\mathcal{L}(V)(dv)$.*

(b) *The solution to the corresponding MSE problem (1.3.5) coincides with the solution of part (a), where bound $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via*

$$r^2 b = \mathrm{E}\left(\left[|A_{\mathrm{rg}}^G v|^2 (\Lambda_f^{\mathrm{loc}})^2 + (A_{\mathrm{sc}}^G)^2 (\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}}^G)^2\right]^{1/2} - b\right)_+ \quad (7.2.23)$$

(c) *Assume*

$$P\left(\begin{pmatrix} A_1 v \Lambda_f^{\mathrm{loc}} \\ A_2(u\Lambda_f^{\mathrm{loc}} - z) \end{pmatrix} = 0\right) = 0 \quad (7.2.24)$$

for all $A_1 \in \mathbb{R}^{k \times k}$ of diagonal form, $A_2, z \in \mathbb{R}$ where A_1, A_2 not both zero. Then, it holds

$$\omega_{c,0}^{\min} = \max\left\{\frac{\sum_{j=1}^k \alpha_j + A_{\mathrm{sc}}^G}{\mathrm{E}\, Y_G}\right\} \quad (7.2.25)$$

with

$$Y_G(v,u) := \left[|A_{\mathrm{rg}}^G v|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + (A_{\mathrm{sc}}^G)^2 (\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}}^G)^2\right]^{1/2} \quad (7.2.26)$$

where $A_{\rm rg}^G = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$ *and* $\alpha_1, \ldots, \alpha_k, A_{\rm sc}^G \in (0, \infty)$, $z_{\rm sc}^G \in \mathbb{R}$.
There exist $\bar{A}_{\rm rg}^G = \mathrm{diag}(\bar{\alpha}_1, \ldots, \bar{\alpha}_k)$ *with* $\bar{\alpha}_1, \ldots, \bar{\alpha}_k \in (0, \infty)$ *and* $\bar{A}_{\rm sc}^G \in (0, \infty)$, $\bar{z}_{\rm sc}^G \in \mathbb{R}$ *and* $\bar{\eta} \in \Psi_2$ *achieving* $\omega_{c,0}^{\min}$, *respectively. And then necessarily* $\bar{\eta}$ *is of form (7.2.9), where* $\bar{A}_{\rm rg} = G^\tau \bar{A}_{\rm rg}^G G$, $\bar{A}_{\rm sc} = \bar{A}_{\rm sc}^G$ *and* $\bar{z}_{\rm sc} = \bar{z}_{\rm sc}^G$.

Proof Assume (F1) and (7.2.19).

(a) By (F1) we may restrict our considerations to the Hampel-type problem (1.3.7) with side conditions (7.2.11); confer proof of Proposition 7.2.1 (a). But instead of this problem we now consider the problem

$$\mathrm{E}\,|\eta_{\rm rg}^G|^2 = \min! \qquad \eta_{\rm rg}^G \in L_2^k(\tilde{P}),\ \eta_{\rm sc}^G \in L_2(\tilde{P}) \tag{7.2.27}$$

with side conditions

$$\mathrm{E}\,\eta_{{\rm rg},j}^G v_j \Lambda_f^{\rm loc} = 1 \qquad \mathrm{E}\,\eta_{\rm sc}^G = 0 \qquad \mathrm{E}\,\eta_{\rm sc}^G u \Lambda_f^{\rm loc} = 1 \tag{7.2.28}$$

$$|\eta_{\rm rg}^G|^2 + (\eta_{\rm sc}^G)^2 \le b^2 \qquad \text{a.e. } \tilde{P} \tag{7.2.29}$$

where $\eta_{{\rm rg},j}^G$ and v_j are the $j-th$ components of the vectors $\eta_{\rm rg}^G$ and v. That is, we switch to the regression model with unknown scale and ideal regressor distribution $\mathcal{L}(V)$. With arguments as in the proof of Theorem 5.5.1 (a) in Rieder (1994) the solution reads

$$\tilde{\eta}^G(v,u) = \begin{pmatrix} \tilde{\eta}_{\rm rg}^G(v,u) \\ \tilde{\eta}_{\rm sc}^G(v,u) \end{pmatrix} = \begin{pmatrix} A_{\rm rg}^G v \Lambda_f^{\rm loc}(u) \\ A_{\rm sc}^G (u\Lambda_f^{\rm loc}(u) - z_{\rm sc}^G) \end{pmatrix} w_G(v,u) \tag{7.2.30}$$

with

$$w_G(v,u) = \min\left\{1, \frac{b}{\left[|A_{\rm rg}^G v|^2 \Lambda_f^{\rm loc}(u)^2 + (A_{\rm sc}^G)^2 (u\Lambda_f^{\rm loc}(u) - z_{\rm sc}^G)^2\right]^{1/2}}\right\} \tag{7.2.31}$$

where $A_{\rm rg}^G = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$ and the representations (7.2.20)–(7.2.22) hold. We additionally get, $\mathrm{E}\,\tilde{\eta}_{{\rm rg},j}^G v_i \Lambda_f^{\rm loc} = 0$ for $i, j = 1, \ldots, k$, $i \ne j$ since $\tilde{\eta}_{{\rm rg},j}^G$ is an odd function in v_j and $\mathcal{L}(V)$ is symmetric in v_j. By this and (F1) we have, $\tilde{\eta}_{\rm rg}^G(v,u)$ is the unique solution to the Hampel-type problem (1.3.7) in the corresponding regression and scale model with ideal regressor distribution $\mathcal{L}(V)$.
Now, we define $\tilde{\eta}$ of form (7.2.1) where

$$A_{\rm rg} = G^\tau A_{\rm rg}^G G \qquad z_{\rm sc} = z_{\rm sc}^G \qquad A_{\rm sc} = A_{\rm sc}^G \tag{7.2.32}$$

Then, by this definition

$$\mathbb{I}_k = \mathrm{E}\,\tilde{\eta}_{\rm rg}^G v^\tau \Lambda_f^{\rm loc} = \mathrm{E}\,G^\tau \tilde{\eta}_{\rm rg}^G v^\tau G \Lambda_f^{\rm loc} \tag{7.2.33}$$

$$= \mathrm{E}\,G^\tau A_{\rm rg}^G G G^\tau v v^\tau G (\Lambda_f^{\rm loc})^2 \min\left\{1, \frac{b}{\left[|G^\tau A_{\rm rg}^G G G^\tau v|^2 (\Lambda_f^{\rm loc})^2 + (A_{\rm sc}^G)^2 (u\Lambda_f^{\rm loc} - z_{\rm sc}^G)^2\right]^{1/2}}\right\} \tag{7.2.34}$$

$$= \mathrm{E}\,A_{\rm rg} x x^\tau (\Lambda_f^{\rm loc})^2 \min\left\{1, \frac{b}{\left[|A_{\rm rg} x|^2 (\Lambda_f^{\rm loc})^2 + A_{\rm sc}^2 (u\Lambda_f^{\rm loc} - z_{\rm sc})^2\right]^{1/2}}\right\} \tag{7.2.35}$$

$$= \mathrm{E}\,\tilde{\eta}_{\rm rg} x^\tau \Lambda_f^{\rm loc} \tag{7.2.36}$$

and analogously we obtain

$$\mathrm{E}\,\tilde{\eta}_{\mathrm{sc}} = 0 \qquad \mathrm{E}\,\tilde{\eta}_{\mathrm{sc}} u\Lambda_f^{\mathrm{loc}} = 1 \tag{7.2.37}$$

Hence, by (F1) and $\mathrm{E}\,|\tilde{\eta}^G|^2 = \mathrm{E}\,|\tilde{\eta}|^2$ this $\tilde{\eta}$ is the unique solution to the original Hampel-type problem (1.3.7) and the representations (7.2.20)–(7.2.22) hold.

(b) By the arguments provided in part (a) of this proof, the solution is of form (7.2.1) with representations (7.2.20)–(7.2.22) and (7.2.19) and (7.2.32) imply

$$r^2 b = \mathrm{E}\left(\left[|A_{\mathrm{rg}}x|^2(\Lambda_f^{\mathrm{loc}}(u))^2 + A_{\mathrm{sc}}^2(u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}})^2\right]^{1/2} - b\right)_+ \tag{7.2.38}$$

$$= \mathrm{E}\left(\left[|A_{\mathrm{rg}}^G v|^2(\Lambda_f^{\mathrm{loc}}(u))^2 + (A_{\mathrm{sc}}^G)^2(u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}}^G)^2\right]^{1/2} - b\right)_+ \tag{7.2.39}$$

(c) As in part (a) of this proof we first consider the regression model with unknown scale and ideal regressor distribution $\mathcal{L}(V)$ and use only the side conditions (7.2.28). By arguments as in Rieder (1994) (p 198), the minimum bias is always attained. Therefore, if $\bar{\eta}^G$ achieves the minimum bias, then it solves the convex, well-posed problem

$$\sup_P |\eta^G| = \min! \qquad \eta^G \in L_2^{k+1}(\tilde{P}) \tag{7.2.40}$$

under the side conditions (7.2.28). Analogously to the proof of Theorem 5.5.1 (b) in Rieder (1994), the form

$$\bar{\eta}^G(v,u) = \omega_{c,0}^{\min}\begin{pmatrix} \bar{A}_{\mathrm{rg}}^G v\Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}^G(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}^G) \end{pmatrix} Y_G(v,u)^{-1} \tag{7.2.41}$$

on

$$\left\{\begin{pmatrix} \bar{A}_{\mathrm{rg}}^G v\Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}^G(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}^G) \end{pmatrix} \neq 0\right\} \tag{7.2.42}$$

can be obtained and $\bar{A}_{\mathrm{rg}}^G = \mathrm{diag}(\bar{\alpha}_1, \ldots, \bar{\alpha}_k)$, $\bar{z}_{\mathrm{sc}}^G, \bar{A}_{\mathrm{sc}}^G$ achieve the minimum (7.2.25) where $\bar{A}_{\mathrm{rg}}^G \succeq 0$, $\bar{A}_{\mathrm{sc}}^G \geq 0$ and $\bar{A}_{\mathrm{rg}}^G, \bar{A}_{\mathrm{sc}}^G$ not both zero. Thus, using the additional assumption (7.2.24), we obtain the form (7.2.9) of the solution and we get $\bar{\alpha}_j > 0$ and $\bar{A}_{\mathrm{sc}} > 0$ by the constraints $\mathrm{E}\,\eta_{\mathrm{rg,j}} v_j \Lambda_f^{\mathrm{loc}} = 1$ and $\mathrm{E}\,\eta_{\mathrm{sc}}(x,u)u\Lambda_f^{\mathrm{loc}}(u) = 1$, respectively. Moreover, as in part (a) of this proof, the symmetry of the ideal error distribution F and the symmetry of $\mathcal{L}(V)$ yield

$$\mathrm{E}\,\bar{\eta}_{\mathrm{rg}}^G v^\tau \Lambda_f^{\mathrm{loc}} = \mathbb{I}_k \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{rg}}^G = 0 \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{rg}}^G u\Lambda_f^{\mathrm{loc}} = 0 \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{sc}}^G x^\tau \Lambda_f^{\mathrm{loc}} = 0 \tag{7.2.43}$$

Now, we define $\bar{\eta}$ of form (7.2.9) where $A_{\mathrm{rg}} = G^\tau A_{\mathrm{rg}}^G G$, $z_{\mathrm{sc}} = z_{\mathrm{sc}}^G$ and $A_{\mathrm{sc}} = A_{\mathrm{sc}}^G$. This yields,

$$\mathrm{E}\,\bar{\eta}_{\mathrm{rg}} x^\tau \Lambda_f^{\mathrm{loc}} = \mathbb{I}_k \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{sc}} = 0 \qquad \mathrm{E}\,\bar{\eta}_{\mathrm{sc}} u\Lambda_f^{\mathrm{loc}} = 1 \tag{7.2.44}$$

Moreover, by (F1) and $\omega_{c,0}(\bar{\eta}^G) = \omega_{c,0}(\bar{\eta})$ this $\bar{\eta}$ attains the minimum bias (7.2.25) and is the unique solution to the general problem

$$\omega_{c,0}(\eta) = \min! \qquad \eta \in \Psi_2 \tag{7.2.45}$$

////

Spherically Symmetric Regressor Distribution

We conclude this section with a further specialization of the solution (1.3.16).

Proposition 7.2.4 *Assume (F1) and spherical symmetry of the ideal regressor distribution K.*

(a) *In case $\omega_c^{\min} < b \leq \omega_{c,0}(\eta_h)$, the solution $\tilde{\eta}$ to the Hampel-type problem (1.3.7) is of form (7.2.1) where $A_{\mathrm{rg}} = \alpha_{\mathrm{rg}} \mathbb{I}_k$, $\alpha_{\mathrm{rg}} \in \mathbb{R}$ and the following representations hold*

$$\alpha_{\mathrm{rg}} = \big[\mathrm{E}\, x_1^2 (\Lambda_f^{\mathrm{loc}})^2 w \big]^{-1} \tag{7.2.46}$$

$$z_{\mathrm{sc}} = \mathrm{E}\, \Lambda_f^{\mathrm{sc}} w \big[\mathrm{E}\, w \big]^{-1} \qquad A_{\mathrm{sc}} = \big[\mathrm{E}(u\Lambda_f^{\mathrm{loc}} - z_{\mathrm{sc}})^2 w \big]^{-1} \tag{7.2.47}$$

(b) *The solution to the MSE problem (1.3.5) coincides with the solution of part (a), where bound $b \in (0,\infty)$ and radius $r \in (0,\infty)$ are related via*

$$r^2 b = \mathrm{E} \left(\big[\alpha_{\mathrm{rg}}^2 |x|^2 (\Lambda_f^{\mathrm{loc}})^2(u) + A_{\mathrm{sc}}^2 (u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}})^2 \big]^{1/2} - b \right)_+ \tag{7.2.48}$$

(c) *Assume (7.2.24). It holds,*

$$\omega_{c,0}^{\min} = (k + \bar{\alpha}) \big/ \mathrm{E}\, \bar{Y} \tag{7.2.49}$$

where

$$\bar{Y}(x,u) := \big[|x|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + \bar{\alpha}^2 (u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}})^2 \big]^{1/2} \tag{7.2.50}$$

$$\bar{\alpha} = \mathrm{E}\, x_1^2 (\Lambda_f^{\mathrm{loc}})^2 \bar{Y}^{-1} \big[\mathrm{E} (u\Lambda_f^{\mathrm{loc}} - \bar{z}_{\mathrm{sc}})^2 \bar{Y}^{-1} \big]^{-1} \tag{7.2.51}$$

$$\bar{z}_{\mathrm{sc}} = \mathrm{E}\, u\Lambda_f^{\mathrm{loc}} \bar{Y}^{-1} \big[\mathrm{E}\, \bar{Y}^{-1} \big]^{-1} \tag{7.2.52}$$

And then necessarily

$$\bar{\eta}(x,u) = \omega_{c,0}^{\min} \begin{pmatrix} x\Lambda_f^{\mathrm{loc}}(u) \\ \bar{\alpha}(u\Lambda_f^{\mathrm{loc}} - \bar{z}_{\mathrm{sc}}) \end{pmatrix} \bar{Y}(x,u)^{-1} \tag{7.2.53}$$

achieves this minimum bias and is the unique solution. ////

Proof Assume (F1) and spherically symmetric ideal regressor distribution K.

(a)+(b) Direct consequences of Proposition 7.2.3 (a)+(b).

(c) By arguments as in proof of Proposition 7.2.1 (c), the minimum bias $\omega_{c,0}^{\min}$ is of form (7.2.7) and $\bar{\eta}$ of form (7.2.9) attains this minimum bias. Hence, we have to solve the following convex optimization problem

$$\mathrm{E}\, |\bar{Y}| = \min! \qquad \mathrm{tr}\, (A_{\mathrm{rg}}) + A_{\mathrm{sc}} = 1 \tag{7.2.54}$$

which leads to

$$\mathrm{E}\, |\bar{Y}| - \gamma (\mathrm{tr}\, (A_{\mathrm{rg}}) + A_{\mathrm{sc}}) = \min! \tag{7.2.55}$$

with some multiplier $\gamma \in \mathbb{R}$. As in proof of Proposition 7.2.3 (c) we may restrict our attention to diagonal matrices $A_{\mathrm{rg}} = \mathrm{diag}(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^{k \times k}$ because of the elliptical and particularly spherical symmetry of K. Consequentially,

$$\mathrm{E} \Big[\Big(\sum_{i=1}^{k} \alpha_i^2 x_i^2 \Big) (\Lambda_f^{\mathrm{loc}})^2 + \alpha_{k+1}^2 (u\Lambda_f^{\mathrm{loc}} - z_{\mathrm{sc}})^2 \Big]^{1/2} - \gamma \sum_{i=1}^{k+1} \alpha_i = \min! \tag{7.2.56}$$

where $\alpha_{k+1} := A_{\mathrm{sc}}$ and minimization via differentiation with respect to z_{sc} and α_j ($j = 1, \ldots, k+1$) yields (7.2.52),

$$\gamma \;=\; \alpha_j \,\mathrm{E}\, x_j^2 (\Lambda_f^{\mathrm{loc}})^2 Y^{-1} \qquad \forall j = 1, \ldots, k \tag{7.2.57}$$

and

$$\gamma \;=\; \alpha_{k+1} \,\mathrm{E}(u\Lambda_f^{\mathrm{loc}} - z_{\mathrm{sc}})^2 Y^{-1} \tag{7.2.58}$$

where

$$\bar{Y}(x,u) := \Big[\Big(\sum_{i=1}^{k} \alpha_i^2 x_i^2 \Big) \Lambda_f^{\mathrm{loc}}(u)^2 + \alpha_{k+1}^2 (u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}})^2 \Big]^{1/2} \tag{7.2.59}$$

By (7.2.57) and the spherical symmetry of K we get $\alpha_1 = \ldots = \alpha_k =: \tilde{\alpha}$, together with (7.2.58) and $\bar{\alpha} := \alpha_{k+1}/\tilde{\alpha}$ this yields (7.2.49), (7.2.51) and (7.2.53). ////

7.2.1.2 Average Conditional Contamination Neighborhoods

In this subsubsection we specialize and extend the solutions to problems (A.2.1) and (A.2.5) given in Theorem A.2.3 by assuming a symmetric ideal error distribution F and additionally considering linear regression and scale.

Remark 7.2.5 To distinguish between the various AL estimators, we call the optimal AL estimators for average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) ALc estimators. ////

Proposition 7.2.6 *Assume (F1).*

(a) *In case* $\omega_{c,1}^{\min} < b < \omega_{c,1}(\eta_h)$, *the unique solution to Problem (A.2.1) is*

$$\tilde{\eta}(x,u) = \begin{pmatrix} \tilde{\eta}_{\mathrm{rg}}(x,u) \\ \tilde{\eta}_{\mathrm{sc}}(x,u) \end{pmatrix} = \begin{pmatrix} A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u) \\ A_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}}(x)) \end{pmatrix} w(x,u) \tag{7.2.60}$$

with

$$w(x,u) = \min \left\{ 1, \frac{b}{\big[|A_{\mathrm{rg}} x|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + A_{\mathrm{sc}}^2 (u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}}(x))^2 \big]^{1/2}} \right\} \tag{7.2.61}$$

where

$$A_{\mathrm{rg}} = \big[\mathrm{E}\, x x^\tau (\Lambda_f^{\mathrm{loc}})^2 w \big]^{-1} \tag{7.2.62}$$

$$z_{\mathrm{sc}} = \mathrm{E}_\bullet \Lambda_f^{\mathrm{sc}} w \big[\mathrm{E}_\bullet w \big]^{-1} \qquad A_{\mathrm{sc}} = \big[\mathrm{E}(u\Lambda_f^{\mathrm{loc}} - z_{\mathrm{sc}})^2 w \big]^{-1} \tag{7.2.63}$$

(b) *The solution to the MSE problem (A.2.5) coincides with the solution (7.2.60), where bound* $b \in (0, \infty)$ *and radius* $r \in (0, \infty)$ *are related via*

$$r^2 b = \mathrm{E} \left(\big[|A_{\mathrm{rg}} x|^2 (\Lambda_f^{\mathrm{loc}})^2 + A_{\mathrm{sc}}^2 (u\Lambda_f^{\mathrm{loc}} - z_{\mathrm{sc}})^2 \big]^{1/2} - b \right)_+ \tag{7.2.64}$$

(c) *Assume (7.2.6) with* $z \colon \mathbb{R}^k \to \mathbb{R}$. *Then, it holds*

$$\omega_{c,1}^{\min} = \max \left\{ \frac{\mathrm{tr}\, A_{\mathrm{rg}} + A_{\mathrm{sc}}}{\mathrm{E}\, Y} \right\} \tag{7.2.65}$$

with

$$Y(x,u) := \left[|A_{\mathrm{rg}}x|^2\Lambda_f^{\mathrm{loc}}(u)^2 + A_{\mathrm{sc}}^2(u\Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{sc}}(x))^2\right]^{1/2} \tag{7.2.66}$$

where $A_{\mathrm{rg}} \in \mathbb{R}^{k\times k}\backslash\{0\}$, $A_{\mathrm{sc}} \in (0,\infty)$ *and* $z_{\mathrm{sc}}\colon \mathbb{R}^k \to \mathbb{R}$.
There exist $\bar{A}_{\mathrm{rg}} \in \mathbb{R}^{k\times k}$ $(\bar{A}_{\mathrm{rg}} \succ 0)$, $\bar{A}_{\mathrm{sc}} \in (0,\infty)$, $\bar{z}_{\mathrm{sc}}\colon \mathbb{R}^k \to \mathbb{R}$, *and* $\bar{\eta} \in \Psi_{2\bullet}$ *achieving* $\omega_{c,1}^{\min}$, *respectively. And then necessarily*

$$\bar{\eta}(x,u) = \omega_{c,1}^{\min}\begin{pmatrix} \bar{A}_{\mathrm{rg}}x\Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}(x)) \end{pmatrix}\bar{Y}(x,u)^{-1} \tag{7.2.67}$$

with

$$\bar{Y}(x,u) := \left[|\bar{A}_{\mathrm{rg}}x|^2\Lambda_f^{\mathrm{loc}}(u)^2 + \bar{A}_{\mathrm{sc}}^2(u\Lambda_f^{\mathrm{loc}}(u) - \bar{z}_{\mathrm{sc}}(x))^2\right]^{1/2} \tag{7.2.68}$$

is the unique solution.

Proof Assume (F1).

(a) We split $\eta \in L_2^{k+1}(P)$ into the two parts $\eta_{\mathrm{rg}} \in L_2^k(P)$ and $\eta_{\mathrm{sc}} \in L_2(P)$ and modify the Hampel type problem (A.2.1) using only the side conditions

$$\mathrm{E}\,\eta_{\mathrm{rg}}x^\tau\Lambda_f^{\mathrm{loc}} = \mathbb{I}_k \quad \mathrm{E}_\bullet\,\eta_{\mathrm{sc}} = 0 \quad \mathrm{E}\,\eta_{\mathrm{sc}}u\Lambda_f^{\mathrm{loc}} = 1 \tag{7.2.69}$$

besides

$$|\eta_{\mathrm{rg}}|^2 + \eta_{\mathrm{sc}}^2 \le b^2 \qquad \text{a.e. } P \tag{7.2.70}$$

By Lagrange arguments analogous to the proof of Theorem 7.4.13 (a) in Rieder (1994) (corresponds to Theorem A.2.3 (a)), the solution $\tilde{\eta}$ to this weaker problem is of form (7.2.60) and the representations (7.2.62) and (7.2.63) hold. Since

$$\int x\Lambda_f^{\mathrm{loc}}(u)w(x,u)\,F(du) = 0 \qquad \text{for all } x \in \mathbb{R}^k \tag{7.2.71}$$

by symmetry of F, we get $\mathrm{E}_\bullet\,\tilde{\eta} = 0$, if $z_{\mathrm{sc}}(x)$ is determined such that (7.2.63) holds. Moreover, we have

$$\int xu\Lambda_f^{\mathrm{loc}}(u)^2w(x,u)\,F(du) = 0 \qquad \text{for all } x \in \mathbb{R}^k \tag{7.2.72}$$

by symmetry of F and therefore we get $\mathrm{E}\,\tilde{\eta}\Lambda^\tau = \mathbb{I}_{k+1}$, if A_{rg} and A_{sc} are calculated via (7.2.62) and (7.2.63). Thus, this special $\tilde{\eta}$ of form (7.2.60), fulfills the constraints of the original problem (A.2.1) and hence analogously to the sufficient part of Theorem A.2.3 (a) is the solution to that problem, too. The uniqueness of the solution $\tilde{\eta}$ is a consequence of the convexity of $\Psi_{2\bullet}$ and $\omega_{c,1}$ (cf. Remark A.2.1 (b)) and the strict convexity of the norm.

(b) By analogous arguments as in part (a) of this proof and as in the proof of Theorem 5.5.7 in Rieder (1994), the solution is of form (7.2.60) and (A.2.9) becomes (7.2.64).

(c) We split $\eta \in L_2^{k+1}(P)$ into the two parts $\eta_{\mathrm{rg}} \in L_2^k(P)$ and $\eta_{\mathrm{sc}} \in L_2(P)$ and use only the constraints (7.2.69). By arguments as in Rieder (1994) (p 198, 277), the minimum bias is always attained. Therefore, if $\bar{\eta}$ achieves the minimum bias, then it solves the convex, well-posed problem

$$\omega_{c,1}(\eta) = \min! \qquad \eta \in L_2^{k+1}(P) \tag{7.2.73}$$

under the weaker side conditions (7.2.69). Combining the arguments of part (a) of this proof and of the proof of Theorem 5.5.1 (b) in Rieder (1994), the form

$$\bar{\eta}(x,u) = \omega_{c,1}^{\min} \begin{pmatrix} \bar{A}_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}(\Lambda_f^{\mathrm{sc}}(u) - \bar{z}_{\mathrm{sc}}(x)) \end{pmatrix} \bar{Y}^{-1} \, \mathrm{I}\left\{ \begin{pmatrix} \bar{A}_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u) \\ \bar{A}_{\mathrm{sc}}(\Lambda_f^{\mathrm{sc}}(u) - \bar{z}_{\mathrm{sc}}(x)) \end{pmatrix} \neq 0 \right\} \tag{7.2.74}$$

is obtained and $\bar{A}_{\mathrm{rg}}, \bar{A}_{\mathrm{sc}}, \bar{z}_{\mathrm{sc}}$ achieve the minimum (7.2.65) where $\bar{A}_{\mathrm{rg}} \succeq 0$, $\bar{A}_{\mathrm{sc}} \geq 0$ and $\bar{A}_{\mathrm{rg}}, \bar{A}_{\mathrm{sc}}$ not both zero. Thus, using the additional assumption (7.2.6) with $z\colon \mathbb{R}^k \to \mathbb{R}$, we obtain the form (7.2.67) of the solution. Hence, we get $\bar{A}_{\mathrm{rg}} \succ 0$ and $\bar{A}_{\mathrm{sc}} > 0$ by the constraints $\mathrm{E}\, \eta_{\mathrm{rg}} x^\tau \Lambda_f^{\mathrm{loc}} = \mathbb{I}_k$ and $\mathrm{E}\, \eta_{\mathrm{sc}}(x,u) u \Lambda_f^{\mathrm{loc}}(u) = 1$, respectively. Moreover, as in part (a) of this proof, the symmetry of the ideal error distribution F yields

$$\mathrm{E}_\bullet\, \bar{\eta}_{\mathrm{rg}} = 0 \qquad \mathrm{E}\, \bar{\eta}_{\mathrm{rg}} u \Lambda_f^{\mathrm{loc}} = 0 \qquad \mathrm{E}\, \bar{\eta}_{\mathrm{sc}} x^\tau \Lambda_f^{\mathrm{loc}} = 0 \tag{7.2.75}$$

i.e., $\bar{\eta} \in \Psi_{2\bullet}$. ////

Remark 7.2.7 **(a)** By the symmetry of F, the regression part $\tilde{\eta}_{\mathrm{rg}}$ is automatically centered. However, as $\tilde{\eta}_{\mathrm{rg}}$ and $\tilde{\eta}_{\mathrm{sc}}$ are connected via the weight function w, z_{sc} also depends on x. Moreover, by this connection of the two parts of the optimally robust IC, unlike to the proof of Theorem 7.4.13 in Rieder (1994), we can not apply Lemma C.2.4 of Rieder (1994). Hence, the (product) measurability of $z_{\mathrm{sc}}(x)$ remains to be shown.

(b) The regression part $\tilde{\eta}_{\mathrm{rg}}$ of the solution $\tilde{\eta}$ is a redescending function in u given x as in the unconditional case; confer Remark 7.2.2 (a).

(c) Similarly to the unconditional case (cf. Remark 7.2.2 (b)), we get the following connection between the Lagrange multipliers contained in the optimal ICs at θ_0 and θ

$$A_{\mathrm{rg},\theta} = \sigma^2 A_{\mathrm{rg},\theta_0} \qquad A_{\mathrm{sc},\theta} = \sigma^2 A_{\mathrm{sc},\theta_0} \qquad z_\theta(x) = \sigma^{-1} z_{\theta_0}(x) \qquad b_\theta = \sigma b_{\theta_0} \tag{7.2.76}$$

confer also Theorem 2.4.1. That is, $\tilde{\eta}_\theta$ is the solution to the corresponding Hampel type problem (A.2.1) with bound $b_\theta = \sigma b_{\theta_0}$. ////

Elliptically Symmetric Regressor Distribution

Analogously to case of unconditional neighborhoods, the solution (7.2.60) simplifies by additionally assuming that the ideal regressor distribution K is elliptically symmetric; i.e., (7.2.19) holds.

Proposition 7.2.8 *Assume (F1) and (7.2.19).*

(a) *In case $\omega_{c,1}^{\min} < b < \omega_{c,1}(\eta_h)$, the unique solution to Problem (A.2.1) is of form (7.2.60), where $A_{\mathrm{rg}} = G^\tau A_{\mathrm{rg}}^G G$ with $A_{\mathrm{rg}}^G = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$, $z_{\mathrm{sc}}(v) = z_{\mathrm{sc}}^G(v)$,*

$A_{\mathrm{sc}} = A^G_{\mathrm{sc}}$ *and the following representations hold*

$$\alpha_j = \big[\,\mathrm{E}\, v_j^2(\Lambda_f^{\mathrm{loc}})^2 w_{_G}\big]^{-1} \qquad j = 1, \dots k \tag{7.2.77}$$

$$z^G_{\mathrm{sc}} = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} w_{_G} \big[\,\mathrm{E}_\bullet\, w_{_G}\big]^{-1} \qquad A^G_{\mathrm{sc}} = \big[\,\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 w_{_G}\big]^{-1} \tag{7.2.78}$$

Here

$$w_{_G}(v,u) = \min\left\{1, \frac{b}{\big[|A^G_{\mathrm{rg}} v|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + (A^G_{\mathrm{sc}})^2 (\Lambda_f^{\mathrm{sc}}(u) - z^G_{\mathrm{sc}}(v))^2\big]^{1/2}}\right\} \tag{7.2.79}$$

and E *is expectation taken under* $\tilde{P}(du, dv) := f(u)\lambda(du)\,\mathcal{L}(V)(dv)$.

(b) *The solution to the corresponding MSE problem (A.2.5) coincides with the solution of part (a), where bound* $b \in (0, \infty)$ *and radius* $r \in (0, \infty)$ *are related via*

$$r^2 b = \mathrm{E}\left(\big[|A^G_{\mathrm{rg}} v|^2 (\Lambda_f^{\mathrm{loc}})^2 + (A^G_{\mathrm{sc}})^2 (\Lambda_f^{\mathrm{sc}} - z^G_{\mathrm{sc}}(v))^2\big]^{1/2} - b\right)_+ \tag{7.2.80}$$

(c) *Assume (7.2.24) with* $z\colon \mathbb{R}^k \to \mathbb{R}$. *It holds,*

$$\omega_{c,1}^{\min} = \max\left\{\frac{\mathrm{E}\, Y_{_G}}{\sum_{j=1}^k \alpha_j + A^G_{\mathrm{sc}}}\right\} \tag{7.2.81}$$

with

$$Y_{_G}(v,u) := \big[|A^G_{\mathrm{rg}} v|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + (A^G_{\mathrm{sc}})^2 (\Lambda_f^{\mathrm{sc}}(u) - z^G_{\mathrm{sc}}(v))^2\big]^{1/2} \tag{7.2.82}$$

where $A^G_{\mathrm{rg}} = \mathrm{diag}(\alpha_1, \dots, \alpha_k)$, $z^G_{\mathrm{sc}}\colon \mathbb{R}^k \to \mathbb{R}$, $\alpha_1, \dots, \alpha_k, A^G_{\mathrm{sc}} \in (0, \infty)$.
There exist $\bar{A}^G_{\mathrm{rg}} = \mathrm{diag}(\bar{\alpha}_1, \dots, \bar{\alpha}_k)$ *with* $\bar{\alpha}_1, \dots, \bar{\alpha}_k \in (0, \infty)$, $\bar{z}^G_{\mathrm{sc}}\colon \mathbb{R}^k \to \mathbb{R}$, $\bar{A}^G_{\mathrm{sc}} \in (0, \infty)$ *and* $\bar{\eta} \in \Psi_{2\bullet}$ *achieving* $\omega_{c,1}^{\min}$, *respectively. And then necessarily the unique solution* $\bar{\eta}$ *is of form (7.2.67), where* $\bar{A}_{\mathrm{rg}} = G^\tau \bar{A}^G_{\mathrm{rg}} G$, $\bar{z}_{\mathrm{sc}}(v) = \bar{z}^G_{\mathrm{sc}}(v)$ *and* $\bar{A}_{\mathrm{sc}} = \bar{A}^G_{\mathrm{sc}}$.

Proof This can be shown by analogous arguments as in the proofs of Proposition 7.2.6 and Proposition 7.2.3. ////

Spherically Symmetric Regressor Distribution

The solution $\tilde{\eta}$ can be further simplified if the ideal regressor distribution K is spherically symmetric.

Proposition 7.2.9 *Assume (F1) and spherical symmetry of the ideal regressor distribution* K.

(a) *In case* $\omega_{c,1}^{\min} < b < \omega_{c,1}(\eta_h)$, *the solution* $\tilde{\eta}$ *to the Hampel-type problem (A.2.1) is of form (7.2.60) where* $A_{\mathrm{rg}} = \alpha_{\mathrm{rg}} \mathbb{I}_k$, $\alpha_{\mathrm{rg}} \in \mathbb{R}$ *and the following representations hold*

$$\alpha_{\mathrm{rg}} = \big[\,\mathrm{E}\, x_1^2 (\Lambda_f^{\mathrm{loc}})^2 w\big]^{-1} \tag{7.2.83}$$

$$z_{\mathrm{sc}} = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} w \big[\,\mathrm{E}_\bullet\, w\big]^{-1} \qquad A_{\mathrm{sc}} = \big[\,\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 w\big]^{-1} \tag{7.2.84}$$

(b) *The solution to the corresponding MSE problem (A.2.5) coincides with the solution of part (a), where bound* $b \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *are related via*

$$r^2 b = \mathrm{E}\left(\left[\alpha_{\mathrm{rg}}^2 |x|^2 (\Lambda_f^{\mathrm{loc}})^2 + A_{\mathrm{sc}}^2 (\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2\right]^{1/2} - b\right)_+ \tag{7.2.85}$$

(c) *Assume (7.2.24). It holds,*

$$\omega_{c,1}^{\min} = (k + \bar{\alpha}) / \mathrm{E}\,\bar{Y} \tag{7.2.86}$$

where

$$\bar{Y}(x,u) := \left[|x|^2 \Lambda_f^{\mathrm{loc}}(u)^2 + \bar{\alpha}^2 (\Lambda_f^{\mathrm{sc}}(u) - \bar{z}_{\mathrm{sc}}(x))^2\right]^{1/2} \tag{7.2.87}$$

$$\bar{z}_{\mathrm{sc}} = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} \bar{Y}^{-1} \left[\mathrm{E}_\bullet\, \bar{Y}^{-1}\right]^{-1} \tag{7.2.88}$$

$$\bar{\alpha} = \mathrm{E}\, x_1^2 (\Lambda_f^{\mathrm{loc}})^2 \bar{Y}^{-1} \left[\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 \bar{Y}^{-1}\right]^{-1} \tag{7.2.89}$$

And then necessarily

$$\bar{\eta}(x,u) = \omega_{c,1}^{\min} \begin{pmatrix} x \Lambda_f^{\mathrm{loc}}(u) \\ \bar{\alpha}(\Lambda_f^{\mathrm{sc}}(u) - \bar{z}_{\mathrm{sc}}(x)) \end{pmatrix} \bar{Y}(x,u)^{-1} \tag{7.2.90}$$

achieves this minimum bias and is the unique solution. ////

Proof This can be shown by analogous arguments as in the proofs of Proposition 7.2.6 and Proposition 7.2.4 ////

7.2.2 M Estimators

7.2.2.1 Unconditional Contamination Neighborhoods

Since $\mathrm{E}\,|\rho|^2$ is not necessarily convex in ψ, we instead consider ICs ρ of the form

$$\rho(x,u) = \begin{pmatrix} Ax\psi(x,u) \\ \gamma(u\psi(x,u) - 1) \end{pmatrix} \tag{7.2.91}$$

where $A \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma \in \mathbb{R}\setminus\{0\}$ are fixed. Given some regular and symmetric $A \in \mathbb{R}^{k\times k}$ and some $\gamma \in \mathbb{R}\setminus\{0\}$ we solve the problems (1.3.7) and (1.3.5). Thus, in a second step we have to minimize the results in $A \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma \in \mathbb{R}\setminus\{0\}$, to solve the problems for ρ of form (7.1.47). Moreover, given $A \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma \in \mathbb{R}\setminus\{0\}$, we derive sufficient conditions for the minimum bias $\omega_{\mathrm{M}}^{\min}$ and specify the lower case solution $\bar{\rho}$, where we additionally assume ideal error distribution $F = \mathcal{N}(0,1)$.

To simplify notation, we use the vec and vech operators as defined in Appendix B. Moreover, we use

Lemma 7.2.10 *Assume* $\mathrm{E}_K\,|x|^4 < \infty$. *Then, the following statements are pairwise equivalent*

(a) $\mathrm{rk}\,\mathrm{E}\,\mathrm{vec}\,(xx^\tau)\mathrm{vec}(xx^\tau)^\tau < k(k+1)/2$

(b) $\det \mathrm{E}\,\mathrm{vech}\,(xx^\tau)\mathrm{vech}(xx^\tau)^\tau = 0$

(c) x *is a.e.* $K(dx)$ *concentrated on a conic.*

PROOF The proof is similar to the proof of Proposition B.5.1 in Ruckdeschel (2001).
(a) $\Longleftrightarrow$ **(b):** Let

$$\mathrm{E}\,\mathrm{vec}\,(xx^\tau)\mathrm{vec}(xx^\tau)^\tau =: C = (C_{i_1,i_2,j_1,j_2})_{1\le i_1,i_2,j_1,j_2\le k} \tag{7.2.92}$$

that is,

$$\mathrm{E}\,\mathrm{vech}\,(xx^\tau)\mathrm{vech}(xx^\tau)^\tau = (C_{i_1,i_2,j_1,j_2})_{1\le i_1,j_1\le i_2,j_2\le k} =: C_{\mathrm{h}} \tag{7.2.93}$$

Then, by symmetry of xx^τ we get $\mathrm{rk}\,C \le k(k+1)/2$. Moreover, by $\det C_{\mathrm{h}} = 0$, there exists some $0 \ne A \in \mathbb{R}^{k(k+1)/2}$ such that

$$0 = A^\tau C_{\mathrm{h}} A = \sum_{1\le i_1,j_1\le i_2,j_2\le k} A_{i_1,i_2} C_{i_1,i_2,j_1,j_2} A_{j_1,j_2} \tag{7.2.94}$$

$$= \sum_{1\le i_1,i_2,j_1,j_2\le k} \tilde{A}_{i_1,i_2} C_{i_1,i_2,j_1,j_2} \tilde{A}_{j_1,j_2} \tag{7.2.95}$$

$$= \tilde{A}^\tau C \tilde{A} \tag{7.2.96}$$

where $\tilde{A} \in \mathbb{R}^{k^2}$ with

$$\tilde{A}_{i_1,i_2} := \begin{cases} A_{i_1,i_2}/2 & \text{if } i_1 < i_2 \\ A_{i_1,i_2} & \text{if } i_1 = i_2 \\ A_{i_2,i_1}/2 & \text{if } i_1 > i_2 \end{cases} \tag{7.2.97}$$

(b) $\Longleftrightarrow$ **(c):** Equation (7.2.96) yields

$$A^\tau \mathrm{vech}\,(xx^\tau) = \tilde{A}^\tau \mathrm{vec}\,(xx^\tau) = 0 \qquad \text{a.e. } K(dx) \tag{7.2.98}$$

By Remark B.1.4, we get with $B := \mathrm{vech}^{-1}(A) = \mathrm{vec}^{-1}(\tilde{A})$

$$0 = A^\tau \mathrm{vech}\,(xx^\tau) = \tilde{A}^\tau \mathrm{vec}\,(xx^\tau) = \mathrm{tr}\,Bxx^\tau = \mathrm{tr}\,x^\tau Bx = x^\tau Bx \tag{7.2.99}$$

that is, x is a.e. $K(dx)$ concentrated on a conic. ////

This leads us to the following optimal solutions.

Theorem 7.2.11 *Assume (F1). Additionally, let* $\mathrm{E}_K\,|x|^4 < \infty$ *and*

$$\mathrm{rk}\left[\mathrm{E}_K\,\mathrm{vec}\,(xx^\tau)\mathrm{vec}\,(xx^\tau)^\tau\right] = k(k+1)/2 \tag{7.2.100}$$

(a) *In case of* $A \in \mathbb{R}^{k\times k}$ *regular and symmetric,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma,\infty)$ *fixed, there exist some* $\alpha_1, \alpha_3 \in \mathbb{R}$ *and some symmetric* $B \in \mathbb{R}^{k\times k}$ *such that the solution* $\tilde{\rho}$ *to problem (1.3.7) is based via (7.2.91) on* $\tilde{\psi}$ *of the form*

$$\tilde{\psi}(x,u) = \frac{\alpha_1 u + x^\tau Bx \Lambda_f^{\mathrm{loc}}(u) + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)} w(x,u) + \gamma^2 \frac{u}{q^2(x,u)} \tag{7.2.101}$$

with

$$w(x,u) = \min\Big\{1, \frac{b(x,u)}{|g(x,u)|}\Big\} \tag{7.2.102}$$

where

$$q(x,u) = \sqrt{|Ax|^2 + \gamma^2 u^2} \tag{7.2.103}$$

$$g(x,u) = \frac{\alpha_1 u + x^\tau B x \Lambda_f^{\mathrm{loc}}(u) + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}(u)}{q(x,u)} \tag{7.2.104}$$

$$b(x,u) = \Big[b^2 - \gamma^2 \frac{|Ax|^2}{q^2(x,u)}\Big]^{1/2} \tag{7.2.105}$$

Conversely, if there is some $\tilde\rho \in \Psi_{\mathrm{M}}$ *with* $\tilde\psi$ *of form (7.2.101) for any* $A \in \mathbb{R}^{k\times k}$ *regular and symmetric,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma,\infty)$ *fixed,* $\alpha_1, \alpha_3 \in \mathbb{R}$, $B \in \mathbb{R}^{k\times k}$ *symmetric, then* $\tilde\rho$ *is the solution, and the following representations hold,*

$$\alpha_1 = \Big[C_1 - \mathrm{E}\,\frac{x^\tau B x u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q^2} w\Big] \Big/ \mathrm{E}\,\frac{u^2}{q^2} w \tag{7.2.106}$$

$$\operatorname{vech} B = \Big[H_k\, \mathrm{E}\operatorname{vec}(xx^\tau)\operatorname{vec}(xx^\tau)^\tau \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2} w\, G_k\Big]^{-1} \operatorname{vech} D \tag{7.2.107}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\,\frac{\alpha_1 u^3 \Lambda_f^{\mathrm{loc}} + x^\tau B x u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\,\frac{u^4 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.108}$$

where G_k *is the duplication matrix (cf. Definition B.2.3),* H_k *is an arbitrary left-inverse of* G_k *with full row rank and*

$$C_1 = \mathrm{E}\,\frac{|Ax|^2}{q^2} \in \mathbb{R} \tag{7.2.109}$$

$$C_2 = A^{-1} - \gamma^2\, \mathrm{E}\, xx^\tau \frac{u\Lambda_f^{\mathrm{loc}}}{q^2} \in \mathbb{R}^{k\times k} \quad \textit{(symmetric)} \tag{7.2.110}$$

$$D = C_2 - \mathrm{E}\, xx^\tau \frac{\alpha_1 u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \in \mathbb{R}^{k\times k} \quad \textit{(symmetric)} \tag{7.2.111}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2\, \mathrm{E}\,\frac{u^3 \Lambda_f^{\mathrm{loc}}}{q^2} \in \mathbb{R} \tag{7.2.112}$$

(b) *The unique solution to the corresponding MSE problem (1.3.5) coincides with the solution to problem (1.3.7), where bound* $b \in [\gamma,\infty)$ *and radius* $r \in (0,\infty)$ *are related via*

$$r^2 = \mathrm{E}\Big(\frac{|g(x,u)|}{b(x,u)} - 1\Big)_+ \qquad \textit{if } b > \gamma \tag{7.2.113}$$

respectively

$$r^2 \geq \mathrm{E}\Big(\frac{|g(x,u)|}{b(x,u)} - 1\Big)_+ \qquad \textit{if } b = \gamma \tag{7.2.114}$$

(c) *Now, additionally assume ideal error distribution* $F = \mathcal{N}(0,1)$.
Given $A \in \mathbb{R}^{k\times k}$ *regular and symmetric and* $\gamma \in \mathbb{R}\setminus\{0\}$ *let* $\bar b(A,\gamma) \in [\gamma,\infty)$ *be*

minimal such that

$$C_1 \le \mathrm{E}\,\tfrac{|u|}{q(x,u)}\bar{b}(x,u) \tag{7.2.115}$$

$$0 \preceq \mathrm{E}\,xx^\tau \tfrac{|u|}{q(x,u)}\bar{b}(x,u) - C_2 \tag{7.2.116}$$

and

$$C_3 \le \mathrm{E}\,\tfrac{|u|^3}{q(x,u)}\bar{b}(x,u) \tag{7.2.117}$$

where

$$\bar{b}(x,u) = \Big[\bar{b}(A,\gamma)^2 - \gamma^2 \tfrac{|Ax|^2}{q^2(x,u)}\Big]^{1/2} \tag{7.2.118}$$

Then, $\omega_{\mathrm{M}}^{\min} \ge \bar{b}(A,\gamma)$*. If there is equality in (7.2.115)-(7.2.117) for*

$$\bar{b}(\bar{A},\bar{\gamma}) = \min_{A,\gamma} \bar{b}(A,\gamma) \tag{7.2.119}$$

then $\bar{\zeta}$ *of form*

$$\bar{\zeta}(x,u) = b_{\min}(x,u)\,\mathrm{sign}\,(u) \qquad b_{\min}(x,u) = \Big[\big(\omega_{\mathrm{M}}^{\min}\big)^2 - \bar{\gamma}^2 \tfrac{|\bar{A}x|^2}{|Ax|^2+\bar{\gamma}^2u^2}\Big]^{1/2} \tag{7.2.120}$$

is the unique solution, $\omega_{\mathrm{M}}^{\min} = \bar{b}(\bar{A},\bar{\gamma})$*, and the corresponding* $\bar{\rho}$ *attains this minimum bias.*

Proof Assume (F1).

(a) First we may rewrite

$$|\rho(x,u)|^2 = \zeta^2(x,u) + \gamma^2 \frac{|Ax|^2}{q^2(x,u)} \tag{7.2.121}$$

where

$$\zeta(x,u) = q(x,u)\psi(x,u) - \gamma^2 \frac{u}{q(x,u)} \tag{7.2.122}$$

Hence, problem (1.3.7) is equivalent to

$$\mathrm{E}\,\zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta},\ \sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0 \tag{7.2.123}$$

where

$$\Psi_{\mathrm{M}_\zeta} = \big\{\zeta \in L_2(P) \,\big|\, \zeta \text{ fulfills } (\mathrm{M}_\zeta 1)\text{–}(\mathrm{M}_\zeta 5)\big\} \tag{7.2.124}$$

and the conditions $(\mathrm{M}_\zeta 1)$–$(\mathrm{M}_\zeta 5)$ translate from (M1)–(M5) on ψ to ζ

$(\mathbf{M}_\zeta\,\mathbf{1})$ $\mathrm{E}\,\frac{x}{q(x,u)}\zeta(x,u) = 0$

$(\mathbf{M}_\zeta\,\mathbf{2})$ $\mathrm{E}\,\frac{u}{q(x,u)}\zeta(x,u) = \mathrm{E}\,\frac{|Ax|^2}{q^2(x,u)} =: C_1 \in \mathbb{R}$

$(\mathbf{M}_\zeta\,\mathbf{3})$ $\mathrm{E}\,xx^\tau \frac{\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)}\zeta(x,u) = A^{-1} - \gamma^2\,\mathrm{E}\,xx^\tau \frac{u\Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)} =: C_2 \in \mathbb{R}^{k\times k}$ (symmetric)

$(\mathbf{M}_\zeta\,\mathbf{4})$ $\mathrm{E}\,\frac{u^2\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)}\zeta(x,u) = 1 + \gamma^{-1} - \gamma^2\,\mathrm{E}\,\frac{u^3\Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)} =: C_3 \in \mathbb{R}$

($\mathbf{M}_\zeta$ 5) $\mathrm{E}\, x \frac{u\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)}\zeta(x,u) = 0$

We will solve this convex problem under (M_ζ 2)–(M_ζ 4) and then (M_ζ 1) and (M_ζ 5) will be fulfilled automatically.
Since expectation and scalar products are linear weakly continuous and since also $\sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0$ is convex and weakly lower semicontinuous analogously to Lemma 1.3.4, the problem (7.2.123) is a minimum norm problem over a convex closed subset in Hilbert space. This problem has a unique solution $\tilde{\zeta}$ provided there exists some $\zeta \in \Psi_{\mathrm{M}_\zeta}$ that satisfies

$$\sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0 \tag{7.2.125}$$

To show that this optimization problem is well-posed in sense of Definition B.2.9 in Rieder (1994), we need to verify the well-posedness assumptions B.2(18), B.2(22) and B.2(23) of Theorem B.2.6 of Rieder (1994).
Given $b > \omega_{\mathrm{M}}^{\min}$, we assume some ζ_1 with (M_ζ 2)–(M_ζ 4) and

$$\sup_P \big[|\zeta_1(x,u)| - b_1(x,u)\big] \le 0 \qquad b_1(x,u) := \Big[b_1^2 - \gamma^2 \frac{|Ax|^2}{q^2(x,u)}\Big]^{1/2} \tag{7.2.126}$$

where $\omega_{\mathrm{M}}^{\min} < b_1 < b$. Therefore, by concavity of the root

$$b(x,u) - b_1(x,u) \ge b - b_1 > 0 \tag{7.2.127}$$

Now, consider

$$\zeta_Q(x,u) := \zeta_1(x,u) + x^\tau Q x\, \mathrm{I}(|x| \le M_1)|u|\Lambda_f^{\mathrm{loc}}\, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \tag{7.2.128}$$

where $M_1, M_2 \in (0,\infty)$ and $Q \in \mathbb{R}^{k\times k}$ symmetric with $\|Q\|_{\mathrm{op}} \le \frac{b-b_1}{M_1^2 M_2}$. Then, by ($\mathrm{M}_\zeta$ 2)

$$\begin{aligned} &\mathrm{E}\, \frac{u}{q(x,u)}\, \zeta_Q(x,u) = \\ &\quad = C_1 + \mathrm{E}_K\, x^\tau Q x\, \mathrm{I}(|x| \le M_1)\, \mathrm{E}_\bullet\, \frac{|u|}{q(x,u)} u\Lambda_f^{\mathrm{loc}}\, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \end{aligned} \tag{7.2.129}$$

where

$$0 < \mathrm{E}_\bullet\, \frac{|u|}{q(x,u)} u\Lambda_f^{\mathrm{loc}}\, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \le |\gamma|^{-1}\, \mathrm{E}_\bullet (u\Lambda_f^{\mathrm{loc}})^+ < \infty \tag{7.2.130}$$

since $\mathrm{E}_\bullet\, u\Lambda_f^{\mathrm{loc}} = 1$. Moreover, by ($\mathrm{M}_\zeta$ 4)

$$\begin{aligned} &\mathrm{E}\, \frac{u^2\Lambda_f^{\mathrm{loc}}}{q(x,u)}\, \zeta_Q(x,u) = \\ &\quad = C_3 + \mathrm{E}_K\, x^\tau Q x\, \mathrm{I}(|x| \le M_1)\, \mathrm{E}_\bullet\, \frac{|u|(u\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)}\, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \end{aligned} \tag{7.2.131}$$

where

$$0 < \mathrm{E}_\bullet \frac{|u|}{q(x,u)} (u\Lambda_f^{\mathrm{loc}}(u))^2 \, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \le |\gamma|^{-1} (\mathcal{I}_f^{\mathrm{sc}} + 1) < \infty \tag{7.2.132}$$

by (F3). Hence, in both cases we have to consider expressions of the form

$$\mathrm{E}_K \, xQx^\tau h(x) \, \mathrm{I}(|x| \le M_1) = \mathrm{tr} \left[\mathrm{E}_K \, xx^\tau h(x) \, \mathrm{I}(|x| \le M_1) \, Q \right] \tag{7.2.133}$$

where h is positive and uniformly bounded. Thus, we get by dominated convergence and assumption (K), which is $\mathrm{rk}\, \mathrm{E}_K \, xx^\tau = k$,

$$\lim_{M_1 \to \infty} \mathrm{E}_K \, xx^\tau h(x) \, \mathrm{I}(|x| \le M_1) = \mathrm{E}_K \, xx^\tau h(x) \succ 0 \tag{7.2.134}$$

which by continuity of the determinant and for M_1 sufficiently large yields

$$\mathrm{E}_K \, xx^\tau h(x) \, \mathrm{I}(|x| \le M_1) \succ 0 \tag{7.2.135}$$

Then, letting Q vary over the symmetric $k \times k$-matrices of the special form $\alpha \mathbb{I}_k$, where $\alpha \in \mathbb{R}$ with $|\alpha| \le \frac{b-b_1}{M_1^2 M_2}$, the corresponding set of functions ζ_Q already covers a full neighborhood of C_1 and C_3. Moreover, we get by ($\mathrm{M}_\zeta 3$)

$$\begin{aligned} &\mathrm{E}\, xx^\tau \frac{\Lambda_f^{\mathrm{loc}}}{q(x,u)} \zeta_Q(x,u) \\ &\quad = C_2 + \mathrm{E}_K \, xx^\tau (x^\tau Q x) \, \mathrm{I}(|x| \le M_1) \, \mathrm{E}_\bullet \frac{|u| (\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)} \, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \end{aligned} \tag{7.2.136}$$

where

$$0 < h(x) := \mathrm{E}_\bullet \frac{|u| (\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)} \, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) \le |\gamma|^{-1} \mathcal{I}_f^{\mathrm{loc}} < \infty \tag{7.2.137}$$

by (F2). Using the vec and vech operators, equation (7.2.136) becomes

$$\begin{aligned} &\mathrm{vech}\, \mathrm{E}\, xx^\tau \frac{\Lambda_f^{\mathrm{loc}}}{q(x,u)} \zeta_Q(x,u) \\ &\qquad = \mathrm{vech}\, C_2 + \mathrm{E}_K \, \mathrm{vech}\, [(xx^\tau) Q (xx^\tau)] h(x) \, \mathrm{I}(|x| \le M_1) \end{aligned} \tag{7.2.138}$$

By Theorem B.1.3 and $(xx^\tau) \otimes (xx^\tau) = \mathrm{vec}\,(xx^\tau) \mathrm{vec}\,(xx^\tau)^\tau$, we get

$$\begin{aligned} &\mathrm{E}_K \, \mathrm{vech}\, [(xx^\tau) Q (xx^\tau)] h(x) \, \mathrm{I}(|x| \le M_1) \\ &\qquad = \mathrm{E}_K \, H_k \mathrm{vec}\, [(xx^\tau) Q (xx^\tau)] h(x) \, \mathrm{I}(|x| \le M_1) && (7.2.139) \\ &\qquad = H_k \, \mathrm{E}_K \, \mathrm{vec}\,(xx^\tau) \mathrm{vec}\,(xx^\tau)^\tau \mathrm{vec}\, Q \, h(x) \, \mathrm{I}(|x| \le M_1) && (7.2.140) \\ &\qquad = H_k \, \mathrm{E}_K \, \mathrm{vec}\,(xx^\tau) \mathrm{vec}\,(xx^\tau)^\tau h(x) \, \mathrm{I}(|x| \le M_1) G_k \mathrm{vech}\, Q && (7.2.141) \end{aligned}$$

confer also Remark B.2.4 (c). Here $G_k \in \mathbb{R}^{k^2 \times k(k+1)/2}$ is the duplication matrix as introduced in Definition B.2.3, which has full column rank and $H_k \in \mathbb{R}^{k(k+1)/2 \times k^2}$

is an arbitrary left inverse of G_k (one choice is $H_k = (G_k^\tau G_k)^{-1} G_k^\tau$), which has full row rank. Since $\operatorname{vech} Q$ covers a full neighborhood of $0 \in \mathbb{R}^{k(k+1)/2}$, the corresponding set of functions ζ_Q covers a full neighborhood of C_2, if

$$\operatorname{rk} \mathrm{E}_K \operatorname{vec}(xx^\tau) \operatorname{vec}(xx^\tau)^\tau h(x)\, \mathrm{I}(|x| \le M_1) = \frac{k(k+1)}{2} \tag{7.2.142}$$

By Lemma 7.2.10 assumption (7.2.100) is equivalent to

$$\det \mathrm{E}_K \operatorname{vech}(xx^\tau) \operatorname{vech}(xx^\tau)^\tau > 0 \tag{7.2.143}$$

and as $h(x)$ is positive this is also equivalent to

$$\det \mathrm{E}_K \operatorname{vech}(xx^\tau) \operatorname{vech}(xx^\tau)^\tau h(x) > 0 \tag{7.2.144}$$

Thus, by dominated convergence and continuity of the determinant we get for sufficiently large M_1

$$\det \mathrm{E}_K \operatorname{vech}(xx^\tau) \operatorname{vech}(xx^\tau)^\tau h(x)\, \mathrm{I}(|x| \le M_1) > 0 \tag{7.2.145}$$

which by arguments analogously to the proof of Lemma 7.2.10 is equivalent to (7.2.142). Finally, it holds

$$\begin{aligned} &|\zeta_Q| - b(x,u) \\ &\quad \le |\zeta_1| + |x^\tau Q x|\, \mathrm{I}(|x| \le M_1) |u \Lambda_f^{\mathrm{loc}}(u)|\, \mathrm{I}(0 < u\Lambda_f^{\mathrm{loc}} \le M_2) - b(x,u) && (7.2.146) \\ &\quad \le b_1(x,u) - b(x,u) + \|Q\|_{\mathrm{op}} M_1^2 M_2 \le b_1 - b + b - b_1 = 0 && (7.2.147) \end{aligned}$$

Therefore, the set $\left\{ \zeta_Q \,\middle|\, Q \in \mathbb{R}^{k\times k} \text{ symmetric}, \|Q\|_{\mathrm{op}} \le \frac{b-b_1}{M_1^2 M_2} \right\}$ covers full neighborhoods of C_1, C_2 and C_3, and $\sup_P \left[|\zeta_Q(x,u)| - b(x,u) \right] \le 0$. Hence, problem (7.2.123) is well-posed and Theorem B.2.6 of Rieder (1994) is applicable. Thus, there exists some multipliers $\alpha_1, \alpha_3 \in \mathbb{R}$ and $B \in \mathbb{R}^{k\times k}$ symmetric, such that the solution $\tilde\zeta$ minimizes the Lagrangian

$$\begin{aligned} L(\zeta) &= \mathrm{E}\, \zeta^2 - 2\alpha_1 \mathrm{E}\, \tfrac{u}{q} \zeta - 2\, \mathrm{E}\, x^\tau B x \tfrac{\Lambda_f^{\mathrm{loc}}}{q} \zeta - 2\alpha_3 \mathrm{E}\, \tfrac{u^2 \Lambda_f^{\mathrm{loc}}}{q} \zeta && (7.2.148) \\ &= \mathrm{E} \left[\zeta - \frac{\alpha_1 u + x^\tau B x \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q} \right]^2 + \mathrm{const} && (7.2.149) \end{aligned}$$

among all $\zeta \in L_2(P_{\theta_0})$ satisfying (7.2.125). This Lagrangian can be minimized by pointwise minimization of the integrand subject to (7.2.125). Then, by (7.2.122) the asserted form (7.2.101) of the optimal $\tilde\psi$ is attained. Since $\tilde\zeta(x,.)$ is odd in u a.e. $K(dx)$, ($\mathrm{M}_\zeta 1$) and ($\mathrm{M}_\zeta 5$) are automatically fulfilled; confer Remark 7.1.9 (b).

Conversely, if there is some $\tilde\rho \in \Psi_{\mathrm{M}}$ with $\tilde\psi$ of form (7.2.101) for any $\alpha_1, \alpha_3 \in \mathbb{R}$ and $B \in \mathbb{R}^{k\times k}$ symmetric, then $\tilde\zeta$ minimizes the corresponding Lagrangian L in (7.2.149) among all $\zeta \in L_2(P)$ subject to (7.2.125) and therefore $\tilde\zeta$ solves (7.2.123).

To get the representations (7.2.106) and (7.2.108), respectively we solve ($\mathrm{M}_\zeta 2$) and

$(\mathrm{M}_\zeta 4)$ for α_1 and α_3, respectively. To prove (7.2.107), we rewrite $(\mathrm{M}_\zeta 3)$ using the vec and vech operator

$$\operatorname{vech} D = \operatorname{vech} \mathrm{E}(xx^\tau)B(xx^\tau)\tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w \tag{7.2.150}$$

$$= \mathrm{E}\, H_k \operatorname{vec}\left[(xx^\tau)B(xx^\tau)\right]\tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w \tag{7.2.151}$$

$$= H_k\, \mathrm{E} \operatorname{vec}(xx^\tau)\operatorname{vec}(xx^\tau)^\tau \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w\ \operatorname{vec} B \tag{7.2.152}$$

$$= H_k\, \mathrm{E} \operatorname{vec}(xx^\tau)\operatorname{vec}(xx^\tau)^\tau \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w\ G_k \operatorname{vech} B \tag{7.2.153}$$

where D is the symmetric $k\times k$-matrix defined in (7.2.110) and H_k is an arbitrary left inverse of the duplication matrix G_k. Since

$$\mathrm{E}_\bullet \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w > 0 \tag{7.2.154}$$

we get by arguments as in the proof of the well-posedness before

$$\operatorname{rg}\ \mathrm{E}_K \operatorname{vec}(xx^\tau)\operatorname{vec}(xx^\tau)^\tau \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2}w = \tfrac{k(k+1)}{2} \tag{7.2.155}$$

Therefore, equation (7.2.107) follows by the full column rank of G_k and the full row rank of H_k.

(b) The MSE problem (1.3.5) can equivalently be rewritten as

$$\mathrm{E}\,\zeta^2 + r^2\omega_{\mathrm{M}}^2(\rho) + \mathrm{const} = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta} \tag{7.2.156}$$

Putting $b := \omega_{\mathrm{M}}(\tilde\rho)$, hence $\tilde\zeta$ also solves the problem

$$\mathrm{E}\,\zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta},\ \sup_P \left[|\zeta(x,u)| - b(x,u)\right] \le 0 \tag{7.2.157}$$

which has already occurred as problem (7.2.123) and the solution $\tilde\zeta$, respectively $\tilde\psi$ is of form (7.2.101). Plugged into the Lagrangian L, we get a convex function of the bound b

$$L_1(b) = \mathrm{E}\left(|g(x,u)| - b(x,u)\right)_+^2 + r^2b^2 \tag{7.2.158}$$

Analogously to Lemma C.2.3 of Rieder (1994), we may differentiate L_1 by b, which yields (7.2.113), respectively (7.2.114).

(c) Now, additionally assume ideal error distribution $F = \mathcal{N}(0,1)$.

Before proving the sufficient conditions (7.2.115)–(7.2.117), we give the sketch of a proof using Lagrange arguments as in the proof of Theorem 5.5.1 (b) in Rieder (1994).

By arguments as in Rieder (1994) (p 198) the minimum bias is attained; i.e., if $\bar\zeta$ achieves this minimum bias, then it solves the convex, well-posed problem

$$\sup_P |\zeta| = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta} \tag{7.2.159}$$

Thus, there exists some multipliers $\alpha_1, \alpha_3 \in \mathbb{R}$ and $B \in \mathbb{R}^{k\times k}$ symmetric, such that the solution $\bar\zeta$ maximizes the Lagrangian

$$L(\zeta) = \mathrm{E}\,\zeta\Big[\frac{\alpha_1 u + x^\tau Bxu + \alpha_3 u^3}{q}\Big] \tag{7.2.160}$$

among all $\zeta \in L_2(P)$ and under the constraint $\sup_P [|\zeta| - \bar{b}] \leq 0$. This can be solved by pointwise maximization of the integrand and we get

$$\bar{\zeta}(x,u) = \bar{b}(x,u) \operatorname{sign}\left[\frac{(\alpha_1 + x^\tau B x)u + \alpha_3 u^3}{q}\right] \tag{7.2.161}$$

$$= \bar{b}(x,u) \operatorname{sign}(u) \tag{7.2.162}$$

if $\alpha(x) := \alpha_1 + x^\tau B x \geq 0$ for all x and $\alpha_3 \geq 0$.
However, we can not guarantee by Lagrange arguments that $\alpha(x)$ and α_3 are both positive. Hence, we now give the proof for the sufficient conditions (7.2.115)–(7.2.117), which imply this form of the solution.
For arbitrary $\alpha_1, \alpha_3 \geq 0$, $B \in \mathbb{R}^{k\times k}$ positive semidefinite and symmetric and $\zeta \in \Psi_{\mathrm{M}_\zeta}$ it holds

$$\alpha_1 C_1 + \operatorname{tr}(BC_2) + \alpha_3 C_3 = \mathrm{E}\,\frac{\alpha_1 u + x^\tau B x u + \alpha_3 u^3}{q(x,u)}\zeta \tag{7.2.163}$$

$$\leq \mathrm{E}\,\frac{\alpha(x)|u| + \alpha_3|u|^3}{q(x,u)}|\zeta| \tag{7.2.164}$$

$$\leq \mathrm{E}\,\frac{\alpha(x)|u| + \alpha_3|u|^3}{q(x,u)}b(x,u) \tag{7.2.165}$$

$$= \mathrm{E}\,\alpha(x)\tfrac{|u|}{q(x,u)}b(x,u) + \alpha_3\,\mathrm{E}\,\tfrac{|u|^3}{q(x,u)}b(x,u) \tag{7.2.166}$$

As the right hand side of this equation is positive, if not all three α_1, α_3, B are equal to zero, we obtain the following lower bound for b

$$1 \geq \frac{\alpha_1 C_1 + \operatorname{tr}(BC_2) + \alpha_3 C_3}{\mathrm{E}\,\alpha(x)\tfrac{|u|}{q(x,u)}b(x,u) + \alpha_3\,\mathrm{E}\,\tfrac{|u|^3}{q(x,u)}b(x,u)} =: \frac{d_1 + d_2\alpha_3}{d_3 + d_4\alpha_3} \tag{7.2.167}$$

Given $\alpha(\cdot) \geq 0$ the maximization of the right hand side via differentiation in α_3 yields that the maximum is attained for $\alpha_3 = 0$, if $d_2 d_3 < d_1 d_4$, respectively, for $\alpha_3 = \infty$, if $d_2 d_3 > d_1 d_4$. In case of $\alpha_3 = \infty$, (7.2.167) yields the bound (7.2.117) for b. If $\alpha_3 = 0$, we get

$$1 \geq \frac{\alpha_1 C_1 + \operatorname{tr}(BC_2)}{\mathrm{E}\,\alpha(x)\tfrac{|u|}{q(x,u)}b(x,u)} \tag{7.2.168}$$

$$= \frac{\alpha_1 C_1 + \operatorname{tr}(BC_2)}{\alpha_1\,\mathrm{E}\,\tfrac{|u|}{q(x,u)}b(x,u) + \mathrm{E}\,x^\tau B x\tfrac{|u|}{q(x,u)}b(x,u)} =: \frac{d_1 + d_2\alpha_1}{d_3 + d_4\alpha_1} \tag{7.2.169}$$

Now, given B positive semidefinite and symmetric the maximization of the right hand side via differentiation in α_1 yields that the maximum is attained for $\alpha_1 = 0$, if $d_2 d_3 < d_1 d_4$, respectively, for $\alpha_1 = \infty$, if $d_2 d_3 > d_1 d_4$. In case of $\alpha_1 = \infty$, (7.2.169) yields the bound (7.2.115) for b. If $\alpha_1 = 0$, we get

$$1 \geq \frac{\operatorname{tr}(BC_2)}{\mathrm{E}\,x^\tau B x\tfrac{|u|}{q(x,u)}b(x,u)} \tag{7.2.170}$$

which is equivalent to

$$0 \leq \operatorname{tr} B\big(\operatorname{E} xx^\tau \tfrac{|u|}{q(x,u)} b(x,u) - C_2\big) =: \operatorname{tr} B\Delta \tag{7.2.171}$$

Since B is symmetric and positive semidefinite, there exists an orthogonal matrix U, such that $UBU^\tau = S = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$, where $\lambda_j \geq 0$ for all $j = 1, \ldots, k$. Then, it holds for all matrices B symmetric and positive semidefinite

$$\operatorname{tr} B\Delta = \operatorname{tr} SU\Delta U^\tau =: \operatorname{tr} SM = \sum_{j=1}^{k} \lambda_j m_{jj} \geq 0 \tag{7.2.172}$$

with m_{jj} the j-th diagonal element of M. This is equivalent to $m_{jj} \geq 0$ for all $j = 1, \ldots, k$ and implies the bound (7.2.116), if we choose B such that $U\Delta U^\tau$ is diagonal.
We finally determine $\bar{b}(\bar{A}, \bar{\gamma})$. By (7.2.115)–(7.2.117) necessarily, $\omega_{\mathrm{M}}^{\min} \geq \bar{b}(\bar{A}, \bar{\gamma})$ since $(\mathrm{M}_\zeta 2)$–$(\mathrm{M}_\zeta 4)$ must hold for every $\zeta \in \Psi_{\mathrm{M}_\zeta}$. But, if we have equality in (7.2.115)–(7.2.117) then, $\omega_{\mathrm{M}}^{\min} = \bar{b}(\bar{A}, \bar{\gamma})$, since $\bar{\zeta}(x,u)$ as defined in (7.2.120) fulfills $(\mathrm{M}_\zeta 1)$–$(\mathrm{M}_\zeta 5)$. Hence, $\bar{\zeta}$ is the solution to problem (7.2.159) and the corresponding $\bar{\rho}$ achieves the minimum bias $\omega_{\mathrm{M}}^{\min}$. ////

Remark 7.2.12 (a) As the optimization in A and γ becomes more and more complicated with increasing dimension k, we also consider M estimators with $A = \mathcal{K}^{-1}$ fixed; i.e., we only minimize the results in $\gamma \in \mathbb{R} \setminus \{0\}$. We therefore call this estimators MK estimators. In case of location and scale the MK estimators are identical to the general M estimators. However, there are differences in regression and scale. The additional optimization in A supplies a clear reduction of the MSE; confer Section 7.5.

(b) Starting from $\tilde{\psi}_{\theta_0}$ and defining $\tilde{\psi}_\theta$ by (7.1.54) and fixing the same regular and symmetric $A \in \mathbb{R}^{k \times k}$ and the same $\gamma \in \mathbb{R}$, we get that this $\tilde{\psi}_\theta$ is also of form (7.2.101) with

$$\alpha_{1,\theta} = \sigma^2 \alpha_{1,\theta_0} \qquad \alpha_{3,\theta} = \sigma^2 \alpha_{3,\theta_0} \qquad B_\theta = \sigma^2 B_{\theta_0} \qquad b_\theta = \sigma b_{\theta_0} \tag{7.2.173}$$

Hence, $\tilde{\psi}_\theta$ is the solution to the corresponding Hampel type problem (1.3.7) with bound $b_\theta = \sigma b_{\theta_0}$. ////

Elliptically Symmetric Regressor Distribution

We now specialize the solutions given in Theorem 7.2.11 by additionally assuming, that the ideal regressor distribution K is elliptically symmetric; i.e., (7.2.19) holds. Thus

$$A_\psi^{-1} = \operatorname{E} xx^\tau \Lambda_f^{\mathrm{loc}} \psi(x,u) = G^\tau \operatorname{E} vv^\tau \Lambda_f^{\mathrm{loc}} \psi(G^\tau v, u) G \tag{7.2.174}$$

Moreover, $\mathcal{L}(V)$ is symmetric in v_j for all $j = 1, \ldots, k$; that is, $\mathcal{L}(V)$ is invariant under multiplication by matrices B_j of the form $B_j := \operatorname{diag}(1, \ldots, 1, -1, 1, \ldots, 1)$ with $j{-}th$ diagonal element equal to -1. Therefore, we get the following invariance

$$\tilde{A}_\psi^{-1} := GA_\psi^{-1}G^\tau = \operatorname{E} vv^\tau \Lambda_f^{\mathrm{loc}} \psi = B_j \operatorname{E} vv^\tau \Lambda_f^{\mathrm{loc}} \psi B_j = B_j \tilde{A}_\psi^{-1} B_j \tag{7.2.175}$$

for all $j = 1, \ldots, k$, which implies $\tilde{A}_\psi = \operatorname{diag}(a_1, \ldots, a_k)$. Hence, we only need to optimize over all regular diagonal matrices $A_G \in \mathbb{R}^{k\times k}$ and then obtain the optimal matrix A by $A = G^\tau A_G G$.

Proposition 7.2.13 *Assume (F1) and (7.2.19). Moreover, let* $\mathrm{E}_{\mathcal{L}(V)} |v|^4 < \infty$ *and* E *be expectation taken under* $\tilde{P}(du, dv) := f(u)\lambda(du)\,\mathcal{L}(V)(dv)$.

(a) *In case of* $A_G = \operatorname{diag}(a_1, \ldots, a_k) \in \mathbb{R}^{k\times k}$ *regular,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma, \infty)$ *fixed, the unique solution of Problem (1.3.7) is based via (7.2.91) on* $\tilde{\psi}$ *of the form (7.2.101), where* $A = G^\tau A_G G$, $B = G^\tau B_G G$ *with* $B_G = \operatorname{diag}(b_1, \ldots, b_k)$ *and the following representations hold* ($j = 1, \ldots, k$)

$$\alpha_1 = \Big[C_1 - \mathrm{E}\,\tfrac{v^\tau B_G v u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q_G^2} w_G\Big] \Big/ \mathrm{E}\,\tfrac{u^2}{q_G^2} w_G \tag{7.2.176}$$

$$b_j = \Big[C_{2,j} - \mathrm{E}\,\tfrac{\alpha_1 u \Lambda_f^{\mathrm{loc}} + \sum_{i=1, i\neq j}^k b_i v_i^2 v_j^2 (\Lambda_f^{\mathrm{loc}})^2 + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G\Big] \Big/ \mathrm{E}\,\tfrac{v_j^4 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G \tag{7.2.177}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\,\tfrac{\alpha_1 u^3 \Lambda_f^{\mathrm{loc}} + v^\tau B_G v u^2 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G\Big] \Big/ \mathrm{E}\,\tfrac{u^4 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G \tag{7.2.178}$$

where

$$q_G(v,u) = \big[|A_G v|^2 + \gamma^2 u^2\big]^{1/2} \qquad w_G(v,u) = \min\Big\{1, \frac{b_G(v,u)}{|g_G(v,u)|}\Big\} \tag{7.2.179}$$

$$b_G(v,u) = \Big[b^2 - \gamma^2 \tfrac{|A_G v|^2}{q_G(v,u)^2}\Big]^{1/2} \qquad g_G(v,u) = \tfrac{\alpha_1 u + v^\tau B_G v \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q_G(v,u)} \tag{7.2.180}$$

and

$$C_1 = \mathrm{E}\,\tfrac{|A_G v|^2}{q_G^2} \tag{7.2.181}$$

$$C_{2,j} = a_j^{-1} - \gamma^2\,\mathrm{E}\,v_j^2 \tfrac{u \Lambda_f^{\mathrm{loc}}}{q_G^2} \qquad j = 1, \ldots, k \tag{7.2.182}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2\,\mathrm{E}\,\tfrac{u^3 \Lambda_f^{\mathrm{loc}}}{q_G^2} \tag{7.2.183}$$

(b) *The unique solution to the corresponding MSE problem (1.3.5) coincide with the solution of problem (1.3.7), where bound* $b \in [\gamma, \infty)$ *and radius* $r \in (0, \infty)$ *are related via*

$$r^2 = \mathrm{E}\,\Big(\frac{|g_G(v,u)|}{b_G(v,u)} - 1\Big)_+ \qquad \text{if } b > \gamma \tag{7.2.184}$$

respectively

$$r^2 \geq \mathrm{E}\,\Big(\frac{|g_G(v,u)|}{b_G(v,u)} - 1\Big)_+ \qquad \text{if } b = \gamma \tag{7.2.185}$$

(c) *Now, additionally assume ideal error distribution* $F = \mathcal{N}(0,1)$.
Given $A_G = \operatorname{diag}(a_1, \ldots, a_k) \in \mathbb{R}^{k\times k}$ *regular and* $\gamma \in \mathbb{R}\setminus\{0\}$ *let* $\bar{b}(A_G, \gamma) \in [\gamma, \infty)$

be minimal such that

$$C_1 \leq \mathrm{E}\, \tfrac{|u|}{q_G(v,u)} \bar{b}_G(v,u) \tag{7.2.186}$$

$$C_{2,j} \leq \mathrm{E}\, v_j^2 \tfrac{|u|}{q_G(v,u)} \bar{b}_G(v,u) \qquad j = 1,\ldots,k \tag{7.2.187}$$

and

$$C_3 \leq \mathrm{E}\, \tfrac{|u|^3}{q_G(v,u)} \bar{b}_G(v,u) \tag{7.2.188}$$

where

$$\bar{b}_G(v,u) = \left[\bar{b}(A_G,\gamma)^2 - \gamma^2 \tfrac{|A_G v|^2}{q_G(v,u)^2}\right]^{1/2} \tag{7.2.189}$$

Then, $\omega_{\mathrm{M}}^{\min} \geq \bar{b}(A_G,\gamma)$*. If there is equality in (7.2.186)-(7.2.188) for*

$$\bar{b}(\bar{A}_G,\bar{\gamma}) = \min_{A_G,\gamma} \bar{b}(A_G,\gamma) \tag{7.2.190}$$

then $\bar{\zeta}$ *of form*

$$\bar{\zeta}(x,u) = b_{\min}(x,u)\,\mathrm{sign}\,(u) \qquad b_{\min}(x,u) = \left[\left(\omega_{\mathrm{M}}^{\min}\right)^2 - \bar{\gamma}^2 \tfrac{|\bar{A}x|^2}{|Ax|^2+\bar{\gamma}^2u^2}\right]^{1/2} \tag{7.2.191}$$

with $\bar{A} = G^\tau \bar{A}_G G$ *is the unique solution,* $\omega_{\mathrm{M}}^{\min} = \bar{b}(\bar{A}_G,\bar{\gamma})$*, and the corresponding* $\bar{\rho}$ *attains this minimum bias.*

Proof This may be shown by analogous arguments as in the proofs of Proposition 7.2.3 and Theorem 7.2.11. ////

Spherically Symmetric Regressor Distribution

We again give a second specialization of the solutions given in Theorem 7.2.11 by additionally assuming, that the ideal regressor distribution K is spherically symmetric; i.e, $\mathcal{L}(GX) = \mathcal{L}(X)$ for all orthogonal matrices $G \in \mathbb{R}^{k\times k}$. Thus

$$A_\psi^{-1} = \mathrm{E}\, xx^\tau \Lambda_f^{\mathrm{loc}} \psi(x,u) = G\,\mathrm{E}\, xx^\tau \Lambda_f^{\mathrm{loc}} \psi(x,u) G^\tau = G A_\psi^{-1} G^\tau \tag{7.2.192}$$

for all orthogonal matrices $G \in \mathbb{R}^{k\times k}$, which by Lemma 3.2 of Rieder et al. (2001) implies $A_\psi = a\mathbb{I}_k$ with $a \in \mathbb{R}\setminus\{0\}$. Hence, we only need to optimize over all constants $a \in \mathbb{R}\setminus\{0\}$ and then obtain the optimal matrix A by $A = a\mathbb{I}_k$.

Proposition 7.2.14 Assume (F1) and spherical symmetry of the ideal regressor distribution K, where $\mathrm{E}_K\,|x|^4 < \infty$.

(a) In case of $a \in \mathbb{R}\setminus\{0\}$, $\gamma \in \mathbb{R}\setminus\{0\}$ and $b \in [\gamma,\infty)$ fixed, the unique solution of Problem (1.3.7) is based via (7.2.91) on $\tilde{\psi}$ of the form (7.2.101), where $A = a\mathbb{I}_k$,

$B = d\mathbb{I}_k$ with $d \in \mathbb{R}$ and the following representations hold

$$\alpha_1 = \Big[C_1 - \mathrm{E}\,\tfrac{d|x|^2 u\Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q^2} w\Big] \Big/ \mathrm{E}\,\tfrac{u^2}{q^2} w \tag{7.2.193}$$

$$d = \Big[C_2 - \mathrm{E}\,\tfrac{\alpha_1 u\Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\,\tfrac{x_1^2|x|^2(\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.194}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\,\tfrac{\alpha_1 u^3\Lambda_f^{\mathrm{loc}} + d|x|^2 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\,\tfrac{u^4(\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.195}$$

where

$$q(x,u) = \big[a^2|x|^2 + \gamma^2 u^2\big]^{1/2} \qquad w(x,u) = \min\Big\{1, \frac{b(x,u)}{|g(x,u)|}\Big\} \tag{7.2.196}$$

$$b(x,u) = \Big[b^2 - \gamma^2 \tfrac{a^2|x|^2}{q(x,u)^2}\Big]^{1/2} \qquad g(x,u) = \tfrac{\alpha_1 u + d|x|^2\Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q(x,u)} \tag{7.2.197}$$

and

$$C_1 = \mathrm{E}\,\tfrac{a^2|x|^2}{q^2} \qquad C_2 = a^{-1} - \gamma^2\,\mathrm{E}\,x_1^2 \tfrac{u\Lambda_f^{\mathrm{loc}}}{q^2} \tag{7.2.198}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2\,\mathrm{E}\,\tfrac{u^3\Lambda_f^{\mathrm{loc}}}{q^2} \tag{7.2.199}$$

(b) The unique solution to the corresponding MSE problem (1.3.5) coincide with the solution of problem (1.3.7), where bound $b \in [\gamma, \infty)$ and radius $r \in (0, \infty)$ are related via (7.2.113), respectively (7.2.114).

(c) Now, additionally assume ideal error distribution $F = \mathcal{N}(0,1)$.
Given $a \in \mathbb{R} \setminus \{0\}$ and $\gamma \in \mathbb{R} \setminus \{0\}$ let $\bar{b}(a,\gamma) \in [\gamma, \infty)$ be minimal such that

$$C_1 \le \mathrm{E}\,\tfrac{|u|}{q(x,u)} \bar{b}(x,u) \tag{7.2.200}$$

$$C_2 \le \mathrm{E}\,x_1^2 \tfrac{|u|}{q(x,u)} \bar{b}(x,u) \tag{7.2.201}$$

and

$$C_3 \le \mathrm{E}\,\tfrac{|u|^3}{q(x,u)} \bar{b}(x,u) \tag{7.2.202}$$

where

$$\bar{b}(x,u) = \Big[\bar{b}(a,\gamma)^2 - \gamma^2 \tfrac{a^2|x|^2}{q(x,u)^2}\Big]^{1/2} \tag{7.2.203}$$

Then, $\omega_{\mathrm{M}}^{\min} \ge \bar{b}(a,\gamma)$. If there is equality in (7.2.200)–(7.2.202) for

$$\bar{b}(\bar{a},\bar{\gamma}) = \min_{a,\gamma} \bar{b}(a,\gamma) \tag{7.2.204}$$

then $\bar{\zeta}$ of form

$$\bar{\zeta}(x,u) = b_{\min}(x,u)\,\mathrm{sign}\,(u) \qquad b_{\min}(x,u) = \Big[\big(\omega_{\mathrm{M}}^{\min}\big)^2 - \bar{\gamma}^2 \tfrac{\bar{a}^2|x|^2}{\bar{a}^2|x|^2 + \bar{\gamma}^2 u^2}\Big]^{1/2} \tag{7.2.205}$$

is the unique solution, $\omega_{\mathrm{M}}^{\min} = \bar{b}(\bar{a},\bar{\gamma})$, and the corresponding $\bar{\rho}$ attains this minimum bias. ////

Proof This may be shown by analogous arguments as in the proofs of Proposition 7.2.4 and Theorem 7.2.11. ////

7.2.2.2 Average Conditional Contamination Neighborhoods

We again consider ICs ρ of the form (7.2.91). Then, given some regular and symmetric $A \in \mathbb{R}^{k\times k}$ and some $\gamma \in \mathbb{R}\setminus\{0\}$ we solve the problems (A.2.1) and (A.2.5), where we additionally need

$$\begin{aligned} F(\{0 < u\Lambda_f^{\mathrm{loc}} \le 1\} \cap \{|\Lambda_f^{\mathrm{loc}}| \le |u|\}) \neq 0 \\ F(\{1 < u\Lambda_f^{\mathrm{loc}} < \infty\} \cap \{|\Lambda_f^{\mathrm{loc}}| \ge |u|\}) \neq 0 \end{aligned} \tag{7.2.206}$$

for the proof of well-posedness. In a second step we then have to minimize the results in $A \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma \in \mathbb{R}\setminus\{0\}$, to solve this problems for ρ of form (7.1.47). As in case of unconditional neighborhoods, this can be done via numerical minimization.
Moreover, given $A \in \mathbb{R}^{k\times k}$ regular and symmetric and $\gamma \in \mathbb{R}\setminus\{0\}$ we derive sufficient conditions for the minimum bias $\omega_{\mathrm{Mc}}^{\min}$ and specify the lower case solution $\bar{\rho}$, where we additionally assume ideal error distribution $F = \mathcal{N}(0,1)$.

We again use the vec and vech operators defined in Appendix B and Lemma 7.2.10 for deriving the following optimal solutions.

Theorem 7.2.15 *Assume (F1) and (7.2.206). Additionally, let* $\mathrm{E}_K\,|x|^4 < \infty$ *and*

$$\mathrm{rk}\left[\mathrm{E}_K\,\mathrm{vec}\,(xx^\tau)\mathrm{vec}\,(xx^\tau)^\tau\right] = k(k+1)/2 \tag{7.2.207}$$

(a) *In case of* $A \in \mathbb{R}^{k\times k}$ *regular and symmetric,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma,\infty)$ *fixed, there exist some* $\alpha_1\colon \mathbb{R}^k \to \mathbb{R}$, $\alpha_3 \in \mathbb{R}$ *and some symmetric* $B \in \mathbb{R}^{k\times k}$ *such that the solution* $\tilde{\rho}$ *to problem (A.2.1) is based via (7.2.91) on* $\tilde{\psi}$ *of the form*

$$\tilde{\psi}(x,u) = \frac{\alpha_1(x)u + x^\tau Bx\Lambda_f^{\mathrm{loc}}(u) + \alpha_3 u^2\Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)}w(x,u) + \gamma^2\frac{u}{q^2(x,u)} \tag{7.2.208}$$

with

$$w(x,u) = \min\left\{1, \frac{b(x,u)}{|g(x,u)|}\right\} \tag{7.2.209}$$

where

$$q(x,u) = \sqrt{|Ax|^2 + \gamma^2u^2} \tag{7.2.210}$$

$$g(x,u) = \frac{\alpha_1(x)u + x^\tau Bx\Lambda_f^{\mathrm{loc}}(u) + \alpha_3 u^2\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)} \tag{7.2.211}$$

$$b(x,u) = \left[b^2 - \gamma^2\frac{|Ax|^2}{q^2(x,u)}\right]^{1/2} \tag{7.2.212}$$

Conversely, if there is some $\tilde{\rho} \in \Psi_{\mathrm{Mc}}$ *with* $\tilde{\psi}$ *of form (7.2.208) for any* $A \in \mathbb{R}^{k\times k}$ *regular and symmetric,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma,\infty)$, *fixed,* $\alpha_1\colon \mathbb{R}^k \to \mathbb{R}$, $\alpha_3 \in \mathbb{R}$, $B \in \mathbb{R}^{k\times k}$ *symmetric, then* $\tilde{\rho}$ *is the solution, and the following representations*

hold,

$$\alpha_1(x) = \Big[C_1(x) - \mathrm{E}_\bullet \frac{x^\tau B x u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q^2} w\Big] \Big/ \mathrm{E}_\bullet \frac{u^2}{q^2} w \tag{7.2.213}$$

$$\operatorname{vech} B = \Big[H_k \,\mathrm{E}\operatorname{vec}(xx^\tau)\operatorname{vec}(xx^\tau)^\tau \tfrac{(\Lambda_f^{\mathrm{loc}})^2}{q^2} w\, G_k\Big]^{-1} \operatorname{vech} D \tag{7.2.214}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\,\frac{\alpha_1 u^3 \Lambda_f^{\mathrm{loc}} + x^\tau B x u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\,\frac{u^4 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.215}$$

where

$$C_1(x) = \mathrm{E}_\bullet \frac{|Ax|^2}{q^2} \in \mathbb{R} \tag{7.2.216}$$

$$C_2 = A^{-1} - \gamma^2 \,\mathrm{E}\, xx^\tau \frac{u \Lambda_f^{\mathrm{loc}}}{q^2} \in \mathbb{R}^{k\times k} \tag{7.2.217}$$

$$D = C_2 - \mathrm{E}\, xx^\tau \frac{\alpha_1 u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \in \mathbb{R}^{k\times k} \tag{7.2.218}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2 \,\mathrm{E}\,\frac{u^3 \Lambda_f^{\mathrm{loc}}}{q^2} \in \mathbb{R} \tag{7.2.219}$$

(b) *The unique solution to the corresponding MSE problem (A.2.5) coincide with the solution of problem (A.2.1), where bound $b \in [\gamma, \infty)$ and radius $r \in (0, \infty)$ are related via*

$$r^2 = \mathrm{E}\Big(\frac{|g(x,u)|}{b(x,u)} - 1\Big)_+ \qquad \text{if } b > \gamma \tag{7.2.220}$$

respectively

$$r^2 \geq \mathrm{E}\Big(\frac{|g(x,u)|}{b(x,u)} - 1\Big)_+ \qquad \text{if } b = \gamma \tag{7.2.221}$$

Proof Assume (F1) and (7.2.206).

(a) Analogous to the proof of Theorem 7.2.11 problem (A.2.1) is equivalent to

$$\mathrm{E}\,\zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{Mc}_\zeta},\ \sup_P \big[|\zeta(x,u)| - b(x,u)\big] \leq 0 \tag{7.2.222}$$

where

$$\Psi_{\mathrm{Mc}_\zeta} = \big\{\zeta \in L_2(P) \,\big|\, \zeta \text{ fulfills } (\mathrm{Mc}_\zeta 1),\ (\mathrm{Mc}_\zeta 2),\ (\mathrm{M}_\zeta 3)\text{–}(\mathrm{M}_\zeta 5)\big\} \tag{7.2.223}$$

and the conditions $(\mathrm{Mc}_\zeta 1)$, $(\mathrm{Mc}_\zeta 2)$, $(\mathrm{M}_\zeta 3)$–$(\mathrm{M}_\zeta 5)$ translate from (Mc1), (Mc2), (M3)–(M5) on ψ to ζ

$(\mathrm{Mc}_\zeta\,1)$ $\mathrm{E}_\bullet \frac{x}{q(x,u)} \zeta(x,u) = 0$

$(\mathrm{Mc}_\zeta\,2)$ $\mathrm{E}_\bullet \frac{u}{q(x,u)} \zeta(x,u) = \mathrm{E}_\bullet \frac{|Ax|^2}{q^2(x,u)} =: C_1(x) \in \mathbb{R}$

$(\mathrm{M}_\zeta\,3)$ $\mathrm{E}\, xx^\tau \frac{\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)} \zeta(x,u) = A^{-1} - \gamma^2 \,\mathrm{E}\, xx^\tau \frac{u \Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)} =: C_2 \in \mathbb{R}^{k\times k}$ (symmetric)

($\mathbf{M}_\zeta$ 4) $\mathrm{E}\,\frac{u^2\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)}\zeta(x,u) = 1 + \gamma^{-1} - \gamma^2\,\mathrm{E}\,\frac{u^3\Lambda_f^{\mathrm{loc}}(u)}{q^2(x,u)} =: C_3 \in \mathbb{R}$

($\mathbf{M}_\zeta$ 5) $\mathrm{E}\,x\frac{u\Lambda_f^{\mathrm{loc}}(u)}{q(x,u)}\zeta(x,u) = 0$

We will solve this convex problem under (Mc_ζ 2), (M_ζ 3) and (M_ζ 4) and then (Mc_ζ 1) and (M_ζ 5) will be fulfilled automatically. Since expectation and scalar products are linear weakly continuous and also $\sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0$ is convex and weakly lower semicontinuous analogously (cf. Remark A.2.1 (b)), the problem (7.2.222) is a minimum norm problem over a convex closed subset in Hilbert space. This problem has a unique solution $\tilde{\zeta}$ provided there exists some $\zeta \in \Psi_{\mathrm{Mc}_\zeta}(\theta_0)$ that satisfies

$$\sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0 \tag{7.2.224}$$

In general, there is no Lagrange multiplier for the side condition (Mc_ζ 2); confer Rieder (1994) (p 276). Thus, we first keep (Mc_ζ 2) and (7.2.224) back in the domain and show that the resulting optimization problem is well-posed in sense of Definition B.2.9 in Rieder (1994).
Given $b > \omega_{\mathrm{Mc}}^{\min}$ we assume some ζ_1 with (Mc_ζ 2), (M_ζ 3), (M_ζ 4) and

$$\sup_P \big[|\zeta_1(x,u)| - b_1(x,u)\big] \le 0 \qquad b_1(x,u) := \Big[b_1^2 - \gamma^2\frac{|Ax|^2}{q^2(x,u)}\Big]^{1/2} \tag{7.2.225}$$

where $\omega_{\mathrm{Mc}}^{\min} < b_1 < b$. Therefore, by concavity of the root

$$b(x,u) - b_1(x,u) \ge b - b_1 > 0 \tag{7.2.226}$$

Then, consider

$$\zeta_Q(x,u) := \zeta_1(x,u) + x^\tau Q x\,\mathrm{I}(|x| \le M_1)|u|\Lambda_f^{\mathrm{loc}}(u)\big[\,\mathrm{I}_1 - \delta(x)\,\mathrm{I}_2\,\big] \tag{7.2.227}$$

where

$$\mathrm{I}_1 := \mathrm{I}(1 < u\Lambda_f^{\mathrm{loc}} \le M_2)\,\mathrm{I}(|\Lambda_f^{\mathrm{loc}}| \ge |u|) \tag{7.2.228}$$

$$\mathrm{I}_2 := \mathrm{I}(\varepsilon < u\Lambda_f^{\mathrm{loc}} \le 1)\,\mathrm{I}(|\Lambda_f^{\mathrm{loc}}| \le |u|) \tag{7.2.229}$$

$$\delta(x) := \mathrm{E}_\bullet\,\tfrac{|u|}{q}u\Lambda_f^{\mathrm{loc}}\,\mathrm{I}_1 \Big/ \big[\,\mathrm{E}_\bullet\,\tfrac{|u|}{q}u\Lambda_f^{\mathrm{loc}}\,\mathrm{I}_2\,\big]^{-1} \in (0,\infty) \tag{7.2.230}$$

$M_1 \in (0,\infty)$, $M_2 \in (1,\infty)$, $\varepsilon \in (0,1)$ and $Q \in \mathbb{R}^{k\times k}$ symmetric with bounded operator norm $\|Q\|_{\mathrm{op}} \le \frac{b-b_1}{M_1^2\max\{M_2,|\delta|\}}$. By definition of $\delta(x)$ and assumption (7.2.206) we get

$$\mathrm{E}\,\frac{u}{q(x,u)}\zeta_Q(x,u) = 0 \tag{7.2.231}$$

Moreover, by (M_ζ 4)

$$\begin{aligned}&\mathrm{E}\,\frac{u^2\Lambda_f^{\mathrm{loc}}}{q(x,u)}\zeta_Q(x,u)\\ &\quad = C_3 + \mathrm{E}_K\,x^\tau Q x\,\mathrm{I}(|x| \le M_1)\,\mathrm{E}_\bullet\,\frac{|u|(u\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)}\big[\,\mathrm{I}_1 - \delta(x)\,\mathrm{I}_2\,\big]\end{aligned} \tag{7.2.232}$$

where by assumption (7.2.206) and (F3), which is $\mathcal{I}_F^{\mathrm{sc}} < \infty$,

$$0 < h(x) := \mathrm{E}_\bullet \frac{|u|}{q(x,u)} (u\Lambda_f^{\mathrm{loc}}(u))^2 \big[\, \mathrm{I}_1 - \delta(x)\, \mathrm{I}_2 \,\big] \le \mathrm{E}_\bullet \frac{|u|}{q(x,u)} (u\Lambda_f^{\mathrm{loc}}(u))^2\, \mathrm{I}_1 \quad (7.2.233)$$

$$\le |\gamma|^{-1} (\mathcal{I}_f^{\mathrm{sc}} + 1) < \infty \quad (7.2.234)$$

Hence, we have to consider the following expression

$$\mathrm{E}_K\, xQx^\tau h(x)\, \mathrm{I}(|x| \le M_1) = \mathrm{tr}\left[\, \mathrm{E}_K\, xx^\tau h(x)\, \mathrm{I}(|x| \le M_1)\, Q\right] \quad (7.2.235)$$

where h is positive and uniformly bounded. By analogous arguments as in the proof of Theorem 7.2.11, we cover a full neighborhood of C_3 if Q varies over the symmetric $k \times k$-matrices of the special form $\alpha \mathbb{I}_k$, where $\alpha \in \mathbb{R}$ with $|\alpha| \le \frac{b - b_1}{M_1^2 \max\{M_2, |\delta|\}}$. Furthermore, we get by ($\mathrm{M}_\zeta 3$)

$$\begin{aligned} &\mathrm{E}\, xx^\tau \frac{\Lambda_f^{\mathrm{loc}}}{q(x,u)} \zeta_Q(x,u) \\ &\quad = C_2 + \mathrm{E}_K\, xx^\tau (x^\tau Q x)\, \mathrm{I}(|x| \le M_1)\, \mathrm{E}_\bullet \frac{|u| (\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)} \big[\, \mathrm{I}_1 - \delta(x)\, \mathrm{I}_2 \,\big] \end{aligned} \quad (7.2.236)$$

where by assumption (7.2.206) and (F2), which is $\mathcal{I}_F^{\mathrm{loc}} < \infty$,

$$0 < h(x) := \mathrm{E}_\bullet \frac{|u| (\Lambda_f^{\mathrm{loc}}(u))^2}{q(x,u)} \big[\, \mathrm{I}_1 - \delta(x)\, \mathrm{I}_2 \,\big] \le \mathrm{E}_\bullet \frac{|u|}{|\gamma|\,|u|} (\Lambda_f^{\mathrm{loc}}(u))^2\, \mathrm{I}_1 \quad (7.2.237)$$

$$\le |\gamma|^{-1}\, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}}(u))^2 = |\gamma|^{-1} \mathcal{I}_f^{\mathrm{loc}} < \infty \quad (7.2.238)$$

Using the vec and vech operators and the same arguments as in the proof of Theorem 7.2.11 we also cover a full neighborhood of C_2. Finally it holds, that

$$\begin{aligned} &|\zeta_Q| - b(x,u) \\ &\quad \le |\zeta_1| + |x^\tau Q x|\, \mathrm{I}(|x| \le M_1) \big|\, |u\Lambda_f^{\mathrm{loc}}(u)|\, \mathrm{I}_1 - \delta(x) |u\Lambda_f^{\mathrm{loc}}(u)|\, \mathrm{I}_2 \,\big| - b(x,u) && (7.2.239) \\ &\quad \le b_1(x,u) - b(x,u) + \|Q\|\, M_1^2 \max\{M_2, |\delta|\} && (7.2.240) \\ &\quad \le b_1 - b + b - b_1 = 0 && (7.2.241) \end{aligned}$$

Therefore, the set $\big\{\zeta_Q \,\big|\, Q \in \mathbb{R}^{k \times k}$ symmetric, $\|Q\|_{\mathrm{op}} \le \frac{b - b_1}{M_1^2 M_2}\big\}$ covers full neighborhoods of C_2 and C_3, and the constraints ($\mathrm{Mc}_\zeta 2$) and (7.2.224) are fulfilled. That is, problem (7.2.222) is well-posed and Theorem B.2.6 of Rieder (1994) is applicable. Thus, there exist some multipliers $\alpha_3 \in \mathbb{R}$ and $B \in \mathbb{R}^{k \times k}$ symmetric, such that the solution $\tilde{\zeta}$ minimizes the Lagrangian

$$\begin{aligned} L(\zeta) &= \mathrm{E}\, \zeta^2 - 2\, \mathrm{E}\, x^\tau B x \tfrac{\Lambda_f^{\mathrm{loc}}}{q} \zeta - 2\alpha_3\, \mathrm{E}\, \tfrac{u^2 \Lambda_f^{\mathrm{loc}}}{q} \zeta && (7.2.242) \\ &= \mathrm{E}\, \mathrm{E}_\bullet \left[\zeta - 2x^\tau B x \tfrac{\Lambda_f^{\mathrm{loc}}}{q} \zeta - 2\alpha_3 \tfrac{u^2 \Lambda_f^{\mathrm{loc}}}{q} \zeta \right] && (7.2.243) \end{aligned}$$

among all $\zeta \in L_2(P_{\theta_0})$ satisfying (7.2.224). This Lagrangian can be minimized by minimization of the conditional expectation inside, at each point x, subject to ($\mathrm{Mc}_\zeta 2$) and

$$\sup_\bullet \left[|\zeta(x, \cdot)| - b(x, \cdot) \right] \le 0 \quad (7.2.244)$$

The conditional problem given any $x \in \mathbb{R}^k$ refers to the space $L_2(F)$ with expectation under F. This problem is well-posed; consider ζ_Q as defined in (7.2.128). Hence, there exists some $\alpha(x) \in \mathbb{R}$ such that the x section of the solution $\tilde{\zeta}$ may tentatively be found by minimization of the conditional Lagrangian

$$L_\bullet(\zeta) = \mathrm{E}_\bullet \Big[\zeta - \frac{\alpha_1 u + x^\tau B x \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q}\Big]^2 + \mathrm{const} \tag{7.2.245}$$

over all $\zeta(x, \cdot\,) \in L_2(F)$ such that (7.2.244) holds. Then, by (7.2.122) the asserted form (7.2.208) of the optimal section $\tilde{\psi}(x, \cdot\,)$ is attained. Since $\tilde{\zeta}(x, .\,)$ is odd in u a.e. $K(dx)$, ($\mathrm{M}_\zeta 1$) and ($\mathrm{M}_\zeta 5$) are automatically fulfilled; confer Remark 7.1.9 (b).

Conversely, if there is some $\tilde{\rho} \in \Psi_{\mathrm{Mc}}$ with $\tilde{\psi}$ of this form (7.2.208). Then, (K almost) all x sections minimize the corresponding conditional Lagrangian $L_\bullet$ over $L_2(F)$ subject to (7.2.244), hence

$$\mathrm{E}_\bullet \Big[\zeta - 2x^\tau B x \tfrac{\Lambda_f^{\mathrm{loc}}}{q}\zeta - 2\alpha_3 \tfrac{u^2 \Lambda_f^{\mathrm{loc}}}{q}\zeta\Big] \tag{7.2.246}$$

among all $\zeta \in L_2(P)$ satisfying ($\mathrm{Mc}_\zeta 2$) and (7.2.244). On integration with respect to $K(dx)$, then $\tilde{\zeta}$ minimizes the Lagrangian L in (7.2.243) over $L_2(P)$ subject to ($\mathrm{Mc}_\zeta 2$) and (7.2.224). Therefore $\tilde{\zeta}$ solves (7.2.222).
To get the representations (7.2.213)–(7.2.215), proceed as in the proof of Theorem 7.2.11.

(b) The problem (A.2.5) can equivalently be rewritten as

$$\mathrm{E}\,\zeta^2 + r^2 \sup_P |\rho|^2 + \mathrm{const} = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta} \tag{7.2.247}$$

Putting $b := \sup_P |\tilde{\rho}|$, hence $\tilde{\zeta}$ also solves the problem

$$\mathrm{E}\,\zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{M}_\zeta},\ \sup_P \big[|\zeta(x,u)| - b(x,u)\big] \le 0 \tag{7.2.248}$$

which has already occurred as problem (7.2.222) and the solution $\tilde{\zeta}$, respectively $\tilde{\psi}$ is of form (7.2.208). Plugged into the Lagrangian L, we get a convex function of the bound b

$$L_1(b) = \mathrm{E}\,\big(|g(x,u)| - b(x,u)\big)_+^2 + r^2 b^2 \tag{7.2.249}$$

Analogously to Lemma C.2.3 of Rieder (1994), we may differentiate L_1 by b, which yields (7.2.220), respectively (7.2.221). ////

Remark 7.2.16 (a) Similarly to the case of AL estimators (cf. Remark 7.2.7 (a)), the (product) measurability of $\alpha(x)$ is still an open problem.

(b) Starting from $\tilde{\psi}_{\theta_0}$ and defining $\tilde{\psi}_\theta$ by (7.1.54) and fixing the same regular and symmetric $A \in \mathbb{R}^{k\times k}$ and the same $\gamma \in \mathbb{R}$, we get that this $\tilde{\psi}_\theta$ is also of form (7.2.208) with

$$\alpha_{1,\theta}(x) = \sigma^2 \alpha_{1,\theta_0}(x) \qquad \alpha_{3,\theta} = \sigma^2 \alpha_{3,\theta_0} \qquad B_\theta = \sigma^2 B_{\theta_0} \qquad b_\theta = \sigma b_{\theta_0} \tag{7.2.250}$$

Thus, $\tilde{\psi}_\theta$ is the solution to the corresponding Hampel type problem (A.2.1) with bound $b_\theta = \sigma b_{\theta_0}$. ////

Elliptically Symmetric Regressor Distribution

We now specialize the solutions given in Theorem 7.2.15 by additionally assuming that the ideal regressor distribution K is elliptically symmetric; i.e., (7.2.19) holds. As shown in the preceding subsubsection, we only need to optimize over all regular diagonal matrices $A_G \in \mathbb{R}^{k\times k}$ and then obtain the optimal matrix A by $A = G^\tau A_G G$.

Proposition 7.2.17 *Assume (F1), (7.2.206) and (7.2.19) and let* $\mathrm{E}_{\mathcal{L}(V)}\,|v|^4 < \infty$ *and* E *be expectation taken under* $\tilde{P}_{\theta_0}(du, dv) := f(u)\lambda(du)\,\mathcal{L}(V)(dv)$.

(a) *In case of* $A_G = \mathrm{diag}(a_1, \ldots, a_k) \in \mathbb{R}^{k\times k}$ *regular,* $\gamma \in \mathbb{R}\setminus\{0\}$ *and* $b \in [\gamma, \infty)$ *fixed, the unique solution of Problem (A.2.1) is based via (7.2.91) on* $\tilde{\psi}$ *of the form (7.2.208), where* $A = G^\tau A_G G$, $B = G^\tau B_G G$ *with* $B_G = \mathrm{diag}(b_1, \ldots, b_k)$ *and the following representations hold*

$$\alpha_1(x) = \Big[C_1(x) - \mathrm{E}_\bullet \tfrac{v^\tau B_G v u \Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q_G^2} w_G\Big] \Big/ \mathrm{E}_\bullet \tfrac{u^2}{q_G^2} w_G \tag{7.2.251}$$

$$b_j = \frac{C_{2,j} - \mathrm{E}\,\frac{\alpha_1(x) u \Lambda_f^{\mathrm{loc}} + \sum_{i=1, i\neq j}^k b_i v_i^2 v_j^2 (\Lambda_f^{\mathrm{loc}})^2 + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G}{\mathrm{E}\,\frac{v_j^4 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G} \tag{7.2.252}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\,\tfrac{\alpha_1(x) u^3 \Lambda_f^{\mathrm{loc}} + v^\tau B_G v u^2 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G\Big] \Big/ \mathrm{E}\,\tfrac{u^4 (\Lambda_f^{\mathrm{loc}})^2}{q_G^2} w_G \tag{7.2.253}$$

where

$$q_G(v,u) = \big[|A_G v|^2 + \gamma^2 u^2\big]^{1/2} \qquad w_G(v,u) = \min\Big\{1, \frac{b_G(v,u)}{|g_G(v,u)|}\Big\} \tag{7.2.254}$$

$$b_G(v,u) = \Big[b^2 - \gamma^2 \tfrac{|A_G v|^2}{q_G(v,u)^2}\Big]^{1/2} \qquad g_G(v,u) = \tfrac{\alpha_1(x) u + v^\tau B_G v \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q_G(v,u)} \tag{7.2.255}$$

and

$$C_1(x) = \mathrm{E}_\bullet \tfrac{|A_G v|^2}{q_G^2} \tag{7.2.256}$$

$$C_{2,j} = a_j^{-1} - \gamma^2\,\mathrm{E}\, v_j^2 \tfrac{u \Lambda_f^{\mathrm{loc}}}{q_G^2} \qquad j = 1, \ldots, k \tag{7.2.257}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2\,\mathrm{E}\,\tfrac{u^3 \Lambda_f^{\mathrm{loc}}}{q_G^2} \tag{7.2.258}$$

(b) *The unique solution to the corresponding MSE problem (A.2.5) coincide with the solution of problem (A.2.1), where bound* $b \in [\gamma, \infty)$ *and radius* $r \in (0, \infty)$ *are related via*

$$r^2 = \mathrm{E}\,\Big(\frac{|g_G(v,u)|}{b_G(v,u)} - 1\Big)_+ \qquad \text{if } b > \gamma \tag{7.2.259}$$

respectively

$$r^2 \geq \mathrm{E}\,\Big(\frac{|g_G(v,u)|}{b_G(v,u)} - 1\Big)_+ \qquad \text{if } b = \gamma \tag{7.2.260}$$

Proof Follows by analogous arguments as in the proofs of Proposition 7.2.8 and Theorem 7.2.15. ////

Spherically Symmetric Regressor Distribution

We now specialize the solutions given in Theorem 7.2.15 by additionally assuming, that the ideal regressor distribution K is spherically symmetric. As shown in the preceding subsubsection, we only need to optimize over all constants $a \in \mathbb{R} \setminus \{0\}$ in this case and then obtain the optimal matrix A via $A = a\mathbb{I}_k$.

Proposition 7.2.18 Assume (F1), (7.2.206) and spherical symmetry of the ideal regressor distribution K, where $\mathrm{E}_K |x|^4 < \infty$.

(a) In case of $a \in \mathbb{R} \setminus \{0\}$, $\gamma \in \mathbb{R} \setminus \{0\}$ and $b \in [\gamma, \infty)$ fixed, the unique solution of Problem (A.2.1) is based via (7.2.91) on $\tilde{\psi}$ of the form (7.2.208), where $A = a\mathbb{I}_k$, $B = d\mathbb{I}_k$ with $d \in \mathbb{R}$ and the following representations hold

$$\alpha_1(x) = \Big[C_1(x) - \mathrm{E}_\bullet \tfrac{d|x|^2 u\Lambda_f^{\mathrm{loc}} + \alpha_3 u^3 \Lambda_f^{\mathrm{loc}}}{q^2} w\Big] \Big/ \mathrm{E}_\bullet \tfrac{u^2}{q^2} w \tag{7.2.261}$$

$$d = \Big[C_2 - \mathrm{E}\, \tfrac{\alpha_1(x) u\Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\, \tfrac{x_1^2 |x|^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.262}$$

$$\alpha_3 = \Big[C_3 - \mathrm{E}\, \tfrac{\alpha_1(x) u^3 \Lambda_f^{\mathrm{loc}} + d|x|^2 u^2 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w\Big] \Big/ \mathrm{E}\, \tfrac{u^4 (\Lambda_f^{\mathrm{loc}})^2}{q^2} w \tag{7.2.263}$$

where

$$q(x,u) = \big[a^2|x|^2 + \gamma^2 u^2\big]^{1/2} \qquad w(x,u) = \min\Big\{1, \frac{b(x,u)}{|g(x,u)|}\Big\} \tag{7.2.264}$$

$$b(x,u) = \Big[b^2 - \gamma^2 \tfrac{a^2|x|^2}{q(x,u)^2}\Big]^{1/2} \qquad g(x,u) = \tfrac{\alpha_1(x) u + d|x|^2 \Lambda_f^{\mathrm{loc}} + \alpha_3 u^2 \Lambda_f^{\mathrm{loc}}}{q(x,u)} \tag{7.2.265}$$

and

$$C_1(x) = \mathrm{E}_\bullet \tfrac{a^2|x|^2}{q^2} \qquad C_2 = a^{-1} - \gamma^2\, \mathrm{E}\, x_1^2 \tfrac{u\Lambda_f^{\mathrm{loc}}}{q^2} \tag{7.2.266}$$

$$C_3 = 1 + \gamma^{-1} - \gamma^2\, \mathrm{E}\, \tfrac{u^3 \Lambda_f^{\mathrm{loc}}}{q^2} \tag{7.2.267}$$

(b) The unique solution to the corresponding MSE problem (A.2.5) coincide with the solution of problem (A.2.1), where bound $b \in [\gamma, \infty)$ and radius $r \in (0, \infty)$ are related via (7.2.220), respectively (7.2.221). ////

Proof Analogously to the proofs of Proposition 7.2.9 and Theorem 7.2.15. ////

Error Distribution $F = \mathcal{N}(0,1)$

In case of ideal error distribution $F = \mathcal{N}(0,1)$ the conditions (Mc2) and (M3) coincide and $A_\psi^{-1} = \mathcal{K}$; confer Remark 7.1.9 (c).

Theorem 7.2.19 *Assume $F = \mathcal{N}(0,1)$.*

(a) *In case of $\gamma \in \mathbb{R} \setminus \{0\}$ and $b \in [\gamma, \infty)$ fixed, there exist some $\alpha_1 \colon \mathbb{R}^k \to \mathbb{R}$ and $\alpha_3 \in \mathbb{R}$ such that the solution $\tilde{\rho}$ to problem (A.2.1) is based via (7.2.91) on $\tilde{\psi}$ of the form*

$$\tilde{\psi}(x,u) = \frac{\alpha_1(x) u + \alpha_3 u^3}{q^2(x,u)} w(x,u) + \gamma^2 \frac{u}{q^2(x,u)} \tag{7.2.268}$$

with

$$w(x,u) = \min\left\{1, \frac{b(x,u)}{|g(x,u)|}\right\} \tag{7.2.269}$$

where

$$q(x,u) = \sqrt{|\mathcal{K}^{-1}x|^2 + \gamma^2 u^2} \qquad g(x,u) = \frac{\alpha_1(x)u + \alpha_3 u^3}{q(x,u)} \tag{7.2.270}$$

$$b(x,u) = \left[b^2 - \gamma^2 \frac{|\mathcal{K}^{-1}x|^2}{q^2(x,u)}\right]^{1/2} \tag{7.2.271}$$

Conversely, if there is some $\tilde{\rho} \in \Psi_{\mathrm{Mc}}$ *with* $\tilde{\psi}$ *of form (7.2.268) for any* $\gamma \in \mathbb{R} \setminus \{0\}$ *and* $b \in [\gamma, \infty)$ *fixed,* $\alpha_1 \colon \mathbb{R}^k \to \mathbb{R}$, $\alpha_3 \in \mathbb{R}$, *then* $\tilde{\rho}$ *is the solution, and the following representations hold,*

$$\alpha_1(x) = \left[C_1(x) - \mathrm{E}_\bullet \frac{\alpha_3 u^3}{q^2} w\right] \Big/ \mathrm{E}_\bullet \frac{u^2}{q^2} w \tag{7.2.272}$$

$$\alpha_3 = \left[C_3 - \mathrm{E}\, \frac{\alpha_1(x) u^3}{q^2} w\right] \Big/ \mathrm{E}\, \frac{u^6}{q^2} w \tag{7.2.273}$$

where

$$C_1(x) = \mathrm{E}_\bullet \frac{|\mathcal{K}^{-1}x|^2}{q^2} \in \mathbb{R} \qquad C_3 = 1 + \gamma^{-1} - \gamma^2\, \mathrm{E}\, \frac{u^4}{q^2} \in \mathbb{R} \tag{7.2.274}$$

(b) *The unique solution to the corresponding MSE problem (A.2.5) coincide with the solution of problem (A.2.1), where bound* $b \in [\gamma, \infty)$ *and radius* $r \in (0, \infty)$ *are related via*

$$r^2 = \mathrm{E}\left(\frac{|g(x,u)|}{b(x,u)} - 1\right)_+ \qquad \text{if } b > \gamma \tag{7.2.275}$$

respectively

$$r^2 \geq \mathrm{E}\left(\frac{|g(x,u)|}{b(x,u)} - 1\right)_+ \qquad \text{if } b = \gamma \tag{7.2.276}$$

(c) *Given* $\gamma \in \mathbb{R} \setminus \{0\}$ *let* $\bar{b}(\gamma) \in [\gamma, \infty)$ *be minimal such that*

$$C_1(x) \leq \mathrm{E}_\bullet \tfrac{|u|}{q(x,u)} \bar{b}(x,u) \qquad C_3 \leq \mathrm{E}\, \tfrac{|u|^3}{q(x,u)} \bar{b}(x,u) \tag{7.2.277}$$

where

$$\bar{b}(x,u) = \left[\bar{b}(\gamma)^2 - \gamma^2 \tfrac{|\mathcal{K}^{-1}x|^2}{q^2(x,u)}\right]^{1/2} \tag{7.2.278}$$

Then, $\omega_{\mathrm{Mc}}^{\min} \geq \bar{b}(\gamma)$. *If there is equality in (7.2.277) for* $\bar{b}(\bar{\gamma}) = \min_\gamma \bar{b}(\gamma)$ *then* $\bar{\zeta}$ *of form*

$$\bar{\zeta}(x,u) = b_{\min}(x,u)\, \mathrm{sign}\,(u) \tag{7.2.279}$$

$$b_{\min}(x,u) = \left[\left(\omega_{\mathrm{Mc}}^{\min}\right)^2 - \bar{\gamma}^2 \tfrac{|\mathcal{K}^{-1}x|^2}{|\mathcal{K}^{-1}x|^2 + \bar{\gamma}^2 u^2}\right]^{1/2} \tag{7.2.280}$$

is the solution to problem (7.2.15), $\omega_{\mathrm{Mc}}^{\min} = \bar{b}(\bar{\gamma})$, *and the corresponding* $\bar{\rho}$ *attains this minimum bias.*

PROOF

(a) The proof is very similar to the proof of Theorem 7.2.15. As already shown, Problem (A.2.1) is equivalent to

$$\mathrm{E}\,\zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{Mc}_\zeta},\ \sup_P \left[|\zeta(x,u)| - b(x,u)\right] \le 0 \tag{7.2.281}$$

where

$$\Psi_{\mathrm{Mc}_\zeta} = \left\{\zeta \in L_2(P) \,\middle|\, \zeta \text{ fulfills } (\mathrm{Mc}_\zeta 1), (\mathrm{Mc}_\zeta 2), (\mathrm{M}_\zeta 4), (\mathrm{M}_\zeta 5)\right\} \tag{7.2.282}$$

We will solve this convex problem under $(\mathrm{Mc}_\zeta\, 2)$ and $(\mathrm{M}_\zeta\, 4)$ and then $(\mathrm{Mc}_\zeta\, 1)$ and $(\mathrm{M}_\zeta\, 5)$ are automatically fulfilled.
Since expectation and scalar products are linear weakly continuous and since also $\sup_{P_{\theta_0}} \left[|\zeta(x,u)| - b(x,u)\right] \le 0$ is convex and weakly lower semicontinuous (cf. Remark A.2.1 (b)), the problem (7.2.281) is a minimum norm problem over a convex closed subset in Hilbert space. This problem has a unique solution $\tilde{\zeta}$ provided there exists some $\zeta \in \Psi_{\mathrm{Mc}_\zeta}$ that satisfies

$$\sup_P \left[|\zeta(x,u)| - b(x,u)\right] \le 0 \tag{7.2.283}$$

In general, there is no Lagrange multiplier for the side condition $(\mathrm{Mc}_\zeta\, 2)$; confer Rieder (1994) (p 276). Thus, we first keep $(\mathrm{Mc}_\zeta\, 2)$ and (7.2.283) back in the domain and show that the resulting optimization problem is well-posed in sense of Definition B.2.9 in Rieder (1994). To verify this, we assume some ζ_1 with $(\mathrm{Mc}_\zeta\, 2)$, $(\mathrm{M}_\zeta\, 4)$ and

$$\sup_P \left[|\zeta_1(x,u)| - b_1(x,u)\right] \le 0 \qquad b_1(x,u) := \left[b_1^2 - \gamma^2 \frac{|\mathcal{K}^{-1}x|^2}{q^2(x,u)}\right]^{1/2} \tag{7.2.284}$$

where $\omega_{\mathrm{Mc}}^{\min} < b_1 < b$ and define the bounded function

$$\Delta(x,u) := \mathrm{I}(0 < u < 1) - \delta(x)\,\mathrm{I}(1 < u < 2) \tag{7.2.285}$$

where

$$\delta(x) := \mathrm{E}_\bullet \frac{u\,\mathrm{I}(0 < u < 1)}{q(x,u)} \left[\mathrm{E}_\bullet \frac{u\,\mathrm{I}(1 < u < 2)}{q(x,u)}\right]^{-1} \tag{7.2.286}$$

that is, $\mathrm{E}_\bullet \frac{u}{q}\Delta = 0$ for all $x \in \mathbb{R}^k$. Then, passing over to $\zeta_\epsilon := \zeta_1 + \epsilon\Delta$ with $|\epsilon| \le \epsilon_0$ and $\epsilon_0|\Delta(x,u)| \le b - b_1$ we cover a full neighborhood of C_3, the constraint $(\mathrm{Mc}_\zeta\, 2)$ holds and

$$\sup_P \left[|\zeta_\epsilon(x,u)| - b(x,u)\right] \le b_1 - b + \epsilon_0 \|\Delta(x,u)\| \le 0 \tag{7.2.287}$$

Therefore, Theorem B.2.6 of Rieder (1994) is applicable and there exist some multiplier $\alpha_3 \in \mathbb{R}$ such that the solution $\tilde{\zeta}$ minimizes the Lagrangian

$$L(\zeta) = \mathrm{E}\,\zeta^2 - 2\alpha_3\,\mathrm{E}\,\frac{u^3}{q}\zeta = \mathrm{E}\,\mathrm{E}_\bullet \left[\zeta^2 - 2\alpha_3 \frac{u^3}{q}\right] \tag{7.2.288}$$

among all $\zeta \in L_2(P)$ satisfying $(\mathrm{Mc}_\zeta\, 2)$ and (7.2.283). This Lagrangian can be minimized by minimization of the conditional expectation inside, at each point x, subject to $(\mathrm{Mc}_\zeta\, 2)$ and

$$\sup_\bullet \left[|\zeta(x,\cdot)| - b(x,\cdot)\right] \le 0 \tag{7.2.289}$$

The conditional problem given any $x \in \mathbb{R}^k$ is also well-posed as $b > \omega_{\mathrm{M}}^{\min}$ and we may verify the well-posedness assumptions with $\zeta_\epsilon := \zeta_1 + \epsilon \operatorname{sign}(u)$ and $|\epsilon| \le b - b_1$. Hence, there exist some $\alpha_1(x) \in \mathbb{R}$ such that the x section of the solution $\tilde{\zeta}$ may tentatively be found by minimization of the conditional Lagrangian

$$L_\bullet(\zeta) = \mathrm{E}_\bullet \left[\zeta^2 - 2\alpha_1(x) \frac{u}{q} \zeta - 2\alpha_3 \frac{u^3}{q} \zeta \right] \tag{7.2.290}$$

$$= \mathrm{E}_\bullet \left[\zeta - \frac{\alpha_1(x) u + \alpha_3 u^3}{q} \right]^2 + \text{const} \tag{7.2.291}$$

over all $\zeta(x, \cdot) \in L_2(\mathcal{N}(0,1))$ subject to (7.2.289). Pointwise minimization of integrands together with (7.2.122) implies the asserted form (7.2.268) of the optimal section $\tilde{\psi}(x, \cdot)$ and $\tilde{\psi}(x, \cdot)$ is odd in u a.e. $K(dx)$.

Conversely, if there is some $\tilde{\rho} \in \Psi_{\mathrm{Mc}_\zeta}$ with $\tilde{\psi}$ of form (7.2.268) for any measurable function $\alpha_1(x)$ from $\mathbb{R}^k$ to $\mathbb{R}$ and any $\alpha_3 \in \mathbb{R}$. Then, (K almost) all x sections minimize the corresponding conditional Lagrangian $L_\bullet$ among all $\zeta(x, \cdot) \in L_2(\mathcal{N}(0,1))$ subject to (7.2.289), hence $\mathrm{E}_\bullet \zeta^2 - 2\alpha_3 \mathrm{E}_\bullet \frac{u^3}{q}$ over $\zeta \in L_2(P)$ subject to (Mc_ζ 2) and (7.2.289). On integration with respect to $K(dx)$, $\tilde{\zeta}$ minimizes the Lagrangian L in (7.2.288) over $\zeta \in L_2(P_{\vartheta_0})$ subject to (Mc_ζ 2) and (7.2.283) and therefore $\tilde{\zeta}$ solves (7.2.281). To prove (7.2.272) and (7.2.273), we solve (Mc_ζ 2) by $\alpha_1(x)$ and (M_ζ 4) by α_3, respectively.

(b) The problem (A.2.5) can equivalently be rewritten as

$$\mathrm{E}\, \zeta^2 + r^2 \sup_P |\rho|^2 + \text{const} = \min! \qquad \zeta \in \Psi_{\mathrm{Mc}_\zeta} \tag{7.2.292}$$

Putting $b := \sup_P |\tilde{\rho}|$, hence $\tilde{\zeta}$ also solves the problem

$$\mathrm{E}\, \zeta^2 = \min! \qquad \zeta \in \Psi_{\mathrm{Mc}_\zeta},\ \sup_P \left[|\zeta(x,u)| - b(x,u) \right] \le 0 \tag{7.2.293}$$

which has already occurred as problem (7.2.281) and the solution $\tilde{\zeta}$, respectively $\tilde{\psi}$ is of form (7.2.268). Plugged into the Lagrangian L, which is $\mathrm{E}_K L_\bullet$ with $L_\bullet$ from (7.2.291), we get a convex function of the bounds

$$L_1(b) = \mathrm{E} \left(|g(x,u)| - b(x,u) \right)_+^2 + r^2 b^2 \tag{7.2.294}$$

Analogously to Lemma C.2.3 of Rieder (1994), we may differentiate L_1 by b, which yields (7.2.275) and (7.2.276), respectively.

(c) For arbitrary measurable $\alpha_1(x) \ge 0$ a.e. $K(dx)$, $a_3 \ge 0$ and $\zeta \in \Psi_{\mathrm{Mc}_\zeta}$ it holds

$$\mathrm{E}\, \alpha_1(x) C_1(x) + \alpha_3 C_3 = \mathrm{E}\, \frac{\alpha_1(x) u + \alpha_3 u^3}{q(x,u)} \zeta \tag{7.2.295}$$

$$\le \mathrm{E}\, \frac{\alpha_1(x)|u| + \alpha_3 |u|^3}{q(x,u)} |\zeta| \tag{7.2.296}$$

$$\le \mathrm{E}\, \frac{\alpha_1(x)|u| + \alpha_3 |u|^3}{q(x,u)} b(x,u) \tag{7.2.297}$$

$$= \mathrm{E}\, \alpha_1(x)\, \mathrm{E}_\bullet \tfrac{|u|}{q(x,u)} b(x,u) + \alpha_3\, \mathrm{E}\, \tfrac{|u|^3}{q(x,u)} b(x,u) \tag{7.2.298}$$

where $b(x,u) = \left[b^2 + \gamma^2 \frac{|\tilde{x}|^2}{q^2(x,u)}\right]^{1/2}$. Thus we obtain the following lower bound for b

$$1 \leq \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u) + \alpha_3\,\mathrm{E}\,\frac{|u|^3}{q(x,u)} b(x,u)}{\mathrm{E}\,\alpha_1(x) C_1(x) + \alpha_3 C_3} =: \frac{d_1 + d_2\alpha_3}{d_3 + d_4\alpha_3} \tag{7.2.299}$$

Given $\alpha_1(\cdot) \geq 0$, the minimization of the right hand side via differentiation in α_3 yields that the minimum is attained for $\alpha_3 = 0$, if $d_2 d_3 > d_1 d_4$, respectively, for $\alpha_3 = \infty$, if $d_2 d_3 < d_1 d_4$. In case of $\alpha_3 = \infty$ (7.2.299) yields the necessary the second lower bound in (7.2.277) for b. If $\alpha_3 = 0$, we get

$$1 \leq \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u)}{\mathrm{E}\,\alpha_1(x) C_1(x)} \tag{7.2.300}$$

We now minimize the right hand side of (7.2.300) in $\alpha_1(\cdot) \geq 0$. If $\mathrm{E}_\bullet \frac{|u|}{q} b \geq m C_1(x)$ a.e. $K(dx)$, it follows

$$\min_{\alpha_1(\cdot)\geq 0} \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u)}{\mathrm{E}\,\alpha_1(x) C_1(x)} \geq \inf_{K(dx)} \frac{\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u)}{C_1(x)} =: m_0 \tag{7.2.301}$$

On the other hand, if $\alpha_1(\cdot)$ is concentrated on points x with

$$\mathrm{E}_\bullet \tfrac{|u|}{q} b \leq (m_0 + \delta) C_1(x) \tag{7.2.302}$$

for some $\delta > 0$, we have

$$\frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u)}{\mathrm{E}\,\alpha_1(x) C_1(x)} \leq m_0 + \delta \tag{7.2.303}$$

Hence,

$$\frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet \frac{|u|}{q(x,u)} b(x,u)}{\mathrm{E}\,\alpha_1(x) C_1(x)} = m_0 \tag{7.2.304}$$

holds, where $m_0 \geq 1$ by (7.2.299). This yields the first lower bound in (7.2.277) for b.

We then determine $\bar{b}(\bar{\gamma})$. By (7.2.277) necessarily, $\omega_{\mathrm{Mc}}^{\min} \geq \bar{b}(\bar{\gamma})$ since $(\mathrm{Mc}_\zeta 2)$ and $(\mathrm{M}_\zeta 4)$ must hold for every $\zeta \in \Psi_{\mathrm{Mc}_\zeta}$. However, if we get equality in (7.2.277) then necessarily, $\omega_{\mathrm{Mc}}^{\min} = \bar{b}(\bar{\gamma})$, since $\bar{\zeta}(x,u)$ as defined in (7.2.279) fulfills $(\mathrm{Mc}_\zeta 1)$, $(\mathrm{Mc}_\zeta 2)$, $(\mathrm{M}_\zeta 4)$ and $(\mathrm{M}_\zeta 5)$. That is, $\bar{\zeta}$ is the unique solution to problem (7.2.159) and the corresponding $\bar{\rho}$ achieves the minimum bias $\omega_{\mathrm{Mc}}^{\min}$. ////

Remark 7.2.20 Since condition (Mc2) and (M3) coincide, the solution in case of elliptically, respectively spherically symmetric ideal regressor distribution K is identical to the solution given in Theorem 7.2.19. ////

7.3 Separate Estimation

We consider the separate estimation of regression and scale. This may be interpreted as follows: Today we want to estimate the regression parameter and consider scale as nuisance parameter, whereas tomorrow we are interested in the estimation of scale and the regression parameter is regarded as nuisance. Doing so only contamination may change.

To lighten the notation, we drop the fixed parameter θ.

7.3.1 AL estimators

We have to solve the following optimization problem

$$\mathrm{E}\,|\eta_{\mathrm{rg}}|^2 + \mathrm{E}\,|\eta_{\mathrm{sc}}|^2 + r^2\big(\omega_{c,t}^2(\eta_{\mathrm{rg}}) + \omega_{c,t}^2(\eta_{\mathrm{sc}})\big) = \min! \tag{7.3.1}$$

with $t = 0$, $t = \alpha = 1$ under the side conditions

$$\mathrm{E}\,\eta_{\mathrm{rg}} = 0, \qquad \mathrm{E}\,\eta_{\mathrm{rg}}\Lambda_f^{\mathrm{loc}} x^\tau = \mathbb{I}_k, \qquad \mathrm{E}\,\eta_{\mathrm{rg}}\Lambda_f^{\mathrm{sc}} = 0 \tag{7.3.2}$$

and

$$\mathrm{E}\,\eta_{\mathrm{sc}} = 0, \qquad \mathrm{E}\,\eta_{\mathrm{sc}}\Lambda_f^{\mathrm{sc}} = 1, \qquad \mathrm{E}\,\eta_{\mathrm{sc}}\Lambda_f^{\mathrm{loc}} x^\tau = 0 \tag{7.3.3}$$

Obviously, this MSE problem consists of the sum of the individual MSE problems of the regression and the scale model. The solutions to these individual problems are given in Subsection 9.2.1.

Remark 7.3.1 **(a)** The setup corresponds to the MSE problems with coordinatewise oscillations terms considered in Rieder (1994) (index $s = 2$) where this is called *one at a time optimaliy* (p 197, ibid.).

(b) To distinguish between the different AL estimators, we call the optimal AL estimators for the separate estimation of regression and scale ALs estimators. ////

Proposition 7.3.2 *Assume (F1). Then, in case of unconditional ($* = c$, $t = 0$) as well as average conditional ($* = c$, $t = \alpha = 1$) contamination neighborhoods, the unique solutions $\tilde\eta_{\mathrm{rg}}$ and $\tilde\eta_{\mathrm{sc}}$ to problem (7.3.1) are*

$$\tilde\eta_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u) \min\Big\{1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u)|}\Big\} \tag{7.3.4}$$

with

$$A_{\mathrm{rg}} = \Big[\mathrm{E}(\Lambda_f^{\mathrm{loc}})^2 x x^\tau \min\Big\{1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}|}\Big\}\Big]^{-1} \tag{7.3.5}$$

and

$$\tilde\eta_{\mathrm{sc}}(u) = A_{\mathrm{sc}}(\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}}) w(u) \qquad w(u) = \min\Big\{1, \frac{c_{\mathrm{sc}}}{|\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}}|}\Big\} \tag{7.3.6}$$

with

$$z_{\mathrm{sc}} = \mathrm{E}\,\Lambda_f^{\mathrm{sc}} w_{\mathrm{sc}} \big[\mathrm{E}\, w_{\mathrm{sc}}\big]^{-1} \qquad A_{\mathrm{sc}} = \big[\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 w_{\mathrm{sc}}\big]^{-1} \tag{7.3.7}$$

where the optimal clipping bounds $b_{\mathrm{rg}}, c_{\mathrm{sc}} \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *are related via*

$$r^2 b_{\mathrm{rg}} = \mathrm{E}\left(|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}| - b_{\mathrm{rg}}\right)_+ \tag{7.3.8}$$

and

$$r^2 c_{\mathrm{sc}} = \mathrm{E}\left(|\Lambda_f^{\mathrm{sc}}| - c_{\mathrm{sc}}\right)_+ \tag{7.3.9}$$

respectively.

PROOF As a consequence of robust adaptivity (cf. equations (9.2.9) and (9.2.14)) problem (7.3.1) in case of unconditional contamination neighborhoods ($* = c$, $t = 0$) is solved by the solutions to the corresponding separate problems specified in Subsubsection 9.2.1.1. Moreover, by the arguments given in Remarks 9.2.10 and 9.2.11, the solutions $\tilde{\eta}_{\mathrm{rg}}$ and $\tilde{\eta}_{\mathrm{sc}}$ in case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) coincide with the corresponding solutions in case of unconditional contamination neighborhoods ($* = c$, $t = 0$). ////

7.3.2 M estimators

As in Bednarski and Müller (2001), we restrict our considerations to ideal error distribution $F = \mathcal{N}(0,1)$. For given $\gamma \in \mathbb{R} \setminus \{0\}$ we have to solve the following Hampel type problem

$$\mathrm{E}\,|\rho|^2 = \min! \qquad \rho \in \Psi_{\mathrm{Ms}},\ \omega_{\mathrm{Ms},2}(\rho) \le b_{\mathrm{rg,sc}} \tag{7.3.10}$$

where ρ is of form (7.2.91) with $A^{-1} = \mathcal{K}$ by $F = \mathcal{N}(0,1)$ and (Mc2), Ψ_{Ms} is defined in (7.1.53) and $\omega_{\mathrm{Ms},2}(\rho) \le b_{\mathrm{rg,sc}}$ stands for

$$\sup_P |\mathcal{K}^{-1} x \psi(x,u)| \le b_{\mathrm{rg}} \qquad \sup_P |u\psi(x,u) - 1| \le b_{\mathrm{sc},\gamma} := b_{\mathrm{sc}}/\gamma \tag{7.3.11}$$

which under (Mc15), (Mc2) and for $b_{\mathrm{sc},\gamma} \ge 1$ is equivalent to

$$|\psi(x,u)| \le b_{\mathrm{rg,sc}}(x,u) := \min\left\{\frac{b_{\mathrm{rg}}}{|\mathcal{K}^{-1}x|}, \frac{1 + b_{\mathrm{sc},\gamma}}{|u|}\right\} \qquad \forall\, x,\ u \tag{7.3.12}$$

Moreover, given $\gamma \in \mathbb{R} \setminus \{0\}$ we solve the corresponding MSE problem

$$\mathrm{E}\,|\rho|^2 + r^2 \omega_{\mathrm{Ms},2}^2(\rho) = \min! \qquad \rho \in \Psi_{\mathrm{Ms}}(\theta_0) \tag{7.3.13}$$

over a conditional contamination neighborhood of (starting) radius $r \in [0,\infty)$. Finally, given $\gamma \in \mathbb{R} \setminus \{0\}$ we derive sufficient conditions for the minimum bias $\omega_{\mathrm{Ms},2}^{\min}$ and the solution $\bar{\rho}$ of the convex problem

$$\sup_P |\rho|^2 = \min! \qquad \rho \in \Psi_{\mathrm{Ms}}(\theta_0) \tag{7.3.14}$$

In a second step we then have to minimize the results in $\gamma \in \mathbb{R} \setminus \{0\}$, to solve these problems for ρ of form (7.1.47). As for the M estimators considered in Subsection 7.2.2, this is done via numerical minimization.

Remark 7.3.3 **(a)** Under symmetry of the ideal error distribution F condition (Mc15) will be fulfilled automatically. Hence, by Remark 7.1.9 (b), one could also use (Mc1) and (M5) instead of (Mc15). Moreover, as we consider only ideal error distribution $F = \mathcal{N}(0,1)$ condition (Mc2) entails condition (M3). That is, in this special case, the solutions to the above optimization problems in case of Ms estimators are identical to the solution in case of the more general Mc estimators.

(b) Since the BM estimators of Bednarski and Müller (2001) may be regarded as special M estimators with conditionally centered ICs, one might want to know how much efficiency they lose with respect to the optimal M estimators with conditionally centered ICs. Due to part (a) of this remark, this can achieved by a comparison with the optimal Ms estimators. As our main interest is the simultaneous estimation of regression and scale, we do not derive more general M solutions for the separate estimation of regression and scale.

(c) Robust adaptivity as defined Section 9.1 in case of M estimators is equivalent to (M5) which is automatically fulfilled by (F1) and (Mc15). However, the coupling of the components of ICs of form (7.2.91) prevents the decomposition of problem (7.3.13) into two separate problems as in case of the AL estimators; confer Subsection 7.3.1. ////

Theorem 7.3.4 *Assume ideal error distribution* $F = \mathcal{N}(0,1)$.

(a) *Let* b_{rg}, $b_{\mathrm{sc},\gamma}$ *be such that there exists some* ψ *with (Mc15), (Mc2), (M4) and (7.3.11). Then, for* $\gamma \in \mathbb{R} \setminus \{0\}$ *fixed, there exist some function* $\alpha_1 \colon \mathbb{R}^k \to \mathbb{R}$ *and some* $\alpha_3 \in \mathbb{R}$ *such that the solution is of the form*

$$\tilde{\psi}(x,u) = g(x,u) \min\Big\{1, \frac{b_{\mathrm{rg,sc}}(x,u)}{|g(x,u)|}\Big\} \tag{7.3.15}$$

where

$$g(x,u) = \frac{\alpha_1(x)u + \alpha_3 u^3}{q^2(x,u)} \tag{7.3.16}$$

$$q(x,u) = \sqrt{|\mathcal{K}^{-1}x|^2 + \gamma^2 u^2} \tag{7.3.17}$$

Conversely, if there is some IC $\tilde{\rho} \in \Psi_{\mathrm{Ms}}$ *with* $\tilde{\psi}$ *of form (7.3.15) for any* $b_{\mathrm{rg}} \in (0,\infty)$, $b_{\mathrm{sc},\gamma} \in (1,\infty)$, $\gamma \in \mathbb{R} \setminus \{0\}$ *fixed,* $\alpha_1 \colon \mathbb{R}^k \to \mathbb{R}$ *and* $\alpha_3 \in \mathbb{R}$, *then* $\tilde{\rho}$ *is the solution, and the following representations hold*

$$\alpha_1(x) = \big[1 - \alpha_3 \,\mathrm{E}_\bullet \tfrac{u^4}{q^2} w\big] \big/ \,\mathrm{E}_\bullet \tfrac{u^2}{q^2} w \tag{7.3.18}$$

$$\alpha_3 = \frac{1 + \gamma^{-1} - \mathrm{E}\Big[\mathrm{E}_\bullet \tfrac{u^4}{q^2} w \big/ \mathrm{E}_\bullet \tfrac{u^2}{q^2} w\Big]}{\mathrm{E}\, \tfrac{u^6}{q^2} w - \mathrm{E}\Big[\big(\mathrm{E}_\bullet \tfrac{u^4}{q^2} w\big)^2 \big/ \mathrm{E}_\bullet \tfrac{u^2}{q^2} w\Big]} \tag{7.3.19}$$

(b) *The unique solution to the corresponding MSE problem (7.3.13) coincide with the solution of problem (7.3.10), where bound* $b_{\mathrm{rg}} \in (0,\infty)$, *respectively bound* $b_{\mathrm{sc},\gamma} \in (1,\infty)$ *and* $r \in (0,\infty)$ *are related via*

$$r^2 b_{\mathrm{rg}} = \mathrm{E}\, \frac{1}{|\mathcal{K}^{-1}x|^2}\, \mathrm{E}_\bullet \big[|\mathcal{K}^{-1}x|\, |\alpha_1(x)u + \alpha_3 u^3| - b_{\mathrm{rg}} q^2\big]_+ \mathrm{I}\{b_{\mathrm{rg}}|u| < (1 + b_{\mathrm{sc},\gamma})|\mathcal{K}^{-1}x|\} \tag{7.3.20}$$

respectively

$$r^2\gamma^2 b_{\mathrm{sc},\gamma} = \mathrm{E}\,\mathrm{E}_\bullet \left[|\alpha_1(x) + \alpha_3 u^2| - (1 + b_{\mathrm{sc},\gamma}) \frac{q^2}{u^2} \right]_+ \mathrm{I}\{ b_{\mathrm{rg}} |u| > (1 + b_{\mathrm{sc},\gamma}) |\mathcal{K}^{-1} x| \} \tag{7.3.21}$$

(c) *Given* $\gamma \in \mathbb{R} \setminus \{0\}$ *let* $\bar{b}_{\mathrm{rg}}(\gamma) \in (0, \infty)$ *and* $\bar{b}_{\mathrm{sc},\gamma}(\gamma) \in (1, \infty)$ *be minimal such that*

$$\mathrm{E}_\bullet |u| \min \left\{ \frac{\bar{b}_{\mathrm{rg}}(\gamma)}{|\mathcal{K}^{-1} x|}, \frac{1 + \bar{b}_{\mathrm{sc},\gamma}(\gamma)}{|u|} \right\} \geq 1 \qquad \text{a.e. } K(dx) \tag{7.3.22}$$

and

$$\mathrm{E}\,|u|^3 \min \left\{ \frac{\bar{b}_{\mathrm{rg}}(\gamma)}{|\mathcal{K}^{-1} x|}, \frac{1 + \bar{b}_{\mathrm{sc},\gamma}(\gamma)}{|u|} \right\} \geq 1 + \gamma^{-1} \tag{7.3.23}$$

Then, $\left(\omega^{\min}_{\mathrm{Ms},2}\right)^2 \geq \bar{b}_{\mathrm{rg}}(\gamma)^2 + \gamma^2 \bar{b}_{\mathrm{sc},\gamma}(\gamma)^2$ *. If there is equality in (7.3.22) and (7.3.23) for*

$$\bar{\gamma} = \operatorname{argmin} \left\{ \bar{b}_{\mathrm{rg}}(\gamma)^2 + \gamma^2 \bar{b}_{\mathrm{sc},\gamma}(\gamma)^2 \right\} \tag{7.3.24}$$

then $\bar{\psi}$ *of form*

$$\bar{\psi}(x,u) = \bar{b}_{\mathrm{rg,sc}}(x,u) \operatorname{sign}(u) \qquad \bar{b}_{\mathrm{rg,sc}}(x,u) = \min \left\{ \frac{\bar{b}_{\mathrm{rg}}(\bar{\gamma})}{|\mathcal{K}^{-1} x|}, \frac{1 + \bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma})}{|u|} \right\} \tag{7.3.25}$$

is the solution to problem (7.3.14), where

$$\left(\omega^{\min}_{\mathrm{Ms},2}\right)^2 = \bar{b}_{\mathrm{rg}}(\bar{\gamma})^2 + \bar{\gamma}^2 \bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma})^2 = \bar{b}_{\mathrm{rg}}(\bar{\gamma})^2 + \bar{b}_{\mathrm{sc}}(\bar{\gamma})^2 \tag{7.3.26}$$

and the corresponding $\bar{\rho}$ *attains this minimum bias.*

Proof Assume ideal error distribution $F = \mathcal{N}(0,1)$.

(a) By similar arguments as in proof of Theorem 7.2.19 this problem has a unique solution and is well-posed in sense of Definition B.2.9 of Rieder (1994). Therefore, Theorem B.2.6 of Rieder (1994) is applicable and there exist some multiplier $\alpha_3 \in \mathbb{R}$ such that the solution $\tilde{\psi}$ minimizes the Lagrangian

$$L(\psi) = \mathrm{E}\,\mathrm{E}_\bullet \left[q^2\psi^2 - 2\psi(\gamma^2 u + \alpha_3 u^3) \right] \ (+\gamma^2) \tag{7.3.27}$$

among all $\psi \in L_2(P)$ satisfying (Mc2) and (7.3.12) (condition (Mc15) will be fulfilled automatically). This Lagrangian can be minimized by minimization of the conditional expectation inside, at each point x, subject to (Mc2) and

$$|\psi(x,u)| \leq b_{\mathrm{rg,sc}}(x,u) \qquad \forall\, u \tag{7.3.28}$$

The conditional problem given any x is also well-posed similarly to the proof of Theorem 7.2.19. Thus, there exist some $\tilde{\alpha}_1(x) \in \mathbb{R}$ such that the x section of the solution $\tilde{\psi}$ may be found by minimization of the conditional Lagrangian

$$L_\bullet(\psi) = \mathrm{E}_\bullet \left[q^2\psi^2 - 2\psi(\gamma^2 u + \tilde{\alpha}_1(x) u + \alpha_3 u^3) \right] \tag{7.3.29}$$

$$= \mathrm{E}_\bullet\, q^2 \big(\psi - g\big)^2 + \mathrm{const} \tag{7.3.30}$$

over all $\psi(x,\cdot) \in L_2(\mathcal{N}(0,1))$ such that (7.3.28) holds. Pointwise minimization of the integrand yields the asserted form (7.3.15) of the optimal section $\tilde{\psi}(x,\cdot)$ and $\tilde{\psi}(x,\cdot)$ is odd u a.e. $K(dx)$ (i.e., (Mc15) holds), where $\alpha_1(x) = \gamma^2 + \tilde{\alpha}_1(x)$.
Conversely, if there is some $\tilde{\rho} \in \Psi_{\mathrm{Ms}}$ with $\tilde{\psi}$ of form (7.3.15) for any function $\alpha_1(x)$ from $\mathbb{R}^k$ to $\mathbb{R}$ and any $\alpha_3 \in \mathbb{R}$. Then, (K almost) all x sections minimize the corresponding conditional Lagrangian $L_\bullet$ among all $\psi(x,\cdot) \in L_2(\mathcal{N}(0,1))$ subject to (7.3.28), hence $\mathrm{E}_\bullet\left[q^2\psi^2 - 2\gamma^2\psi - 2\alpha_3\psi\right]$ over $\psi \in L_2(P_{\theta_0})$ subject to (Mc2) and (7.3.28). On integration with respect to $K(dx)$, $\tilde{\psi}$ minimizes the Lagrangian L in (7.3.27) over $\psi \in L_2(P_{\theta_0})$ subject to (Mc2) and (7.3.12) and therefore $\tilde{\psi}$ solves (7.3.10).
To prove (7.3.18), we solve (Mc2) for $\alpha_1(x)$ and insert the result in (M4), which we then solve for α_3.

(b) Putting $b_{\mathrm{rg}} := \sup_P |\mathcal{K}^{-1}x\psi|$ and $b_{\mathrm{sc},\gamma} := \sup_P |u\psi - 1|$, $\tilde{\psi}$ also solves the problem

$$\mathrm{E}\,|\rho|^2 = \min! \qquad \rho \in \Psi_{\mathrm{Ms}},\ \omega_{\mathrm{Ms},2}(\rho) \le b_{\mathrm{rg,sc}} \tag{7.3.31}$$

where $\omega_{\mathrm{M},2}(\rho) \le b_{\mathrm{rg,sc}}$ stands for

$$\sup_P |\mathcal{K}^{-1}x\psi| \le b_{\mathrm{rg}} \qquad \sup_P |u\psi - 1| \le b_{\mathrm{sc},\gamma} \tag{7.3.32}$$

which has already occurred as problem (7.3.10) and the solution $\tilde{\psi}$ is of form (7.3.15). Plugged into the corresponding Lagrangian L, which is $\mathrm{E}_K L_\bullet$ with $L_\bullet$ defined in (7.3.30), we get a convex function of the bounds

$$L_1(b_0) = \mathrm{E}\, q^2\big(|g(x,u)| - b_{\mathrm{rg,sc}}(x,u)\big)_+^2 + r^2 b_{\mathrm{rg}}^2 + r^2\gamma^2 b_{\mathrm{sc},\gamma}^2 \tag{7.3.33}$$

Analogously to Lemma C.2.3 of Rieder (1994), differentiation of L_1 by b_{rg}, respectively $b_{\mathrm{sc},\gamma}$ yields (7.3.20) and (7.3.21), respectively.

(c) For arbitrary measurable $\alpha_1(x) \ge 0$ a.e. $K(dx)$, $a_3 \ge 0$ and $\psi \in \Psi_{\mathrm{Ms}}$ it holds

$$\begin{aligned}
\mathrm{E}\,\alpha_1(x) + \alpha_3(1+\gamma^{-1}) &= \mathrm{E}(\alpha_1(x)u + \alpha_3 u^3)\psi && (7.3.34)\\
&\le \mathrm{E}(\alpha_1(x)|u| + \alpha_3|u|^3)|\psi| && (7.3.35)\\
&\le \mathrm{E}(\alpha_1(x)|u| + \alpha_3|u|^3) b_{\mathrm{rg,sc}}(x,u) && (7.3.36)\\
&= \mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u) + \alpha_3\,\mathrm{E}\,|u|^3 b_{\mathrm{rg,sc}}(x,u) && (7.3.37)
\end{aligned}$$

Hence, we obtain the following lower bound for b_{rg} and $b_{\mathrm{sc},\gamma}$

$$1 \le \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u) + \alpha_3\,\mathrm{E}\,|u|^3 b_{\mathrm{rg,sc}}(x,u)}{\mathrm{E}\,\alpha_1(x) + \alpha_3(1+\gamma^{-1})} =: \frac{d_1 + d_2\alpha_3}{d_3 + d_4\alpha_3} \tag{7.3.38}$$

Given $\alpha_1(\cdot) \ge 0$ the minimization of the right hand side via differentiation in α_3 yields that the minimum is attained for $\alpha_3 = 0$, if $d_2 d_3 > d_1 d_4$, respectively, for $\alpha_3 = \infty$, if $d_2 d_3 < d_1 d_4$. In case of $\alpha_3 = \infty$, (7.3.38) yields the bound (7.3.23) for b_{rg} and $b_{\mathrm{sc},\gamma}$. If $\alpha_3 = 0$, we get

$$1 \le \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u)}{\mathrm{E}\,\alpha_1(x)} \tag{7.3.39}$$

We now minimize the right hand side of (7.3.39) in $\alpha_1(\cdot) \geq 0$. If $\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}} \geq m$ a.e. $K(dx)$, it follows

$$\min_{\alpha_1(\cdot)\geq 0} \frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u)}{\mathrm{E}\,\alpha_1(x)} \geq \inf{}_{K(dx)}\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u) =: m_0 \tag{7.3.40}$$

On the other hand, if $\alpha_1(\cdot)$ is concentrated on points x with $\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}} \leq m_0 + \delta$ for some $\delta > 0$, we have

$$\frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u)}{\mathrm{E}\,\alpha_1(x)} \leq m_0 + \delta \tag{7.3.41}$$

Thus

$$\frac{\mathrm{E}\,\alpha_1(x)\,\mathrm{E}_\bullet\,|u| b_{\mathrm{rg,sc}}(x,u)}{\mathrm{E}\,\alpha_1(x)} = m_0 \tag{7.3.42}$$

holds, where $m_0 \geq 1$ by (7.3.38). This yields the second lower bound (7.3.22) for b_{rg} and $b_{\mathrm{sc},\gamma}$. We now determine $\bar{b}_{\mathrm{rg}}(\bar{\gamma})$ and $\bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma})$. By (7.3.22) and (7.3.23) we get $\big(\omega_{\mathrm{Ms},2}^{\min}\big)^2 \geq \bar{b}_{\mathrm{rg}}(\gamma)^2 + \gamma^2 \bar{b}_{\mathrm{sc},\gamma}(\gamma)^2$, since (Mc2) and (M4) must hold for every $\psi \in \Psi_{\mathrm{Ms}}$. However, if we get equality in (7.3.22) and (7.3.23) then $\big(\omega_{\mathrm{Ms},2}^{\min}\big)^2 = \bar{b}_{\mathrm{rg}}(\bar{\gamma})^2 + \gamma^2 \bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma})^2$, since $\bar{\psi}(x,u)$ as defined in (7.3.25) fulfills (Mc15), (Mc2) and (M4). That is, $\bar{\psi}$ is the solution to problem (7.3.14) and the corresponding $\bar{\rho}$ achieves the minimum bias $\omega_{\mathrm{Ms},2}^{\min}$. ////

Remark 7.3.5 **(a)** To ensure that $\tilde{\psi}$ in part (a) of the proof is measurable and to justify the minimization of expectations by pointwise minimization of integrands, the function $\alpha_1(x)$ has to be measurable. We can not guarantee this by Lagrange arguments, but our fixed point algorithm supplies continuous functions $\alpha_1(x)$.

(b) Starting from $\tilde{\psi}_{\theta_0}$ and defining $\tilde{\psi}_\theta$ by (7.1.54) and fixing the same $\gamma \in \mathbb{R}$, we get that this $\tilde{\psi}_\theta$ is also of form (7.3.15) with

$$\alpha_{1,\theta}(x) = \sigma^2 \alpha_{1,\theta_0}(x) \qquad \alpha_{3,\theta} = \sigma^2 \alpha_{3,\theta_0} \qquad b_{\mathrm{rg},\theta} = \sigma b_{\mathrm{rg},\theta_0} \qquad b_{\mathrm{sc},\gamma,\theta} = \sigma b_{\mathrm{sc},\gamma,\theta_0} \tag{7.3.43}$$

Thus, $\tilde{\psi}_\theta$ is the solution to the corresponding Hampel type problem (7.3.10) with bounds $\sigma b_{\mathrm{rg},\theta_0}$ and $\sigma b_{\mathrm{sc},\theta_0}$.

(c) In case of the Ms-estimators we don't give the specializations of the solutions to elliptically and spherically symmetric ideal regressor distribution K, as the forms of the solutions stay unchanged. But, indeed, this specializations lead to a simplification of the numerical algorithms.

(d) We determine the minimax asymptotic MSE solution via numerical optimization in b_{loc} and b_{sc} where

$$\sup{}_P\,|\rho|^2 = \sup{}_P\,[\tilde{\psi}^2 + \gamma^2 (u\tilde{\psi} - 1)^2] = b_{\mathrm{loc}}^2 + \gamma^2 b_{\mathrm{sc},0}^2 \tag{7.3.44}$$

(e) In the normal location and scale model, our numerical calculations yield equality (up to an error less than 10^{-8}) in (7.3.22) and (7.3.23) for the optimal $\bar{\gamma}$. Hence,

$$\bar{\psi}(u) = \bar{b}_{\mathrm{loc,sc}}(u)\,\mathrm{sign}\,(u) \qquad \bar{b}_{\mathrm{loc,sc}}(u) = \min\Big\{\bar{b}_{\mathrm{loc}}(\bar{\gamma}), \frac{1 + \bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma})}{|u|}\Big\} \tag{7.3.45}$$

with $\bar{\gamma} \approx 1.395$, $\bar{b}_{\mathrm{loc}}(\bar{\gamma}) \approx 1.360$ and $\bar{b}_{\mathrm{sc},\gamma}(\bar{\gamma}) = 1$; confer also Remark 7.3.8 (c). ////

7.3.3 BM Estimators

We now consider the estimators for regression and scale as invented by Bednarski and Müller (2001), which we call BM estimators. They introduce a class of (suboptimal) M estimators with ICs as given in (7.2.91) and propose estimators for the regression model with unknown scale and ideal error distribution $F = \mathcal{N}(0,1)$ by solving a location and scale problem at each design point x. Therefore, this proposal is limited to ideal regressor distributions K with finite support.
Bednarski and Müller (2001) separately minimize $\mathrm{E}_\bullet\, \psi^2$ and $\mathrm{E}_\bullet\, |u\psi - 1|^2$, where (as before) $\mathrm{E}_\bullet$ is expectation with respect to $F(du)$ given x. More precisely, they use the side conditions (Mc15),(Mc2) and (Mc4) for $\psi \in L_2(F)$. Then, they first solve the Hampel type problem

$$\mathrm{E}_\bullet\, \psi^2 = \min! \qquad \text{(Mc15), (Mc2)}, \ \omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}} \tag{7.3.46}$$

where x and $\gamma_x \in \mathbb{R}$ are fixed and $\omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}}$ stands for

$$\sup_\bullet |\psi| \le b_{\mathrm{rg},x} := b_{\mathrm{rg}}/|\mathcal{K}^{-1}x| \qquad \sup_\bullet |u\psi - 1| \le b_{\mathrm{sc},0,x} := b_{\mathrm{sc}}/\gamma_x \tag{7.3.47}$$

which under (Mc15),(Mc2) and for $b_{\mathrm{sc},0,x} \ge 1$ is equivalent to

$$|\psi(x,u)| \le b_{\mathrm{rg,sc}}(x,u) := \min\left\{b_{\mathrm{rg},x}, \frac{1 + b_{\mathrm{sc},0,x}}{|u|}\right\} \qquad \forall u \in \mathbb{R} \tag{7.3.48}$$

Additionally, for fixed x and fixed γ_x we derive the minimal bound $\bar{b}_{\mathrm{rg},x}$, which is attainable for some given bound $b_{\mathrm{sc},0,x} \ge 1$. That is, if $\bar{\psi}$ attains this minimal $\bar{b}_{\mathrm{rg},x}$, then it is the solution to

$$\sup_\bullet |\psi| = \min! \qquad \text{(Mc15), (Mc2)}, \ \sup_\bullet |u\psi - 1| \le b_{\mathrm{sc},0,x} \tag{7.3.49}$$

The solutions to these problems may be read off from Proposition 1, respectively Theorem 1 in Bednarski and Müller (2001). However, they use an interior point assumption to guarantee well-posedness of these problems and give only incomplete Lagrange arguments. Hence, we will give our own proof to the following Proposition.

Proposition 7.3.6 *Assume ideal error distribution $F = \mathcal{N}(0,1)$.*

(a) *Let x and $\gamma_x \in \mathbb{R}$ be fixed and let $b_{\mathrm{rg},x}$, $b_{\mathrm{sc},0,x}$ be such that there exists some ψ with (Mc15),(Mc2) and (7.3.48). Then, the unique solution $\tilde{\psi}$ to problem (7.3.46) is*

$$\tilde{\psi}(x,u) = \mathrm{sign}\,(u) \min\left\{\alpha_x|u|, b_{\mathrm{rg},x}, \frac{1 + b_{\mathrm{sc},0,x}}{|u|}\right\} \tag{7.3.50}$$

where $\alpha_x \ge 0$ for all x and the following representation holds

$$1 = \mathrm{E}_\bullet \min\left\{\alpha_x u^2, b_{\mathrm{rg},x}|u|, (1 + b_{\mathrm{sc},0,x})\right\} \tag{7.3.51}$$

(b) *For fixed* x *and given* $\gamma_x \in \mathbb{R}$ *it holds that*

$$1 = 2\bar{b}_{\mathrm{rg},x}\left[1/\sqrt{2\pi} - \varphi(\bar{\delta}_{\mathrm{rg},x}) + \bar{\delta}_{\mathrm{rg},x}\Phi(-\bar{\delta}_{\mathrm{rg},x})\right] \tag{7.3.52}$$

where $\bar{\delta}_{\mathrm{rg},x} := \bar{b}^{-1}_{\mathrm{rg},x}(1+b_{\mathrm{sc},0,x})$ *and* $b_{\mathrm{sc},0,x} \geq 1$. *Here* φ *and* Φ *are the density and the cumulative distribution function of* $\mathcal{N}(0,1)$.
There exists $\bar{\psi}$ *with (Mc15), (Mc2) and* $\sup_\bullet |u\psi - 1| \leq b_{\mathrm{sc},0,x}$ *achieving* $\bar{b}_{\mathrm{rg},x}$. *And then necessarily*

$$\bar{\psi}(x,u) = \operatorname{sign}(u) \min\left\{\bar{b}_{\mathrm{rg},x}, \frac{1+b_{\mathrm{sc},0,x}}{|u|}\right\} \tag{7.3.53}$$

is the unique solution.

Proof Assume ideal error distribution $F = \mathcal{N}(0,1)$.

(a) We neglect conditions (M1) and (M5), respectively condition (Mc15) as they will be fulfilled automatically. Then, by analogous arguments as in proof of Theorem 7.2.11 problem (7.3.46) has a unique solution. Moreover, well-posedness in sense of Definition B.2.9 in Rieder (1994) can also be shown similarly. Given some $b_{\mathrm{rg},x}$ and $b_{\mathrm{sc},0,x}$ such that there exists some ψ with (Mc2) and (7.3.48), assume some ψ_1 with (Mc2) and

$$|\psi_1| \leq b'_{\mathrm{rg,sc}}(x,u) := \min\left\{b'_{\mathrm{rg},x}, \frac{1+b'_{\mathrm{sc},0,x}}{|u|}\right\} \qquad \forall u \in \mathbb{R} \tag{7.3.54}$$

where $b'_{\mathrm{rg},x} < b_{\mathrm{rg},x}$ and $b'_{\mathrm{sc},0,x} < b_{\mathrm{sc},0,x}$. Then, the set of functions

$$\psi_\epsilon(x,u) := \psi_1(x,u) + \epsilon \operatorname{sign}(u)\, \mathrm{I}(|u| \leq 1) \tag{7.3.55}$$

with $0 < |\epsilon| \leq \min\{b_{\mathrm{rg},x} - b'_{\mathrm{rg},x}, b_{\mathrm{sc},0,x} - b'_{\mathrm{sc},0,x}\}$ covers a full neighborhood of 1 in (Mc2) and $|\psi_\epsilon(x,u)| \leq b_{\mathrm{rg,sc}}(x,u)$ for all u.
Furthermore, the constraints $\sup_\bullet |u\psi - 1| \leq b_{\mathrm{sc},0,x}$ and $\sup_\bullet |\psi| \leq b_{\mathrm{rg},x}$ are compatible with rescaling of ψ by some $\delta \in (0,1)$. Therefore, we may change (Mc2) to

$$\mathrm{E}_\bullet\, u\psi \geq 1 \tag{7.3.56}$$

Now, Theorem B.2.6 of Rieder (1994) yields some Lagrange multiplier $\alpha_x \geq 0$ such that the solution $\tilde{\psi}$ minimizes

$$L_\bullet(\psi) = \mathrm{E}_\bullet\, \psi^2 - 2\alpha_x\, \mathrm{E}_\bullet\, u\psi = \mathrm{E}_\bullet(\psi - \alpha_x u)^2 + \mathrm{const} \tag{7.3.57}$$

subject to $|\psi| \leq b_{\mathrm{rg,sc}}$ for fixed x and given $\gamma_x \in \mathbb{R}$. This can be solved by pointwise minimization of the integrand and the solution $\tilde{\psi}$ is of form (7.3.50) which is an odd function in u for given x. To get the representation (7.3.51), insert $\tilde{\psi}$ in (Mc2).

(b) By arguments as in Rieder (1994) (p 198) the minimal $\bar{b}_{\mathrm{rg},x}$ is attained. Therefore, if $\bar{\psi}$ achieves $\bar{b}_{\mathrm{rg},x}$ it solves the convex, well-posed problem (7.3.49).

Now, Theorem B.2.3 of Rieder (1994) is applicable and there exists some multiplier $\alpha_x \in \mathbb{R}$ such that $\bar{\psi}$ is the solution to

$$L_\bullet(\psi) = \sup_\bullet |\psi| - \alpha_x \, \mathrm{E}_\bullet \, u\psi = \min! \qquad \text{s.t. } \sup_\bullet |u\psi - 1| \le b_{\mathrm{sc},0,x} \tag{7.3.58}$$

where x and $\gamma_x \in \mathbb{R}$ are fixed. At this point, we may again neglect the constraint (Mc15) as it will be fulfilled automatically. Now, as mentioned in part (a) of this proof, the constraints $\sup_\bullet |u\psi - 1| \le b_{\mathrm{sc},0,x}$ and $\sup_\bullet |\psi| \le b_{\mathrm{rg},x}$ are compatible with rescaling of ψ. Therefore, we may change (Mc2) to (7.3.56) and get $\alpha_x \ge 0$. Thus (7.3.58) is equivalent to

$$\mathrm{E}_\bullet \, u\psi = \max! \qquad \text{s.t. } \sup_\bullet |\psi| \le \bar{b}_{\mathrm{rg},x}, \ \sup_\bullet |u\psi - 1| \le b_{\mathrm{sc},0,x} \tag{7.3.59}$$

This can be solved by pointwise maximization of the integral and the unique solution $\bar{\psi}$ is of form (7.3.53), which is odd in u for given x. To obtain (7.3.52), insert $\bar{\psi}$ in (Mc2) and use $F = \mathcal{N}(0,1)$. ////

Second, Bednarski and Müller (2001) give the solution to the Hampel type problem

$$\mathrm{E}_\bullet \, |u\psi - 1|^2 = \min! \qquad \text{(Mc15),(Mc2)}, \ \omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}} \tag{7.3.60}$$

for special $\gamma_x \in \mathbb{R}$. Additionally, for fixed x and given $\gamma_x \in \mathbb{R}$ we derive the minimal bound $\bar{b}_{\mathrm{sc},0,x}$, which is attainable for some given bound $b_{\mathrm{rg},x} \ge \sqrt{\pi/2}$. That is, if $\bar{\psi}$ attains this minimal $\bar{b}_{\mathrm{sc},0,x}$, then it is the solution to

$$\sup_\bullet |u\psi - 1| = \min! \qquad \text{(Mc15), (Mc2), } \sup_\bullet |\psi| \le b_{\mathrm{rg},x} \tag{7.3.61}$$

The solutions may be found in Proposition 2, respectively Theorem 2 of Bednarski and Müller (2001). However, they again use an interior point assumption to guarantee well-posedness and again give only incomplete Lagrange arguments. Hence, we will give our own proof to the following Proposition, too.

Proposition 7.3.7 *Assume ideal error distribution* $F = \mathcal{N}(0,1)$.

(a) *Let* x *be fixed and determine* $\gamma_x \in \mathbb{R}$ *for the solution to the corresponding problem (7.3.46) via (Mc4). Then, the unique solution to problem (7.3.46) also solves problem (7.3.60) and the representation (7.3.51) holds.*

(b) *For fixed* x *and given* $\gamma_x \in \mathbb{R}$ *it holds that*

$$1 \le 2b_{\mathrm{rg},x}\big[1/\sqrt{2\pi} \, - \varphi(\bar{\delta}_{\mathrm{sc},0,x}) + \bar{\delta}_{\mathrm{sc},0,x}\Phi(-\bar{\delta}_{\mathrm{sc},0,x})\big] \tag{7.3.62}$$

where $\bar{\delta}_{\mathrm{sc},0,x} := b_{\mathrm{rg},x}^{-1}(1 + \bar{b}_{\mathrm{sc},0,x})$ *and* $b_{\mathrm{rg},x} \ge \sqrt{\pi/2}$. *Here* φ *is the density and* Φ *the cumulative distribution function of* $\mathcal{N}(0,1)$.
There exist $\bar{\psi}$ *with (Mc15), (Mc2) and* $\sup_\bullet |\psi| \le b_{\mathrm{rg},x}$ *achieving* $\bar{b}_{\mathrm{sc},0,x}$. *And then necessarily*

$$\bar{\psi}(x,u) = \mathrm{sign}\,(u) \min\left\{b_{\mathrm{rg},x}, \frac{1 + \bar{b}_{\mathrm{sc},0,x}}{|u|}\right\} \tag{7.3.63}$$

it the unique solution.

PROOF Assume ideal error distribution $F = \mathcal{N}(0,1)$.

(a) Using (Mc2) for fixed x and given some $\gamma_x \in \mathbb{R}$ problem (7.3.60) may be equivalently rewritten as

$$\mathrm{E}_\bullet\, u^2\psi^2 = \min! \qquad \text{(Mc15),(Mc2)}, \ \omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}} \tag{7.3.64}$$

By (Mc4) this equivalent to

$$\mathrm{E}_\bullet(\alpha_x u^2 - u\psi)^2 = \min! \qquad \text{(Mc15),(Mc2)}, \ \omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}} \tag{7.3.65}$$

where α_x is the multiplier determined in the corresponding problem (7.3.46). Now, we neglect the constraints (Mc15) and (Mc2) and consider the problem

$$\mathrm{E}_\bullet(\alpha_x u^2 - u\psi)^2 = \min! \qquad \omega_{\mathrm{BM}}(\psi) \le b_{\mathrm{rg,sc}} \tag{7.3.66}$$

If we fix the special γ_x, which corresponds to the unique solution $\tilde{\psi}$ of problem (7.3.46) via (Mc4), the solutions to problem (7.3.66) and problem (7.3.46) coincide. Moreover (Mc15) and (Mc2) are fulfilled and the representation (7.3.51) holds. Thus, $\tilde{\psi}$ also solves problem (7.3.60).

(b) By arguments as in the proof of Proposition 7.3.6, problem (7.3.61) may be equivalently rewritten as problem (7.3.59) with $b_{\mathrm{sc},0,x}$ replaced by $\bar{b}_{\mathrm{sc},0,x}$ and $\bar{b}_{\mathrm{rg},x}$ replaced by $b_{\mathrm{rg},x}$, respectively. Thus, the solution is of form (7.3.63) and (7.3.62) is obtained. ////

Remark 7.3.8 **(a)** Bednarski and Müller (2001) pretend to work with condition (M4) instead of (Mc4) which makes no difference in case of location and scale. However, if one wants to apply Proposition 2 (ibid.) to regression and scale, that is, apply the results of location and scale to each x, the stricter condition (Mc4) is needed. More precisely, a closer look at equation (3.15) (ibid.) (cf. also (7.3.65)) shows that it is not possible to choose one m for all x but one gets some m_x for each solution $\tilde{\psi}_x$ to problem (7.3.46) where m_x is our γ_x^{-1}. As a consequence, the gross error sensitivity bound s for the scale part which occurs in Section 4 of Bednarski and Müller (2001) depends on x, too.

(b) The lower bound for the regression gross error sensitivity given on p 368 of Bednarski and Müller (2001) indeed applies to linear regression (without scale), but, only if one, in addition to identifiability, considers a discrete design with linearly independent regressors; confer Theorem 2 of Kurotschka and Müller (1992). That is, in case of a one-dimensional regressor this result essentially covers only location and scale.

(c) As the subsequent arguments show, the solutions for the simultaneous estimation of regression and scale in case of BM estimators (contrary to AL and M estimators) coincide with the solutions for the separate estimation. It holds,

$$\sup_{x,u} |\mathcal{K}^{-1} x\tilde{\psi}(x,u)| = b_{\mathrm{rg}} \qquad \text{and} \qquad \sup_{x,u} |\gamma_x(u\tilde{\psi}(x,u) - 1)| = b_{\mathrm{sc}} \tag{7.3.67}$$

For fixed x we obtain

$$|\tilde{\psi}(x,u)| = b_{\mathrm{rg},x} \qquad \text{if} \quad \frac{b_{\mathrm{rg},x}}{\alpha_x} \le |u| \le \frac{1 + b_{\mathrm{sc},0,x}}{b_{\mathrm{rg},x}} \tag{7.3.68}$$

Consequently,

$$\alpha_x \geq \frac{b_{\mathrm{rg},x}^2}{1 + b_{\mathrm{sc},0,x}} \tag{7.3.69}$$

Moreover,

$$|u\tilde{\psi}(x,u) - 1| = b_{\mathrm{sc},0,x} \qquad \text{if} \quad |u| \geq \max\left\{\sqrt{\frac{1 + b_{\mathrm{sc},0,x}}{\alpha_x}}\,,\ \frac{1 + b_{\mathrm{sc},0,x}}{b_{\mathrm{rg},x}}\right\} \tag{7.3.70}$$

where

$$\sqrt{\frac{1 + b_{\mathrm{sc},0,x}}{\alpha_x}} > \frac{1 + b_{\mathrm{sc},0,x}}{b_{\mathrm{rg},x}} \qquad \Longleftrightarrow \qquad \alpha_x < \frac{b_{\mathrm{rg},x}^2}{1 + b_{\mathrm{sc},0,x}} \tag{7.3.71}$$

Thus, (7.3.69) entails

$$|u\tilde{\psi}(x,u) - 1| = b_{\mathrm{sc},0,x} \qquad \text{if} \quad |u| \geq \frac{1 + b_{\mathrm{sc},0,x}}{b_{\mathrm{rg},x}} \tag{7.3.72}$$

That is, $|\tilde{\psi}(x,u)| = b_{\mathrm{rg},x}$ and $|u\tilde{\psi}(x,u) - 1| = b_{\mathrm{sc},0,x}$ if $|u| = \frac{1+b_{\mathrm{sc},0,x}}{b_{\mathrm{rg},x}}$ which implies

$$\sup_{x,u} |\tilde{\rho}(x,u)|^2 = b_{\mathrm{rg}}^2 + b_{\mathrm{sc}}^2 \tag{7.3.73}$$

(d) Since BM estimators are derived by combining the different solutions for each design point x, we do not specialize the solutions to elliptically, respectively spherically symmetric ideal regressor distribution K.

(e) Starting from $\tilde{\psi}_{\theta_0}$ and defining $\tilde{\psi}_\theta$ by (7.1.54) for given x and fixing the same $\gamma_x \in \mathbb{R}$, we get that this $\tilde{\psi}_\theta$ is also of form (7.3.50) with

$$\alpha_\theta = \sigma^2 \alpha_{\theta_0} \qquad b_{\mathrm{rg},x,\theta} = \sigma b_{\mathrm{rg},x,\theta_0} \qquad b_{\mathrm{sc},0,x,\theta} = \sigma b_{\mathrm{sc},0,x,\theta_0} \tag{7.3.74}$$

Thus, $\tilde{\psi}_\theta$ is the solution to the corresponding Hampel type problem (7.3.46), respectively (7.3.60) with bounds $b_{\mathrm{rg},x,\theta} = \sigma b_{\mathrm{rg},x,\theta_0}$ and $b_{\mathrm{sc},0,x,\theta} = \sigma b_{\mathrm{sc},0,x,\theta_0}$.

(f) The optimal BM estimators are much harder to compute than the optimal Ms estimators. In case of Ms estimators the outer optimization is one-dimensional and can easily be carried out using the `R` function `optimize`; confer R Development Core Team (2005). In contrast, the outer optimization in case of BM estimators (via γ_x) depends on the number of design points. That is, if the regressor support consists of m different points, we have to solve an $m+1$-dimensional optimization problem. Consequently, BM estimators are not only restricted to regressors with finite support, but to regressors with finite support and only a moderate number of design points.

(g) In case of normal location and scale, the optimal Ms and BM estimators for $r \in (0,\infty)$ would coincide, if $\alpha_1 = \alpha$ and $\alpha_3 = \alpha\gamma^2$, that is, $\gamma^2 = \alpha_3/\alpha_1$. But in all cases, in which we numerically evaluated α, γ with bounds b_{loc}, $b_{\mathrm{sc},\gamma}$ optimally chosen from equations (7.3.20) and (7.3.21), we didn't get this coincidence; confer Table 7.1. However, the differences are only very small and the solutions indeed coincide for $r = \infty$; confer Remarks 7.3.5 and 8.4.1. In contrast, the efficiency losses in case of regression and scale, for which the BM estimators were announced, are clearly larger; confer Section 7.5. ////

r	α_3/α_1	γ^2_{BM}	γ_{Ms}	γ_{BM}	$\mathrm{MSE}_{\mathrm{Ms}}$	$\mathrm{MSE}_{\mathrm{BM}}$
0.1	0.326	0.335	0.580	0.578	1.6595	1.6595
0.25	0.473	0.523	0.725	0.723	2.0727	2.0728
0.5	0.708	0.906	0.953	0.952	3.0559	3.0563
1.0	1.067	1.627	1.276	1.276	6.1517	6.1519
1.5	1.545	1.862	1.365	1.365	10.9396	10.9396
2.0	2.281	1.884	1.373	1.373	17.6024	17.6024

Table 7.1: Comparison between Ms and BM estimators in case of normal location and scale.

7.4 One-Step Construction

We restrict our considerations to AL estimators and unconditional contamination neighborhoods. The one-step construction in case of conditional neighborhoods is a separate issue we do not treat in this thesis.

7.4.1 Normal Regression

We consider the linear regression model (7.1.1) without scale parameter σ. By Theorem 2.4.6 of Rieder (1994), this model is L_2 differentiable at every $\theta \in \mathbb{R}^k$ with L_2 derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by

$$\Lambda_\theta(x,y) = x\Lambda_f^{\mathrm{loc}}(y - x^\tau\theta) \qquad \text{and} \qquad \mathcal{I}_\theta = \mathcal{I}_F^{\mathrm{loc}}\mathcal{K} \tag{7.4.1}$$

Now, we want to apply Theorem 2.3.3. Given $\theta \in \mathbb{R}^k$ let $\theta_n = \theta + h/\sqrt{n}$ for some $h \in \mathbb{R}^k$. Then,

$$\mathcal{L}_{P_{\theta_n}}(\Lambda_{\theta_n}) = \mathcal{L}_{P_\theta}(\Lambda_\theta) \qquad \text{and} \qquad \mathcal{I}_{\theta_n} = \mathcal{I}_\theta \tag{7.4.2}$$

i.e., (2.1.102) and (2.1.103) hold. Assuming $F = \mathcal{N}(0,1)$, we get $\Lambda_\theta = x(y - x^\tau\theta)$ and conditions (2.3.10) and (2.3.11) read

$$\sup_{\mathcal{D}} \left\{ |xx^\tau h/\sqrt{n}\,| \right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{7.4.3}$$

and

$$\sup_{^c\mathcal{D}} \left\{ \frac{|xx^\tau h/\sqrt{n}\,|}{|A_\theta x(y - x^\tau\theta)|} \right\} \longrightarrow 0 \qquad \text{as } n \to \infty \tag{7.4.4}$$

where

$$\mathcal{D} = \left\{ (x,y) \in \mathbb{R}^{k+1} \,\middle|\, |A_{\theta_n}\Lambda_{\theta_n}| \le b_{\theta_n} \right\} \cup \left\{ (x,y) \in \mathbb{R}^{k+1} \,\middle|\, |A_\theta\Lambda_\theta| \le b_\theta \right\} \tag{7.4.5}$$

and $\mathcal{D} = \mathcal{D}_n$. To prove conditions (7.4.3) and (7.4.4), we additionally assume $\sup_\Omega |x| =: M < \infty$. Then,

$$\sup_{\mathcal{D}} \left\{ |xx^\tau h/\sqrt{n}\,| \right\} \le |h|/\sqrt{n} \sup_{\mathcal{D}} |x|^2 = |h|/\sqrt{n}\, M^2 = \mathrm{o}(1/\sqrt{n}\,) \tag{7.4.6}$$

and

$$\sup_{c\mathcal{D}}\left\{\frac{|xx^\tau h/\sqrt{n}\,|}{|A_\theta x(y-x^\tau\theta)|}\right\} \le M^2|h|/\sqrt{n}\,\sup_{c\mathcal{D}}\left\{\frac{1}{|A_\theta x(y-x^\tau\theta)|}\right\} \tag{7.4.7}$$

$$\le \frac{M^2|h|}{\sqrt{n}\,b_\theta} = \mathrm{o}(1/\sqrt{n}\,) \tag{7.4.8}$$

Remark 7.4.1 **(a)** Thus, in case of normal regression with bounded regressors, we can construct the optimally robust estimator as one-step estimator. However, one has to find a strict and – on $\mathcal{U}_c(\theta)$ – $\sqrt{n}$ consistent initial estimator. Unfortunately, minimum distance estimators like the Kolmogorov(–Smirnov) estimator are difficult to handle in the regression setup. Thus, there still is the need for an appropriate starting estimator. The minimum L_1 or the least trimmed square estimator could be possible candidates.

(b) By using $\tilde{x} = (x^\tau, 1)^\tau$ instead of x the above results at once apply to linear regression with intercept which is considered in Subsection 9.2.2.

(c) There might be two possible solutions to cover unbounded regressors, too. The first idea is to consider converging regressor distributions $K_n = K\big|_{\Omega_n}$ (normalized) where $\Omega_n = \{x \in \Omega\,|\,|x| \le M_n\}$ and $M_n = \mathrm{o}(\sqrt[4]{n})$; confer also Remark 7.1.1 (c). The second solution could be to use approximations of the optimal ICs as in Chapter 6.4 of Rieder (1994); confer also Subsection 2.3.1. ////

7.4.2 Normal Regression and Scale

We consider only the optimal AL estimators for the simultaneous estimation of regression and scale. Let $\theta = (\beta^\tau, \sigma)^\tau \in \mathbb{R}^k \times (0,\infty)$ and $\theta_n = \theta + h/\sqrt{n}$ for some $h \in \mathbb{R}^{k+1}$. By (7.1.4)–(7.1.6) we obtain

$$\mathcal{L}_{\theta_n}(\Lambda_{\theta_n}) = \frac{1}{\sigma_n}\,\mathcal{L}_{\theta_0}(\Lambda_{\theta_0}) \qquad \text{and} \qquad \mathcal{I}_{\theta_n} = \sigma_n^{-2}\mathcal{I}_{\theta_0} \tag{7.4.9}$$

i.e., (2.1.102) and (2.1.103) hold. As in Subsection 7.4.1 we assume ideal error distribution $F = \mathcal{N}(0,1)$ and $\sup_\Omega |x| = M < \infty$. We first verify condition (2.3.10); i.e., we have to show

$$\sup_{\mathcal{D}}\left\{\left|\begin{pmatrix} x\left[\frac{y-x^\tau\beta_n}{\sigma_n^2} - \frac{y-x^\tau\beta}{\sigma^2}\right] \\ \sigma_n^{-1}\left(\frac{y-x^\tau\beta_n}{\sigma_n}\right)^2 - \sigma^{-1}\left(\frac{y-x^\tau\beta}{\sigma}\right)^2 - \sigma_n^{-1} + \sigma^{-1} \end{pmatrix}\right|\right\} \longrightarrow 0 \tag{7.4.10}$$

where

$$\mathcal{D} = \left\{|A_{\theta_n}\Lambda_{\theta_n} - a_{\theta_n}| \le b_{\theta_n}\right\} \cup \left\{|A_\theta\Lambda_\theta - a_\theta| \le b_\theta\right\} \tag{7.4.11}$$

Since x is bounded, y is necessarily bounded on $\mathcal{D}$. We obtain

$$\begin{aligned}
&\sup_{\mathcal{D}} \left\{ \left| x \Big[\frac{y - x^\tau \beta_n}{\sigma_n^2} - \frac{y - x^\tau \beta}{\sigma^2} \Big] \right| \right\} \\
&\qquad \leq \sup_{\mathcal{D}} \left\{ |x| \Big[\big| y(\sigma_n^{-2} - \sigma^{-2}) \big| + \big| x x^\tau (\beta_n \sigma_n^{-2} - \beta \sigma^{-2}) \big| \Big] \right\} && (7.4.12) \\
&\qquad \leq M \big| \sigma_n^{-2} - \sigma^{-2} \big| \sup_{\mathcal{D}} |y| + M^2 \big| \beta_n \sigma_n^{-2} - \beta \sigma^{-2} \big| && (7.4.13) \\
&\qquad \longrightarrow 0 && (7.4.14)
\end{aligned}$$

as $n \to \infty$. Furthermore,

$$\begin{aligned}
&\sup_{\mathcal{D}} \left\{ \left| \sigma_n^{-1} \Big(\frac{y - x^\tau \beta_n}{\sigma_n} \Big)^2 - \sigma^{-1} \Big(\frac{y - x^\tau \beta}{\sigma} \Big)^2 - \sigma_n^{-1} + \sigma^{-1} \right| \right\} \\
&\qquad \leq \sup_{\mathcal{D}} \left\{ \left| \sigma_n^{-1} \Big(\frac{y - x^\tau \beta_n}{\sigma_n} \Big)^2 - \sigma^{-1} \Big(\frac{y - x^\tau \beta}{\sigma} \Big)^2 \right| \right\} + \big| \sigma_n^{-1} - \sigma^{-1} \big| && (7.4.15)
\end{aligned}$$

where the second summand on the RHS tends to zero as $n \to \infty$. By the boundedness of x and y on $\mathcal{D}$, we get for the first summand

$$\begin{aligned}
&\sup_{\mathcal{D}} \left\{ y^2 \big| \sigma_n^{-3} - \sigma^{-3} \big| + \big| 2 y x^\tau (\beta_n \sigma_n^{-2} - \beta \sigma^{-2}) \big| + \big| x^\tau (\beta_n \beta_n^\tau \sigma_n^{-3} - \beta \beta^\tau \sigma^{-3}) x \big| \right\} \\
&\qquad \leq \big| \sigma_n^{-3} - \sigma^{-3} \big| \sup_{\mathcal{D}} y^2 + 2M \Big| \frac{\beta_n}{\sigma_n^2} - \frac{\beta}{\sigma^2} \Big| \sup_{\mathcal{D}} |y| + M^2 \Big\| \frac{\beta_n \beta_n^\tau}{\sigma_n^3} - \frac{\beta \beta^\tau}{\sigma^3} \Big\|_{\mathrm{op}} && (7.4.16) \\
&\qquad \longrightarrow 0 && (7.4.17)
\end{aligned}$$

as $n \to \infty$. Since $|(u^\tau, v)^\tau| \leq |u| + |v|$ for all $u \in \mathbb{R}^k$ and all $v \in \mathbb{R}$, condition (2.3.10) holds. In the setup of normal linear regression and scale condition (2.3.11) reads

$$\sup_{{}^c\mathcal{D}} \left\{ \frac{\left| \begin{pmatrix} x \Big[\frac{y - x^\tau \beta_n}{\sigma_n^2} - \frac{y - x^\tau \beta}{\sigma^2} \Big] \\ \sigma_n^{-1} \Big(\frac{y - x^\tau \beta_n}{\sigma_n} \Big)^2 - \sigma^{-1} \Big(\frac{y - x^\tau \beta}{\sigma} \Big)^2 - \sigma_n^{-1} + \sigma^{-1} \end{pmatrix} \right|}{\left| \begin{pmatrix} A_{\mathrm{rg},\theta} x \frac{y - x^\tau \beta}{\sigma^2} \\ A_{\mathrm{sc},\theta} \sigma^{-1} \Big[\Big(\frac{y - x^\tau \beta}{\sigma} \Big)^2 - 1 - \sigma z_{\mathrm{sc},\theta} \Big] \end{pmatrix} \right|} \right\} \longrightarrow 0 \qquad (7.4.18)$$

We obtain,

$$\begin{aligned}
&\sup_{{}^c\mathcal{D}} \left\{ \frac{\left| x \Big[\frac{y - x^\tau \beta_n}{\sigma_n^2} - \frac{y - x^\tau \beta}{\sigma^2} \Big] \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\
&\qquad \leq \sup_{{}^c\mathcal{D}} \left\{ \frac{|x| \big| y(\sigma_n^{-2} - \sigma^{-2}) \big| + \big| x x^\tau (\beta_n \sigma_n^{-2} - \beta \sigma^{-2}) \big|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} && (7.4.19) \\
&\qquad \leq M \big| \sigma_n^{-2} - \sigma^{-2} \big| \sup_{{}^c\mathcal{D}} \left\{ \frac{|y|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\
&\qquad\quad + M^2 \big| \beta_n \sigma_n^{-2} - \beta \sigma^{-2} \big| \sup_{{}^c\mathcal{D}} \left\{ \frac{1}{|A_\theta \Lambda_\theta - a_\theta|} \right\} && (7.4.20)
\end{aligned}$$

where the second summand on the RHS tends to zero as $|A_\theta \Lambda_\theta - a_\theta| \geq b_\theta > 0$ on ${}^c\mathcal{D}$. Moreover, on ${}^c\mathcal{D} \cap \{|y| \leq M_1\}$ with $M_1 \in (0, \infty)$, the first summand on the RHS tends to zero, too. Now, consider $\mathcal{E} := {}^c\mathcal{D} \cap \{|y| > M_1\}$ where we choose M_1 such that

$$M_1 - 2\sigma \sup_{{}^c\mathcal{D}} |x^\tau \beta| - \sigma^2 \frac{\sigma z_{\mathrm{sc},\theta} + 1}{M_1} \geq \delta > 0 \tag{7.4.21}$$

Then,

$$\sup_{\mathcal{E}} \left\{ \frac{|y|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \leq \sup_{\mathcal{E}} \left\{ \frac{\sigma^3 |y|}{A_{\mathrm{sc},\theta} |y^2 - 2\sigma y x^\tau \beta + (x^\tau \beta)^2 - \sigma^2 - \sigma^3 z_{\mathrm{sc},\theta}|} \right\} \tag{7.4.22}$$

$$\leq \frac{\sigma^3}{A_{\mathrm{sc},\theta}} \sup_{\mathcal{E}} \left\{ \frac{1}{\big||y| - 2\sigma \operatorname{sign}(y) x^\tau \beta - \sigma^2(\sigma z_{\mathrm{sc},\theta} + 1)/|y|\big|} \right\} \tag{7.4.23}$$

$$\leq \frac{\sigma^3}{A_{\mathrm{sc}} \delta} < \infty \tag{7.4.24}$$

Consequently,

$$\sup_{{}^c\mathcal{D}} \left\{ \frac{\left| x \left[\frac{y - x^\tau \beta_n}{\sigma_n^2} - \frac{y - x^\tau \beta}{\sigma^2} \right] \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \longrightarrow 0 \tag{7.4.25}$$

as $n \to \infty$. Furthermore, we obtain

$$\begin{aligned} &\sup_{{}^c\mathcal{D}} \left\{ \frac{\left| \sigma_n^{-1} \left(\frac{y - x^\tau \beta_n}{\sigma_n} \right)^2 - \sigma^{-1} \left(\frac{y - x^\tau \beta}{\sigma} \right)^2 - \sigma_n^{-1} + \sigma^{-1} \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\ &\qquad \leq \sup_{{}^c\mathcal{D}} \left\{ \frac{\left| \sigma_n^{-1} \left(\frac{y - x^\tau \beta_n}{\sigma_n} \right)^2 - \sigma^{-1} \left(\frac{y - x^\tau \beta}{\sigma} \right)^2 \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\ &\qquad\quad + |\sigma_n^{-1} - \sigma^{-1}| \sup_{{}^c\mathcal{D}} \left\{ \frac{1}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \end{aligned} \tag{7.4.26}$$

where the second summand on the RHS tends to zero as $|A_\theta \Lambda_\theta - a_\theta| \geq b_\theta > 0$ on ${}^c\mathcal{D}$. The first summand can be rewritten as

$$\begin{aligned} &\sup_{{}^c\mathcal{D}} \left\{ \frac{y^2 \left| \sigma_n^{-3} - \sigma^{-3} \right| + \left| 2 y x^\tau (\beta_n \sigma_n^{-2} - \beta \sigma^{-2}) \right| + \left| x^\tau (\beta_n \beta_n^\tau \sigma_n^{-3} - \beta \beta^\tau \sigma^{-3}) x \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\ &\quad \leq \left| \sigma_n^{-3} - \sigma^{-3} \right| \sup_{{}^c\mathcal{D}} \left\{ \frac{y^2}{|A_\theta \Lambda_\theta - a_\theta|} \right\} + 2M \left| \frac{\beta_n}{\sigma_n^2} - \frac{\beta}{\sigma^2} \right| \sup_{{}^c\mathcal{D}} \left\{ \frac{|y|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \\ &\qquad + M^2 \left\| \frac{\beta_n \beta_n^\tau}{\sigma_n^3} - \frac{\beta \beta^\tau}{\sigma^3} \right\|_{\mathrm{op}} \sup_{{}^c\mathcal{D}} \left\{ \frac{1}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \end{aligned} \tag{7.4.27}$$

where the second and third summand on the RHS tend to zero by analogous arguments as given above. In addition, on ${}^c\mathcal{D} \cap \{|y| \leq M_2\}$ with $M_2 \in (0, \infty)$, the first

summand on the RHS tends to zero. Now, consider $\mathcal{E} := {}^{c}\mathcal{D} \cap \{|y| > M_2\}$ where we choose M_2 such that

$$1 - 2\sigma \sup_{{}^{c}\mathcal{D}} |x^\tau \beta|/M_2 - \sigma^2(\sigma z_{\mathrm{sc},\theta} + 1)/M_2^2 > \delta > 0 \tag{7.4.28}$$

Then,

$$\sup_{\mathcal{E}} \left\{ \frac{y^2}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \leq \sup_{\mathcal{E}} \left\{ \frac{\sigma^3 y^2}{A_{\mathrm{sc},\theta} |y^2 - 2\sigma y x^\tau \beta + (x^\tau \beta)^2 - \sigma^2 - \sigma^3 z_{\mathrm{sc},\theta}|} \right\} \tag{7.4.29}$$

$$\leq \frac{\sigma^3}{A_{\mathrm{sc}}} \sup_{\mathcal{E}} \left\{ \frac{1}{|1 - 2\sigma x^\tau \beta/|y| - \sigma^2(\sigma z_{\mathrm{sc},\theta} + 1)/y^2|} \right\} \tag{7.4.30}$$

$$\leq \frac{\sigma^3}{A_{\mathrm{sc}} \delta} < \infty \tag{7.4.31}$$

Consequently,

$$\sup_{{}^{c}\mathcal{D}} \left\{ \frac{\left| \sigma_n^{-1} \left(\frac{y - x^\tau \beta_n}{\sigma_n} \right)^2 - \sigma^{-1} \left(\frac{y - x^\tau \beta}{\sigma} \right)^2 - \sigma_n^{-1} + \sigma^{-1} \right|}{|A_\theta \Lambda_\theta - a_\theta|} \right\} \longrightarrow 0 \tag{7.4.32}$$

as $n \to \infty$. Thus, condition (2.3.11) holds and we can construct the optimally robust estimator as one-step estimator. For some additional comments on unbounded regressors, respectively strict and – on $\mathcal{U}_c(\theta)$ – $\sqrt{n}$ consistent initial estimators we refer to Remark 7.4.1.

7.5 Numerical Results

7.5.1 Simultaneous Estimation

We compare AL, M, MK, ALc, Mc, and BM estimators for various simple, one-dimensional, discrete regressor distributions K where we choose ideal error distribution $F = \mathcal{N}(0,1)$.

In all considered situations the differences between AL and M estimators are only small, respectively very small. The maximum efficiency loss we encounter is 9.8% (cf. Table 7.7) but in almost all cases it is clearly smaller. If K is uniform, the MK estimators perform well, too. However, there are situations where this is definitely not true; confer in particular Tables 7.6 and 7.7. Moreover, in almost all situations the results of the MK estimators and Mc estimators show very small differences. The only exception is $K = p \sum_{k=1}^{5} k^{-2} \mathrm{I}_{\{k\}}$ ($p \approx 0.683$); confer Tables 7.12 and 7.13. While the general M estimators seem to perform almost as well as the optimal AL estimators (apart from the computational effort), this does not hold in case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$), where the Mc estimators may have a very large subefficiency ($> 300\%$) with respect to the optimal ALc estimators; confer Tables 7.6 and 7.7.

Since $\Psi_{2\bullet}(\theta) \subset \Psi_2(\theta)$, we expect some subefficiency for the optimal estimators in

case of average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$) with respect to the optimal estimators in case of unconditional contamination neighborhoods ($* = c,\ t = 0$). However, in case of AL and ALc estimators the minimax asymptotic MSEs are very similar in all considered situations. In contrast, we get clear differences in case of M and Mc estimators especially for $K = 25/26 \cdot \mathrm{I}_{\{1\}} + 1/26 \cdot \mathrm{I}_{\{5\}}$ where the maximum subefficiency is about 270%; see Table 7.6.
Definitely, the BM estimators perform worst. These estimators lose up to 70% efficiency with respect to the Mc estimators. Moreover, the maximum subefficiency with respect to the ALc estimators is even about 425%.

Remark 7.5.1 **(a)** In case of BM estimators and regressors with m design points the computation is based on a $(m+1)$-dimensional numerical optimization; confer Remark 7.3.8 (d). Thus, in case of the 5-points regressor, the results already have a certain variability depending on the starting values. Hence, we give only two decimal places in these situations.

(b) As M, respectively Mc estimators in all respects (theoretical, numerical, ...) turn out inferior to AL, respectively ALc estimators in the local setup, and global properties (breakdown, ...) could be preserved by a suitable one-step construction, the M principle seems statistically unjustified in this context and we strongly recommend to use the optimally robust AL and ALc estimators, respectively. ////

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	0.678	0.678	0.678	0.678	0.678	0.682
0.25	0.902	0.903	0.903	0.902	0.903	0.924
0.5	1.382	1.383	1.384	1.382	1.384	1.455
1.0	2.704	2.715	2.715	2.705	2.715	3.028
1.5	4.574	4.639	4.640	4.575	4.640	5.520
2.0	7.099	7.251	7.251	7.100	7.251	8.999

Table 7.2: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = \mathrm{Unif}(\{1.0, 5.0\})$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.000	1.000	1.000	1.006
0.25	1.001	1.000	1.001	1.023
0.5	1.001	1.001	1.001	1.051
1.0	1.004	1.000	1.004	1.115
1.5	1.014	1.000	1.014	1.190
2.0	1.021	1.000	1.021	1.241

Table 7.3: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = \mathrm{Unif}\left(\{1.0, 5.0\}\right)$.

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	0.810	0.810	0.810	0.810	0.810	0.838
0.25	1.059	1.060	1.061	1.060	1.061	1.175
0.5	1.596	1.600	1.607	1.598	1.607	2.009
1.0	3.105	3.139	3.232	3.112	3.232	4.900
1.5	5.246	5.417	5.726	5.251	5.726	9.645
2.0	8.116	8.496	9.098	8.121	9.098	16.288

Table 7.4: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = 5/6 \cdot \mathrm{I}_{\{1\}} + 1/6 \cdot \mathrm{I}_{\{5\}}$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.000	1.000	1.000	1.035
0.25	1.001	1.001	1.001	1.107
0.5	1.003	1.004	1.006	1.250
1.0	1.011	1.030	1.039	1.516
1.5	1.033	1.057	1.090	1.684
2.0	1.047	1.071	1.120	1.790

Table 7.5: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = 5/6 \cdot \mathrm{I}_{\{1\}} + 1/6 \cdot \mathrm{I}_{\{5\}}$.

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	1.186	1.187	1.213	1.187	1.213	1.295
0.25	1.544	1.550	1.842	1.546	1.842	2.152
0.5	2.277	2.310	3.904	2.280	3.904	4.848
1.0	4.289	4.485	12.029	4.291	12.029	15.387
1.5	7.115	7.682	25.503	7.117	25.503	32.908
2.0	10.947	12.018	44.314	10.949	44.314	57.433

Table 7.6: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = 25/26 \cdot \mathrm{I}_{\{1\}} + 1/26 \cdot \mathrm{I}_{\{5\}}$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.001	1.022	1.022	1.068
0.25	1.004	1.188	1.191	1.168
0.5	1.014	1.690	1.712	1.242
1.0	1.046	2.682	2.803	1.279
1.5	1.080	3.320	3.583	1.290
2.0	1.098	3.687	4.047	1.296

Table 7.7: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = 25/26 \cdot \mathrm{I}_{\{1\}} + 1/26 \cdot \mathrm{I}_{\{5\}}$.

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	0.692	0.692	0.692	0.692	0.692	0.71
0.25	0.919	0.919	0.919	0.919	0.919	0.95
0.5	1.403	1.405	1.405	1.403	1.405	1.50
1.0	2.738	2.751	2.751	2.739	2.751	3.21
1.5	4.623	4.699	4.700	4.624	4.700	5.90
2.0	7.167	7.345	7.345	7.168	7.345	9.62

Table 7.8: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = \mathrm{Unif}\left(\{1.0, 2.0, 3.0, 4.0, 5.0\}\right)$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.000	1.000	1.000	1.03
0.25	1.000	1.000	1.000	1.03
0.5	1.001	1.000	1.001	1.06
1.0	1.005	1.000	1.004	1.17
1.5	1.016	1.000	1.016	1.26
2.0	1.025	1.000	1.025	1.31

Table 7.9: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = \mathrm{Unif}\,(\{1.0, 2.0, 3.0, 4.0, 5.0\})$.

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	0.757	0.757	0.757	0.757	0.757	0.77
0.25	0.994	0.995	0.995	0.994	0.995	1.05
0.5	1.501	1.504	1.505	1.502	1.505	1.74
1.0	2.902	2.919	2.921	2.905	2.923	4.06
1.5	4.877	4.971	4.971	4.880	4.973	7.74
2.0	7.553	7.762	7.762	7.556	7.764	12.96

Table 7.10: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = p \sum_{k=1}^{5} k^{-1}\, \mathrm{I}_{\{k\}}$ $(p \approx 0.438)$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.000	1.000	1.000	1.01
0.25	1.001	1.000	1.001	1.05
0.5	1.002	1.001	1.002	1.15
1.0	1.006	1.001	1.006	1.07
1.5	1.019	1.000	1.019	1.39
2.0	1.028	1.000	1.028	1.67

Table 7.11: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = p \sum_{k=1}^{5} k^{-1}\, \mathrm{I}_{\{k\}}$ $(p \approx 0.438)$.

	$* = c,\ t = 0$			$* = c,\ t = \alpha = 1$		
r	AL	M	MK	ALc	Mc	BM
0.1	0.909	0.909	0.909	0.909	0.910	0.95
0.25	1.174	1.176	1.180	1.175	1.185	1.40
0.5	1.747	1.754	1.777	1.749	1.901	2.61
1.0	3.346	3.401	3.556	3.352	4.652	7.07
1.5	5.599	5.804	6.197	5.603	9.155	14.45
2.0	8.631	9.064	9.747	8.635	15.389	24.80

Table 7.12: Minimax asymptotic MSE for AL, M, MK, ALc, Mc and BM estimators in case $K = p \sum_{k=1}^{5} k^{-2}\, \mathrm{I}_{\{k\}}$ $(p \approx 0.683)$.

	$* = c,\ t = 0$		$* = c,\ t = \alpha = 1$	
r	M vs. AL	MK vs. M	Mc vs. ALc	BM vs. Mc
0.1	1.000	1.000	1.001	1.04
0.25	1.002	1.003	1.009	1.18
0.5	1.004	1.013	1.087	1.37
1.0	1.016	1.046	1.388	1.52
1.5	1.037	1.068	1.634	1.58
2.0	1.050	1.075	1.782	1.61

Table 7.13: MSE–inefficiency for AL, M, MK, ALc, Mc and BM estimators in case $K = p \sum_{k=1}^{5} k^{-2}\, \mathrm{I}_{\{k\}}$ $(p \approx 0.683)$.

7.5.2 Separate Estimation

We compare ALs, Ms and BM estimators. We choose the same simple, one-dimensional, discrete regressor distributions K as in the previous subsection and again consider only ideal error distribution $F = \mathcal{N}(0,1)$. In this context of separate estimation of regression and scale the optimal AL estimators are called ALs estimators. By symmetry of F, the AL solutions in case of unconditional ($* = c$, $t = 0$) and average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) coincide; confer Proposition 7.3.2.
In the exemplary situations considered, the ALs estimators perform clearly better than Ms and BM estimators, respectively. In case of the Ms estimators the maximum subefficiencies vary between about 12% and 315% where the differences are smallest for regressors with uniform distributions. The efficiency loss of BM relative to Ms is moderate and stays below about 15% in all examples considered. However, as noted in Remark 7.3.8 (d), the BM estimators are much harder to compute than the Ms estimators.

	minimax MSE			MSE–inefficiency	
r	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	0.681	0.682	0.682	1.001	1.000
0.25	0.914	0.923	0.924	1.010	1.001
0.5	1.419	1.453	1.455	1.024	1.001
1.0	2.830	3.002	3.028	1.061	1.009
1.5	4.851	5.335	5.520	1.100	1.035
2.0	7.587	8.472	8.999	1.117	1.062

Table 7.14: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = \mathrm{Unif}(\{1.0, 5.0\})$.

	minimax MSE			MSE–inefficiency	
r	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	0.831	0.838	0.838	1.008	1.000
0.25	1.138	1.175	1.175	1.033	1.000
0.5	1.787	1.985	2.009	1.111	1.012
1.0	3.553	4.464	4.900	1.256	1.098
1.5	6.087	8.759	9.645	1.439	1.101
2.0	9.520	14.323	16.288	1.505	1.137

Table 7.15: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = 5/6 \cdot \mathrm{I}_{\{1\}} + 1/6 \cdot \mathrm{I}_{\{5\}}$.

	minimax MSE			MSE–inefficiency	
r	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	1.245	1.295	1.295	1.040	1.000
0.25	1.658	2.150	2.152	1.297	1.001
0.5	2.457	4.750	4.848	1.933	1.021
1.0	4.726	14.483	15.387	3.065	1.062
1.5	8.051	30.263	32.908	3.759	1.087
2.0	12.571	52.073	57.433	4.142	1.103

Table 7.16: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = 25/26 \cdot \mathrm{I}_{\{1\}} + 1/26 \cdot \mathrm{I}_{\{5\}}$.

	minimax MSE			MSE–inefficiency	
r	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	0.695	0.699	0.71	1.006	1.02
0.25	0.930	0.946	0.95	1.017	1.00
0.5	1.433	1.489	1.50	1.039	1.01
1.0	2.839	3.083	3.21	1.086	1.04
1.5	4.856	5.464	5.90	1.125	1.08
2.0	7.590	8.664	9.62	1.142	1.11

Table 7.17: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = \mathrm{Unif}\left(\{1.0, 2.0, 3.0, 4.0, 5.0\}\right)$.

	minimax MSE			MSE–inefficiency	
r	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	0.766	0.773	0.77	1.009	1.00
0.25	1.023	1.055	1.05	1.031	1.00
0.5	1.576	1.712	1.74	1.086	1.02
1.0	3.119	3.798	4.06	1.218	1.07
1.5	5.338	6.980	7.74	1.308	1.11
2.0	8.346	11.265	12.96	1.350	1.15

Table 7.18: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = p \sum_{k=1}^{5} k^{-1} \mathrm{I}_{\{k\}}$ $(p \approx 0.438)$.

r	minimax MSE			MSE–inefficiency	
	ALs	Ms	BM	Ms vs. ALs	BM vs. Ms
0.1	0.932	0.949	0.95	1.018	1.00
0.25	1.242	1.363	1.40	1.097	1.03
0.5	1.894	2.494	2.61	1.317	1.05
1.0	3.714	6.537	7.07	1.760	1.08
1.5	6.347	12.977	14.45	2.045	1.11
2.0	9.920	21.791	24.80	2.197	1.14

Table 7.19: Minimax asymptotic MSE and MSE–inefficiency for ALs, Ms and BM estimators in case $K = p\sum_{k=1}^{5} k^{-2}\,\mathrm{I}_{\{k\}}$ ($p \approx 0.683$).

7.6 Implementation using R

7.6.1 R Package ROptRegTS

The optimal AL estimators for the simultaneous estimation of regression and scale can be computed via our R package ROptRegTS which is in detail described in Appendix D.4. In case of normal regression and scale, we provide the generating function NormLinRegScaleFamily which generates an object of class L2RegTypeFamily. Given some regressor distribution K (an object of class Distribution) one can use this function to easily specify a normal linear regression and scale family with parameters $\beta \in \mathbb{R}^k$ and $\sigma \in (0, \infty)$ via

```
> LM <- NormLinRegScaleFamily(theta = β, scale = σ,
+                             RegDistr = K)
```

One can also specify a transformation $D \in \mathbb{R}^{p\times(k+1)}$ ($p \le k+1$) of the parameter $\theta = (\beta^\tau, \sigma)^\tau$ by adding `trafo =` D to the call of NormLinRegScaleFamily. The classical optimal (partial) IC corresponding to LM can then be computed using

```
> IC0 <- optIC(model = LM, risk = asCov())
```

The result IC0 is an object of class IC which is part of our R package ROptEst; confer Appendix D.3. For the computation of optimally robust ICs we provide the class InfRobRegTypeModel which in addition to an object of class L2RegTypeFamily includes an unconditional or conditional infinitesimal neighborhood (object of class Neighborhood). The call

```
> RobLM <-  InfRobRegTypeModel(center = LM,
+                 neighbor = ContNeighborhood(radius = r))
```

respectively

```
> RobLMc <-  InfRobRegTypeModel(center = LM,
+                  neighbor=Av1CondContNeighborhood(radius = r))
```

generates an instance RobLM, respectively RobLMc of a normal regression and scale family with unconditional ($* = c,\ t = 0$), respectively average conditional contamination neighborhood ($* = c,\ t = \alpha = 1$) and radius $r \in [0, \infty]$. In case of conditional neighborhoods (objects of class CondNeighborhood) one can additionally specify a radius curve; i.e., given some R function fun one can add radiusCurve = fun to the call of corresponding generating function where fun has to be a function of one argument named x.
The optimally robust IC with respect to the maximum asymptotic MSE (simultaneous estimation of regression and scale) can then be computed via

```
> IC1 <- optIC(model = RobLM, risk = asMSE())
```

respectively

```
> IC1c <- optIC(model = RobLMc, risk = asMSE())
```

That is, an object of class ContIC, respectively Av1CondContIC is returned. In case $r = 0$, the classical optimal IC and in case $r = \infty$, the minimum bias solution are computed, respectively.

Remark 7.6.1 In case of regression and scale, the minimum bias solutions are, so far, only available for unconditional contamination neighborhoods ($* = c,\ t = 0$).
////

The minimum bias solution can also be obtained by calling

```
> IC2 <- optIC(model = RobLM, risk = asBias())
```

If the radius is unknown, respectively unknown except to belong to some interval $[a, b]$ with $a \in [0, \infty)$ and $b \in (a, \infty]$, one can call

```
> IC3 <- radiusMinimaxIC(L2Fam = LM, neighbor=ContNeighborhood(),
+                        risk = asMSE(), loRad = a, upRad = b)
```

respectively

```
> IC3c <- radiusMinimaxIC(L2Fam = LM,
+                         neighbor = Av1CondContNeighborhood(),
+                         risk = asMSE(), loRad = a, upRad = b)
```

which computes the IC of the AL, respectively ALc estimator which is radius–minimax in sense of Section 2.2. Finally, one can determine the least favorable radius r_ρ ($\rho \in (0, 1)$) and the corresponding MSE–inefficiency using

```
> r.rho1 <- leastFavorableRadius(L2Fam = LM,
+                                neighbor = ContNeighborhood(),
+                                risk = asMSE(), rho = ρ)
```

respectively

```
> r.rho2 <- leastFavorableRadius(L2Fam = LM,
+                       neighbor = Av1CondContNeighborhood(),
+                       risk = asMSE(), rho = ρ)
```

which returns a `list` with members `rho`, `leastFavorableRadius` and `ineff`.

Remark 7.6.2 (a) In the examples considered in Section 7.5 the computations take about 60 in case of unconditional ($* = c,\ t = 0$) as well as average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$) on an AMD Athlon with 2.5 GHz and 512 MB RAM using `R 2.0.1` (cf. R Development Core Team (2005)). Already in these simple examples the determination of the radius minimax IC therefore lasts about 15–30 minutes for the AL as well as the ALc estimators. Consequentially, the computation of the least favorable radius takes several hours (about 4). In case of regressors with much more design points, respectively for absolutely continuous univariate regressors the computation of the radius–minimax IC may already last several hours. Hence, the determination of the least favorable radius in such cases probably takes 1–2 days.

(b) The methods for the generic function `optIC` work not only for regression and scale with normal, respectively symmetric distributed errors but for all error distributions with (F2) and (F3). Moreover, they even work for all L_2 differentiable regression-type families as introduced in Appendix A.1. These models cover regression (with unknown scale and/or intercept) as well as certain time series models (with unknown scale and/or shift). Some further examples are given in Subsection 9.2.2 and Section 9.3, respectively. ////

Having some (contaminated) data `X` of class `matrix` and some robust initial estimate `est0`, we can perform a robust estimation by means of the one-step method. If we are sure about the neighborhood radius $r \in (0, \infty)$, we can use

```
> LM1 <- NormLinRegScaleFamily(RegDistr = K, theta = est0[1:k],
+                              scale = est0[k+1])
> RobLM1 <-InfRobRegTypeModel(center = LM,
+                      neighbor = ContNeighborhood(radius = r))
> IC4 <- optIC(model = RobLM1, risk = asMSE())
> est1 <- oneStepEstimator(x = X, IC = IC4, start = est0)
```

respectively

```
> RobLM1c <-InfRobRegTypeModel(center = LM,
+                 neighbor = Av1CondContNeighborhood(radius = r))
> IC4c <- optIC(model = RobLM1c, risk = asMSE())
> est1c <- oneStepEstimator(x = X, IC = IC4c, start = est0)
```

However, if the neighborhood radius is unknown except to lie in some interval $[a, b]$ ($a \in [0, \infty)$, $b \in (a, \infty]$), we instead can proceed as follows

```
> IC5 <- radiusMinimaxIC(L2Fam = LM,
+                        neighbor = ContNeighborhood(),
+                        risk = asMSE(), loRad = a, upRad = b)
```

```
> est2 <- oneStepEstimator(x = X, IC = IC5, start = est0)
```

respectively

```
> IC5c <- radiusMinimaxIC(L2Fam = LM,
+                         neighbor = Av1CondContNeighborhood(),
+                         risk = asMSE(), loRad = a, upRad = b)
> est2c <- oneStepEstimator(x = X, IC = IC5c, start = est0)
```

Remark 7.6.3 After installing our R bundle RobASt one can find the R script NormLinRegScale.R, which contains some examples for the normal regression and scale model, in the directory "…/RHome/library/ROptRegTS/scripts/" where RHome stands for the local home directory of R. ////

7.6.2 R Package RobRex

The optimally robust ICs for the AL and ALc estimators in case of normal regression and scale can also be computed via our R package RobRex which in addition provides functions for the computation of the optimal ICs for all estimators introduced throughout this chapter. We implemented functions of the form rgsOptIC.Est where Est has to be replaced by the short form of the corresponding estimators; i.e., Est = AL, ALc, M, MK, Mc, ALs, Ms, BM. Given some neighborhood radius $r \in (0, \infty)$ and some regressor distribution K (object of class Distribution), these functions are called via

```
> IC1 <- rgsOptIC.Est(r = r, K = K)
```

Remark 7.6.4 In case of AL, M, MK and ALs estimators K may be an arbitrary distribution with (K). But, in case of the other classes of estimators there are certain additional limitations concerning K. The function rgsOptIC.ALc works for all discrete distributions and the functions rgsOptIC.Est with Est = Mc, Ms, BM are only implemented for all discrete univariate distributions. ////

One can also specify additional parameters in the calls for these functions but at least for small radii the default values work fine. For more details about these further parameters we refer to the help pages of package RobRex. The rgsOptIC.Est functions return an object of class ContIC in case of the AL estimators, respectively an object of class Av1CondContIC in case of the ALc estimators. In case of the other estimators objects of class IC, respectively CondIC are returned.

Remark 7.6.5 (a) In the examples considered in Section 7.5 the computations of the optimal robust ICs in case of the AL and ALc estimators using the functions rgsOptIC.AL and rgsOptIC.ALc, respectively, takes about 5 seconds on an AMD Athlon with 2.5 GHz and 512 MB RAM using R 2.0.1 (cf. R Development Core Team (2005)). Hence, is about 10–15 times faster than via optIC of package ROptRegTS; confer Remark 7.6.2 (a). Hence, we can say, this is the price we have to pay for the greater generality that optIC works for not only for regression and

scale with normal errors but for all error distributions with (F2) and (F3) and even for various other models; confer Remark 7.6.2 (b).

(b) Since the functions `rgsOptIC.Est` return an object of class `IC`, respectively `CondIC`, all generic functions of the packages `ROptEst` and `ROptRegTS` with `IC`, respectively `CondIC` in its signature can be applied to the corresponding results. ////

The algorithm for the computation of the optimally robust ICs in case of the AL, ALc and ALs estimators is analogous to Algorithm D.3.1. In case of the M, MK, Mc and Ms estimators we use iteration algorithms which are constructed analogously to the AL, respectively ALc case. However, there are additional outer optimizations in A and γ, respectively γ necessary which are done using the `R` function `optim` with `method = "Nelder-Mead"`, respectively using the `R` function `optimize`; confer R Development Core Team (2005).

Remark 7.6.6 Because of the mentioned additional outer optimizations in case of M, MK, Mc and Ms estimators, the computation of the optimal influence curves may take quite a long time for these estimators and is in addition numerically less stable than in case of the AL, ALc and ALs estimators. ////

In case of the BM estimators α_x and γ_x can be calculated by the equations (7.3.51) and (Mc4). To find the optimal clipping bounds b_{rg} and $b_{\mathrm{sc},0,x}$, we use the `R` function `optim` with `method = "Nelder-Mead"`; confer R Development Core Team (2005). To control the constraints on b_{rg} and $b_{\mathrm{sc},0,x}$, we introduce an additional parameter `MAX`. That is, if `optim` uses values beyond the admitted ones, the function which is minimized returns `MAX`. Thus, `MAX` should be (clearly) larger than the expected minimax asymptotic MSE to obtain valid results. The default `MAX =` 100 works well in many cases.

Remark 7.6.7 There are example sections in the help files to our `R` package `RobRex` which demonstrate the use of these functions. The corresponding examples can be executed via

```
> example(rgsOptIC.Est)
```

where `Est = AL, ALc, M, MK, Mc, ALs, Ms, BM`. ////

Chapter 8

Normal Location and Scale – a Comparative Study

The purpose of this chapter is a comparison of several well-known location and scale estimators proposed in literature. All estimators studied are asymptotically linear; i.e., belong to the class of AL estimators introduced in Section 1.1. Hence, we expect that they have a larger maximum asymptotic risk than the optimal AL estimators specified in Section 8.2. However, it is not clear how much efficiency they lose. In Section 8.1 we briefly introduce the setup and define the notion of absolute and relative information. We then present the ICs of the optimal AL, M and BM estimators as well as the ICs of several estimators proposed in literature; confer Sections 8.2–8.5. Subsequently, we compare these estimators with respect to their (numerical) maximum asymptotic MSEs; confer Section 8.7. In case of the AL estimators we additionally verify that we can construct the optimally robust estimators by means of one-step constructions; confer Section 8.8. We conclude this chapter with a description of our `R` package `RobLox` which includes routines for the computation of the optimal ICs for all considered estimators; confer Section 8.9.

8.1 Setup

We assume the normal location and scale model

$$y_i = \mu + \sigma u_i \qquad (i = 1, \ldots, n) \tag{8.1.1}$$

where $u_1, \ldots, u_n$ are i.i.d. copies of the error $u \sim \mathcal{N}(0,1)$. The L_2 derivative and Fisher information of location and scale at $\theta_0 = (0,1)^\tau$ are

$$\Lambda_{\theta_0} = \begin{pmatrix} u \\ u^2 - 1 \end{pmatrix} \qquad \mathcal{I}_{\theta_0} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \tag{8.1.2}$$

Moreover, we consider infinitesimal contamination neighborhoods of starting radius $r \in [0, \infty)$ which at sample size n consist of all

$$Q_{n,\theta} = (1 - r/\sqrt{n})\mathcal{N}(\mu, \sigma^2) + r/\sqrt{n}\, H_{n,\theta} \tag{8.1.3}$$

where $H_{n,\theta}$ may be arbitrary unknown probabilities; confer Section 1.2.

Remark 8.1.1 **(a)** We consider the simultaneous estimation of location and scale. Whereas, in many papers scale is often considered as a nuisance parameter and some of the considered estimators are rather proposed for this situation.

(b) Of course, the results of regression and scale derived in Chapter 7 apply to the normal location and scale model. Thus, the optimal AL, M and BM estimators can be read off from the corresponding results derived in Section 7.2.

(c) It might be interesting, to also investigate the finite-sample behavior of the considered estimators in a large simulation study in combination with numerical calculations using FFT as done in Part V, respectively in Ruckdeschel and Kohl (2005) wherever this is possible.

(d) There are far too many estimators for normal location and scale to consider all of them in the framework of this thesis. A large study about finite-sample and asymptotic behavior of 68 estimators was made by Andrews et al. (1972). Another extensive study by Stigler (1977) compares the performance of 11 estimators using real data sets. ////

To further analyze the different estimators, we introduce the notion of absolute and relative information.

Definition 8.1.2 *Let* $\eta \in \Psi_2$ *of the form* $\eta = (\eta_{\text{loc}}, \eta_{\text{sc}})^\tau$.

(a) *The* absolute information *of* η *is* $|\eta|^2$.

(b) *The* relative information of scale *of* η *is* $\eta_{\text{sc}}^2/|\eta|^2$.

Remark 8.1.3 **(a)** The absolute information gives an impression of the amount of information which any asymptotically linear estimator attaches to a certain observation. In addition, the relative information shows how this information is divided up between location and scale; i.e., how much is used for the estimation of scale, respectively for the estimation of location ($1 - \eta_{\text{sc}}^2/|\eta|^2$). The idea for these notions, which lead to some cute interpretations (cf. Sections 8.2–8.5), arose from a discussion with H. Rieder and P. Ruckdeschel.

(b) Obviously, the notion of absolute and relative information can easily be extended to other models. ////

8.2 AL Estimators

The unique MSE optimal IC $\tilde{\eta}_r$ in case of infinitesimal contamination neighborhoods (1.2.4) and radius $r \in (0, \infty)$ provided by Proposition 7.2.1 can be rewritten as

$$\tilde{\eta}_r(u) = \begin{pmatrix} \tilde{\eta}_{\text{loc}}(u) \\ \tilde{\eta}_{\text{sc}}(u) \end{pmatrix} = \begin{pmatrix} A_{\text{loc}} u \\ A_{\text{sc}}(u^2 - z_{\text{sc}}) \end{pmatrix} w(u) \tag{8.2.1}$$

with

$$w(u) = \min\left\{1, \frac{b}{\left[A_{\mathrm{loc}}^2 u^2 + A_{\mathrm{sc}}^2 (u^2 - z_{\mathrm{sc}})^2\right]^{1/2}}\right\} \tag{8.2.2}$$

where

$$A_{\mathrm{loc}} = \left[\mathrm{E}\, u^2 w\right]^{-1} \qquad A_{\mathrm{sc}} = \left[\mathrm{E}(u^2 - z_{\mathrm{sc}})^2 w\right]^{-1} \qquad z_{\mathrm{sc}} = \mathrm{E}\, u^2 w \left[\mathrm{E}\, w\right]^{-1} \tag{8.2.3}$$

and

$$r^2 b = \mathrm{E}\left(\left[A_{\mathrm{loc}}^2 u^2 + A_{\mathrm{sc}}^2 (u^2 - z_{\mathrm{sc}})^2\right]^{1/2} - b\right)_+ \tag{8.2.4}$$

For $r = \infty$, we obtain by Proposition 7.2.4 (c)

$$\tilde{\eta}_\infty(u) = \omega_c^{\min} \begin{pmatrix} u \\ \bar{\alpha}(u^2 - \bar{z}_{\mathrm{sc}}) \end{pmatrix} \bar{Y}^{-1} \tag{8.2.5}$$

with

$$\bar{Y}(x, u) := \left[u^2 + \bar{\alpha}^2 (u^2 - \bar{z}_{\mathrm{sc}})^2\right]^{1/2} \tag{8.2.6}$$

$$\bar{\alpha} = \mathrm{E}\, u^2 \bar{Y}^{-1} \left[\mathrm{E}(u^2 - \bar{z}_{\mathrm{sc}})^2 \bar{Y}^{-1}\right]^{-1} \qquad \bar{z}_{\mathrm{sc}} = \mathrm{E}\, u^2 \bar{Y}^{-1} \left[\mathrm{E}\, \bar{Y}^{-1}\right]^{-1} \tag{8.2.7}$$

and $\tilde{\eta}_\infty$ attains the minimum bias

$$\omega_c^{\min} = (1 + \bar{\alpha}) / \mathrm{E}\, \bar{Y} \tag{8.2.8}$$

For a plot of the location and scale part of the optimally robust IC for different values of r see Figure 8.1.

Remark 8.2.1 (a) We determined $\bar{a}_{\mathrm{sc}}$, $\bar{\alpha}$ and $\omega_c^{\min}$ numerically. The results are $\bar{a}_{\mathrm{sc}} \approx 0.372$, $\bar{\alpha} \approx 0.792$ and $\omega_c^{\min} \approx 1.618$. In addition, our numerical computations show, that for radius $r \geq 1.661$ the optimal clipping bound b as well as a_{sc}, respectively $A_{\mathrm{sc}}/A_{\mathrm{loc}}$ are already very close to $\omega_c^{\min}$, $\bar{a}_{\mathrm{sc}}$ and $\bar{\alpha}$ (difference $< 10^{-7}$), respectively. Hence, numerically the lower case is already attained. However, this is indeed only numerically; confer Proposition 2.1.3.

(b) The least favorable radii in sense of Section 2.2 and the corresponding MSE–inefficiencies are given in Table 8.1.

relMSE $(\tilde{\eta}_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1.746	1.314	1.147	1.072	0.579	0.536	0.591

Table 8.1: Least favorable radius and MSE–inefficiency.

(c) The optimal AL estimator for the separate estimation of location and scale (ALs estimator) is identical to the Hu2 estimator introduced in Subsection 8.5.1.

////

In case of the classical optimal estimators ($r = 0$), the absolute information is a polynomial of degree 4. Hence, large observations are assumed to contain very much information and therefore have a large influence on the estimation; confer

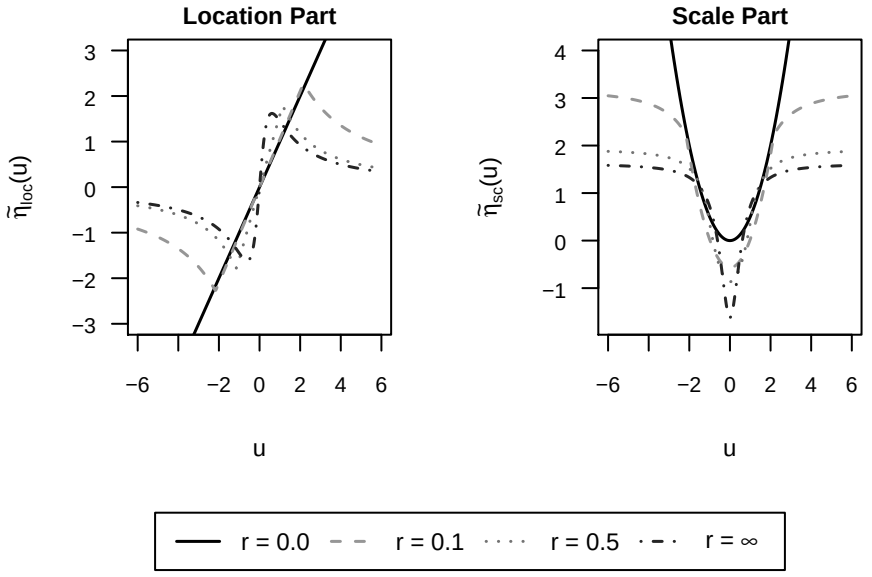

Figure 8.1: Location and scale part of the optimally robust ICs in case of AL estimators and radius $r = 0, 0.1, 0.5, \infty$.

Figure 8.2. In case of the most robust estimator ($r = \infty$), all observations are regarded to include the same amount of information ($(\omega_c^{\min})^2$).
Interestingly, this difference does not show up if one considers relative informations. Although the absolute information in case of the robust AL estimators is bounded and clearly differs from the classical case, the relative information of scale, respectively location is very similar to the classical optimal estimator. That is, the information contained in an observation is divided up between the location and the scale part of the IC in a very similar way; confer Figure 8.2.

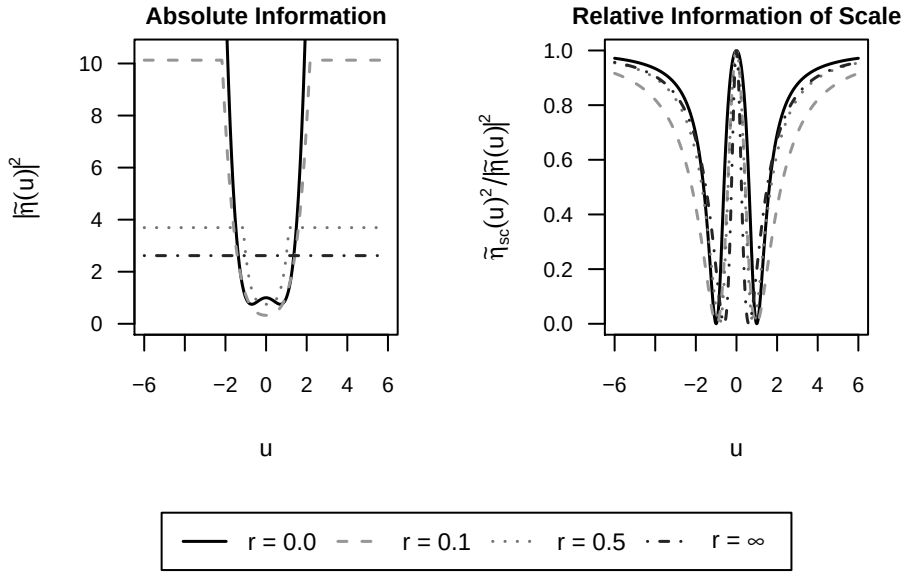

Figure 8.2: Absolute and relative information in case of AL estimators and radius $r = 0, 0.1, 0.5, \infty$.

8.3 M Estimators

The M estimators for normal location and scale have ICs $\rho \in \Psi_2$ of the following form

$$\rho(u) = \begin{pmatrix} \psi(u) \\ \gamma_\psi(u\psi(u) - 1) \end{pmatrix} \qquad \text{with} \qquad \gamma_\psi^{-1} = \mathrm{E}\, u^3\psi - 1 \tag{8.3.1}$$

where $\gamma_\psi \in \mathbb{R} \setminus \{0\}$. For fixed $\gamma \in \mathbb{R} \setminus \{0\}$, the solution given by Theorem 7.2.19 specializes to

$$\tilde{\psi}(u) = \frac{\alpha_1 u + \alpha_3 u^3}{q^2(u)} w(u) + \gamma^2 \frac{u}{q^2(u)} \tag{8.3.2}$$

with

$$w(u) = \min\left\{1, \frac{b(u)}{|g(u)|}\right\} \qquad q(u) = \sqrt{1 + \gamma^2 u^2} \tag{8.3.3}$$

and

$$g(u) = \frac{\alpha_1 u + \alpha_3 u^3}{q(u)} \qquad b(u) = \left[b^2 - \frac{\gamma^2}{q^2(u)}\right]^{1/2} \tag{8.3.4}$$

where

$$r^2 = \mathrm{E}\left(\frac{|g(u)|}{b(u)} - 1\right)_+ \qquad \text{if } b > \gamma \tag{8.3.5}$$

respectively

$$r^2 \geq \mathrm{E}\left(\frac{|g(u)|}{b(u)} - 1\right)_+ \qquad \text{if } b = \gamma \tag{8.3.6}$$

For a plot of the location and scale part of the optimally robust IC for different values of r see Figure 8.3.

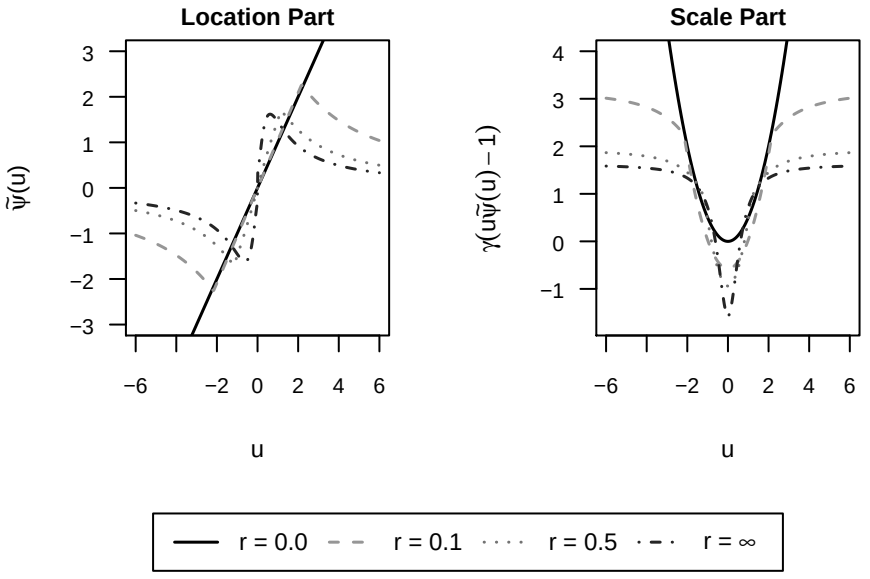

Figure 8.3: Location and scale part of the optimally robust ICs in case of M estimators and radius $r = 0, 0.1, 0.5, \infty$.

Remark 8.3.1 **(a)** Since $A_\psi = 1$ for all $\psi \in L_2(P)$, we only have to minimize the solutions in $\gamma \in \mathbb{R} \setminus \{0\}$, to solve problem (1.3.7) for ρ of form (7.1.47). This optimization is done numerically.

(b) Our numerical calculations yield equality (up to an error less than 10^{-8}) in (7.2.115) and (7.2.117) for the optimal $\bar{\gamma} = \operatorname{argmin} \bar{b}(\gamma)$, and thus

$$\bar{\zeta}(u) = \bar{b}(u)\operatorname{sign}(u) \qquad \bar{b}(u) = \left[\left(\omega_{\mathrm{M}}^{\min}\right)^2 - \tfrac{\bar{\gamma}^2}{1+\bar{\gamma}^2 u^2}\right]^{1/2} \tag{8.3.7}$$

with $\bar{\gamma} \approx 1.589$ and minimum bias $\omega_{\mathrm{M}}^{\min} \approx 1.619$; confer Theorem 7.2.19 (c).

(c) The resulting optimal ICs in case of M estimators are very similar but not identical to the corresponding optimal ICs in case of AL estimators. Hence, also the efficiency loss is very small and stays below 0.5%; confer Section 8.7. However, the optimal ICs in case of AL estimators are much faster to compute as there is no additional outer optimization needed.

(d) We do not consider M estimators with separate bounds (i.e., Ms estimators) as the results are almost identical to the results for BM estimators; confer Remark 7.3.8 (f). ////

Since the optimal ICs in case of M estimators are almost identical to the corresponding ICs in case of AL estimators, the plots of the absolute and relative information also look very similar. However, a closer inspection of the plot of the relative information shows that the deviations from the classical optimal IC are slightly larger than in case of the AL estimators; confer Figure 8.4 and Figure 8.2, respectively.

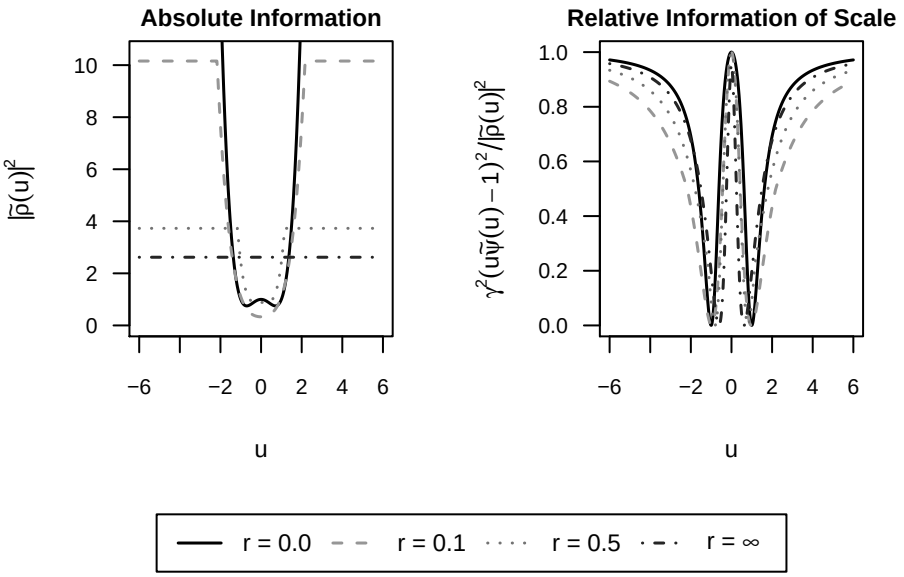

Figure 8.4: Absolute and relative information in case of M estimators and radius $r = 0, 0.1, 0.5, \infty$.

8.4 BM Estimators

The BM estimators for normal location and scale introduced by Bednarski and Müller (2001) have ICs of form (8.3.1). For fixed $\gamma \in \mathbb{R} \setminus \{0\}$, the solution can be read off from Propositions 7.3.6

$$\tilde{\psi}(u) = \operatorname{sign}(u) \min\left\{\alpha|u|, b_{\mathrm{loc}}, \frac{1+b_{\mathrm{sc},0}}{|u|}\right\} \tag{8.4.1}$$

where $\alpha \geq 0$ and the following representation holds

$$1 = \mathrm{E} \min\left\{\alpha u^2, b_{\mathrm{loc}}|u|, (1+b_{\mathrm{sc},0})\right\} \tag{8.4.2}$$

By Proposition 7.3.7, this $\tilde{\psi}$ is the solution to both problems (7.3.46) and (7.3.60) if γ is determined via

$$\gamma = \left[\mathrm{E} \min\left\{\alpha u^4,\, b_{\mathrm{loc}}|u|^3,\, (1+b_{\mathrm{sc},0})u^2\right\} - 1\right]^{-1} \tag{8.4.3}$$

For a plot of the location and scale part of the optimally robust IC for different values of r see Figure 8.5.

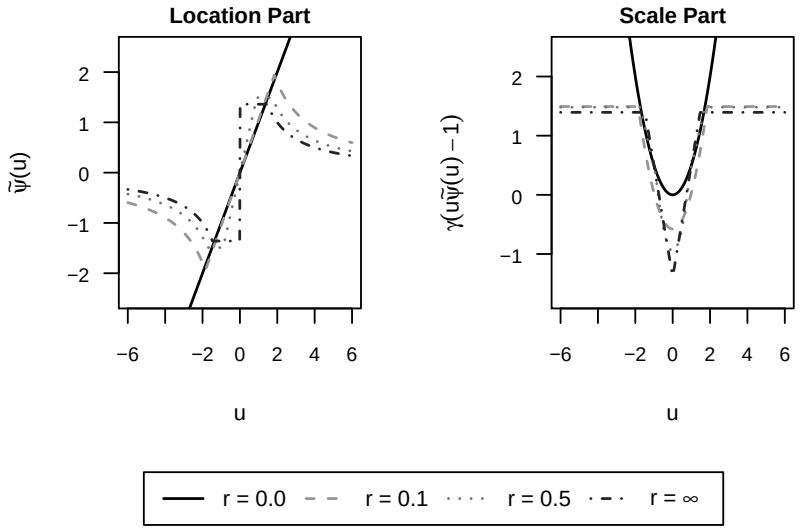

Figure 8.5: Location and scale part of the optimally robust ICs in case of BM estimators and radius $r = 0, 0.1, 0.5, \infty$.

Remark 8.4.1 **(a)** Since there is no outer optimization in γ, the BM estimators are probably only suboptimal M estimators. This is indeed confirmed by our numerical results; confer Section 8.7.

(b) We determine the minimax asymptotic MSE solution via numerical optimization in b_{loc} and b_{sc} where

$$\sup_P |\rho|^2 = \sup_P [\tilde{\psi}^2 + \gamma^2(u\tilde{\psi} - 1)^2] = b_{\mathrm{loc}}^2 + \gamma^2 b_{\mathrm{sc},0}^2 \tag{8.4.4}$$

(c) The minimum bias solution specified in Propositions 7.3.6 and 7.3.7 coincide (at least numerically) and we obtain $\bar{b}_{\mathrm{loc}} \approx 1.360$, $\bar{b}_{\mathrm{sc},0} = 1$ and $\bar{\gamma} \approx 1.395$. Hence, $\omega_{\mathrm{BM}}^{\min} \approx 1.948$; confer Section 8.7. ////

The plot of the relative information indicates that the information contained in the observations is divided up very similarly compared with the AL and M estimators. However, the plot of the absolute information looks very different; confer Figures 8.6, 8.4 and 8.2, respectively. Thus, the suboptimality of the BM estimators could be caused by the fact how they judge the amount of information included in the observations. More precisely, they assume more information around 1 and less information for large (in absolute values) observations. Moreover, the absolute information of the lower case is not constant. Maybe, this reflects the fact that the optimal solution is rather optimal in case of location and is optimal in case of scale only if one chooses γ via (8.4.3).

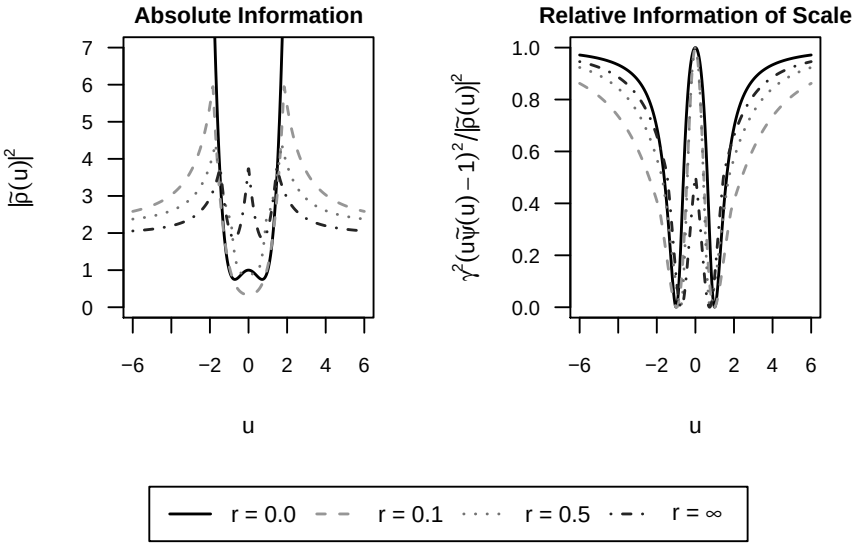

Figure 8.6: Absolute and relative information in case of BM estimators and radius $r = 0, 0.1, 0.5, \infty$.

8.5 Other Proposals

Following Subsection 6.4 of Huber (1981), the simultaneous M estimate of location and scale is any pair of estimators (T_n, S_n) determined by two equations of the form

$$\sum_{i=1}^{n} \psi\Big(\frac{y_i - T_n}{S_n}\Big) = 0 \tag{8.5.1}$$

$$\sum_{i=1}^{n} \chi\Big(\frac{y_i - T_n}{S_n}\Big) = 0 \tag{8.5.2}$$

where ψ is a generalization of the score function Λ_f^{loc} and $\chi = u\psi - 1$ or an arbitrary function which may be independent of ψ.
By choosing odd functions $\psi \in L_2(P)$ (i.e., $\mathrm{E}\,\psi = 0$), respectively even functions $\chi \in L_2(P)$ (i.e., $\mathrm{E}\,\chi = 0$), the corresponding ICs are

$$\mathrm{IC}_\psi(u) = A\psi(u) \qquad \text{with} \qquad A^{-1} = \mathrm{E}\,u\psi(u) \tag{8.5.3}$$

and

$$\mathrm{IC}_\chi(u) = \gamma\chi(u) \qquad \text{with} \qquad \gamma^{-1} = \mathrm{E}\,u^2\chi(u) \tag{8.5.4}$$

Remark 8.5.1 By the assumed symmetry of ψ and χ and the symmetry of $\mathcal{N}(0,1)$, it always holds $\mathrm{E}\, u^2\psi = 0$ and $\mathrm{E}\, u\chi = 0$, respectively. ////

The proposals introduced in the subsequent subsections contain certain tuning constants. We determine the corresponding optimal tuning constants by numerically minimization of the maximum asymptotic MSE

$$\mathrm{E}\,\mathrm{IC}_\psi^2 + \mathrm{E}\,\mathrm{IC}_\chi^2 + r^2 \sup_P [\mathrm{IC}_\psi^2 + \mathrm{IC}_\chi^2] \tag{8.5.5}$$

where the neighborhood radius $r \in (0,\infty)$ is fixed.

8.5.1 Huber Estimators

We call these estimators Huber estimators as they are based on Huber's ψ-function

$$\psi_{\mathrm{Hu1}}(u) = -k \vee u \wedge k \qquad k \in [0,\infty] \tag{8.5.6}$$

derived in his famous work about the robust estimation of a location parameter; confer Huber (1964). Based on ψ_{Hu1} we study the following estimators:

(a) **Hu1 estimator:** This estimator was proposed by Huber (1964) (Proposal 2) and is based on the following ψ and χ functions which both depend on the same parameter $k \in [0,\infty]$

$$\psi(u) = \psi_{\mathrm{Hu1}}(u) \tag{8.5.7}$$
$$\chi(u) = \chi_{\mathrm{Hu1}}(u) = \psi_{\mathrm{Hu1}}(u)^2 - \mathrm{E}\,\psi_{\mathrm{Hu1}}^2 \tag{8.5.8}$$

(b) **Hu2a estimator:** We modify the Hu1 estimator by using two tuning parameters $0 \le k_1 \le k_2 \le \infty$ in Huber's ψ function ψ_{Hu1} in such a way that not only large but also small values of u may be modified

$$\psi(u) = \psi_{\mathrm{Hu2a}}(u) = \mathrm{sign}(u)[k_1 \vee |u| \wedge k_2] \tag{8.5.9}$$
$$\chi(u) = \chi_{\mathrm{Hu2a}}(u) = \psi_{\mathrm{Hu2a}}(u)^2 - \mathrm{E}\,\psi_{\mathrm{Hu2a}}^2 \tag{8.5.10}$$

In particular, this leads to a possible clipping from below in case of χ_{Hu2a} for small values of u as it occurs in case of the optimally robust IC for normal scale starting at radius $r \approx 0.920$; confer Subsection 2.3.1 of Rieder et al. (2001).

(c) **Hu2 estimator:** This estimator was proposed in Example 6.4.1 of Huber (1981) and combines the optimally robust IC of normal location with the optimally robust IC of normal scale (for small radii); i.e., ψ and χ may be calibrated via the parameters $k \in [0,\infty]$ and $c \in [0,\infty]$, respectively

$$\psi(u) = \psi_{\mathrm{Hu1}}(u) \tag{8.5.11}$$
$$\chi(u) = \chi_{\mathrm{Hu2}}(u) = u^2 \wedge c^2 - \mathrm{E}[u^2 \wedge c^2] \tag{8.5.12}$$

The Hu2 estimator corresponds to the ALs estimator introduced in Subsection 7.3.1.

(d) **Hu3 estimator:** Since the optimally robust IC in case of normal scale is also clipped from below for larger radii ($r \geq 0.920$), we introduce a further tuning parameter in χ_{Hu2} to make this possible

$$\psi(u) = \psi_{\text{Hu1}}(u) \tag{8.5.13}$$
$$\chi(u) = \chi_{\text{Hu3}}(u) = c_1^2 \vee u^2 \wedge c_2^2 - \text{E}[c_1^2 \vee u^2 \wedge c_2^2] \tag{8.5.14}$$

where $0 \leq c_1 \leq c_2 \leq \infty$.

(e) **HuMad estimator:** Finally, we combine the optimally robust IC of normal location with the IC of the median of the absolute deviations of the median (MAD)

$$\psi(u) = \psi_{\text{Hu1}}(u) \tag{8.5.15}$$
$$\chi(u) = \chi_{\text{Mad}}(u) = \text{sign}(|u| - \Phi^{-1}(3/4)) \tag{8.5.16}$$

This estimator was proposed by Hampel in the Princeton robustness study; confer Andrews et al. (1972), p 12.

For a plot of the location and scale part of the optimally robust ICs for different values of r see Figure 8.7.

Remark 8.5.2 (a) We do not introduce a Huber estimator with $\chi = u\psi_{\text{Hu1}} - 1$ as this would lead to an unbounded IC.

(b) In all considered cases we obtain $k_1 \approx 0$ for the Hu2a estimator; i.e., the optimally robust IC of the Hu2a estimator is identical to the optimally robust IC of the Hu1 estimator. Moreover, we encounter only very small differences between the optimally robust ICs of the Hu2 and the Hu3 estimator (in all cases $c_1 < 0.05$). Hence, if we choose the optimal (with respect to the maximum asymptotic MSE) tuning parameters, the five proposals lead only to three (numerically) distinguishable estimators namely the Hu1, Hu2 and HuMad estimators; confer also Section 8.7.

(c) The Hu1, Hu2 and HuMad estimators are implemented in the `R`, respectively S-plus package `ROBETH`; confer Chapters 1 and 11 of Marazzi (1993). However, the choice of the tuning parameters is not done with respect to the maximum asymptotic MSE but with respect to the asymptotic variance efficiency at the ideal model. ////

The absolute and relative information in case of the Huber estimators look completely different from the estimators considered so far. This probably reflects the fact that these estimators are rather designed for the separate estimation of location and scale where scale is regarded as a nuisance parameter. The absolute and relative information of the classical optimal estimator seems to be best approximated by the Hu2 estimator among the Huber estimators and this estimator also has the smallest MSE–inefficiency among the Huber estimators; confer Section 8.7.

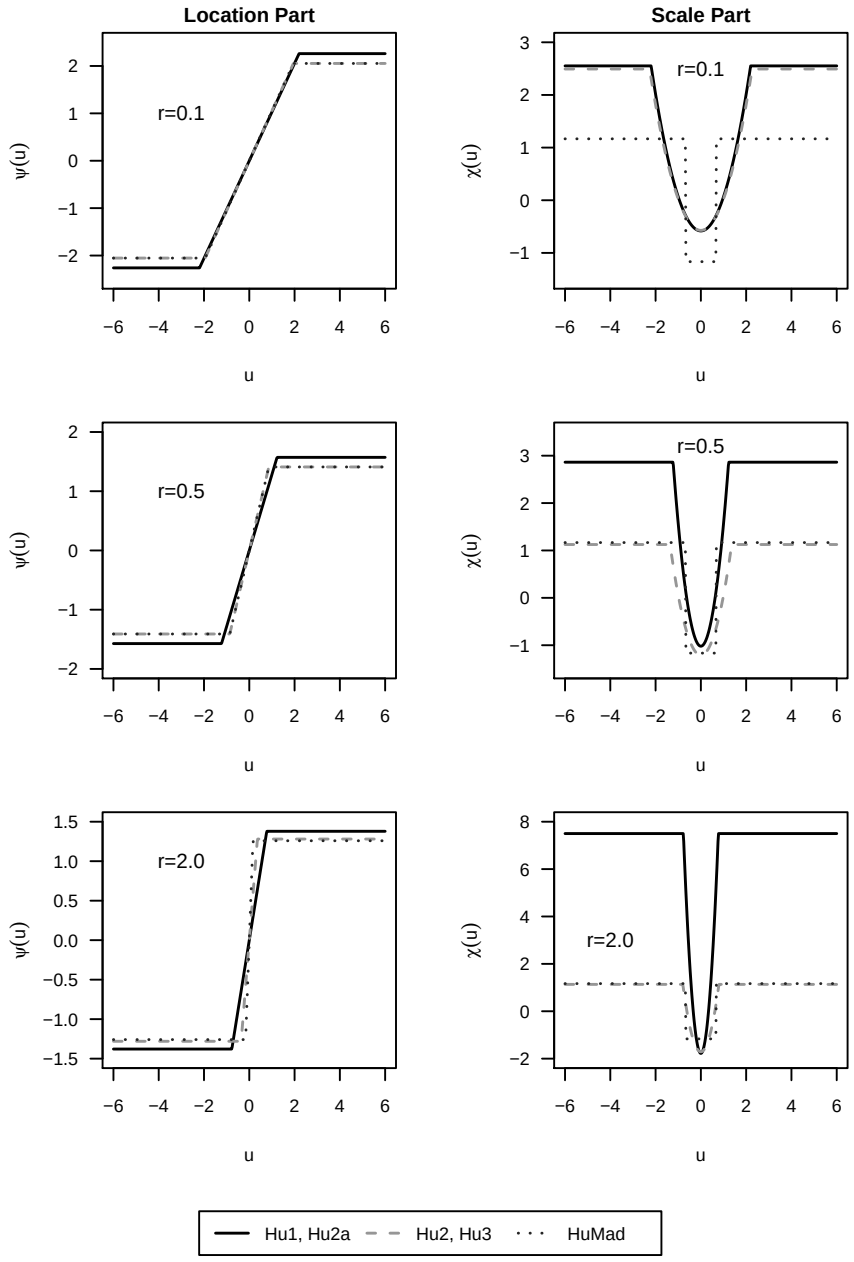

Figure 8.7: Location and scale part of the optimally robust ICs in case of Huber estimators and radius $r = 0.1, 0.5, 2.0$.

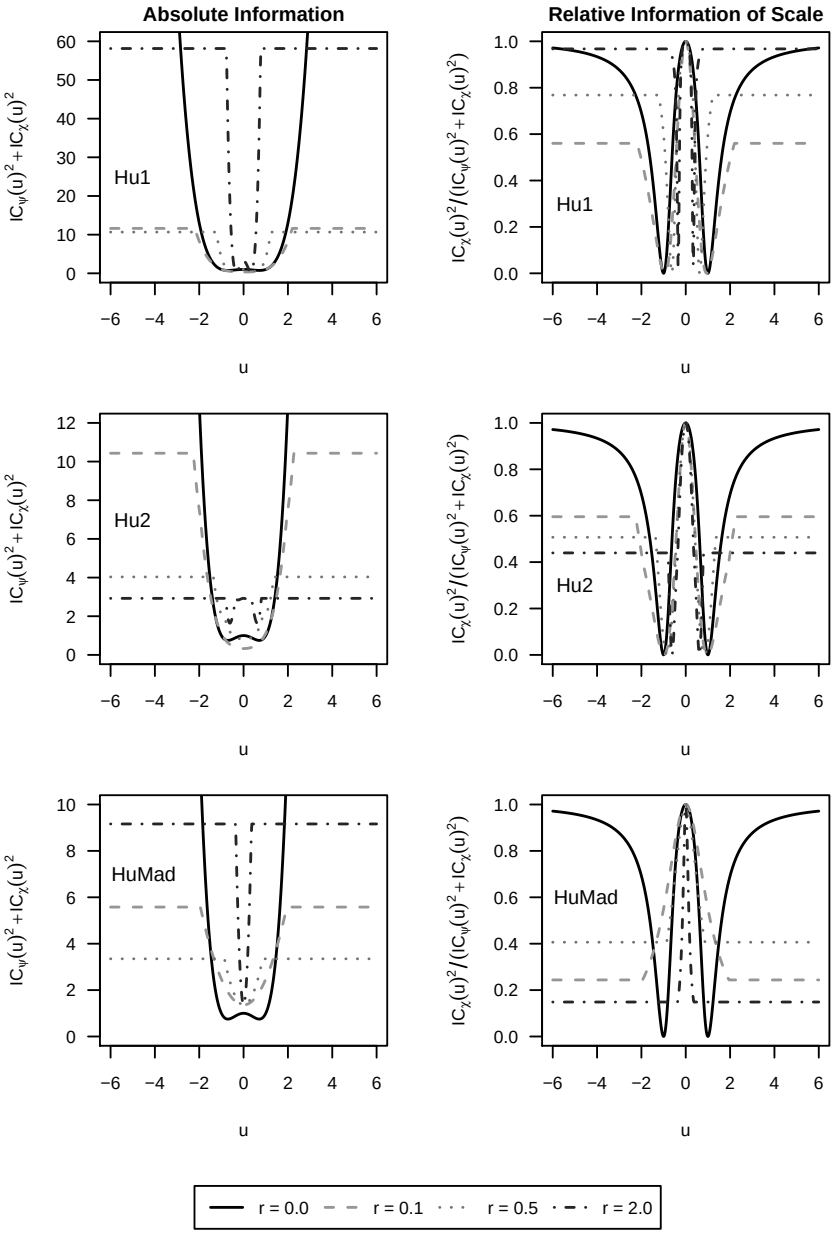

Figure 8.8: Absolute and relative information in case of Hu1, Hu2, HuMad estimators and radius $r = 0, 0.1, 0.5, 2.0$ ($r = 0$ corresponds to the classical optimal ICs).

8.5.2 Hampel Estimators

The Hampel estimators are based on the following three part redescending ψ function introduced by Hampel

$$\psi_{\text{Ha3}}(u) = \text{sign}(u) \begin{cases} |u| & \text{if } 0 \le |u| < a \\ a & \text{if } a \le |u| < b \\ \frac{c-|u|}{c-b}\, a & \text{if } b \le |u| < c \\ 0 & \text{if } |u| \ge c \end{cases} \tag{8.5.17}$$

where $0 \le a \le b \le c \le \infty$; confer Subsections 2C3, 3C3 of Andrews et al. (1972) . Based on ψ_{Ha3} we study the following estimators:

(a) **Ha3 estimator:** This estimator is motivated by the choice $\chi = u\psi - 1$. Hence,

$$\psi(u) = \psi_{\text{Ha3}}(u) \tag{8.5.18}$$
$$\chi(u) = \chi_{\text{Ha3}}(u) = Au\psi_{\text{Ha3}}(u) - 1 \qquad A^{-1} = \text{E}\, u\psi_{\text{Ha3}} \tag{8.5.19}$$

(b) **Ha4 estimator:** We combine Hampel's location estimator with the optimally robust estimator for normal scale; i.e.,

$$\psi(u) = \psi_{\text{Ha3}}(u) \tag{8.5.20}$$
$$\chi(u) = \chi_{\text{Hu1}}(u) = u^2 \wedge k^2 - \text{E}[u^2 \wedge k^2] \tag{8.5.21}$$

where $k \in [0, \infty]$; confer also Chapter 11 of Marazzi (1993).

(c) **HaMad estimator:** This estimator was considered in Andrews et al. (1972); confer also Example 6.6.2 of Huber (1981). It is based on

$$\psi(u) = \psi_{\text{Ha3}}(u) \tag{8.5.22}$$
$$\chi(u) = \chi_{\text{Mad}}(u) = \text{sign}(|u| - \Phi^{-1}(3/4)) \tag{8.5.23}$$

For a plot of the location and scale part of the optimally robust ICs of the Ha3 estimator for different values of r see Figure 8.9.

Remark 8.5.3 (a) In case of the Ha4 and the HaMad estimators, the optimal tuning constants b and c are very large and hence have no effect on the results; i.e., with respect to the maximum asymptotic MSE the redescending is denied and the Ha4 and HaMad estimators coincide (at least numerically) with the Hu2 and HuMad estimators, respectively; confer also Section 8.7.

(b) The Ha4 and HaMad estimators are implemented in the `R`, respectively S-plus package `ROBETH`; confer Chapters 1 and 11 of Marazzi (1993). However, the choice of the tuning parameters is not done with respect to the maximum asymptotic MSE but with respect to the asymptotic variance efficiency at the ideal model. ////

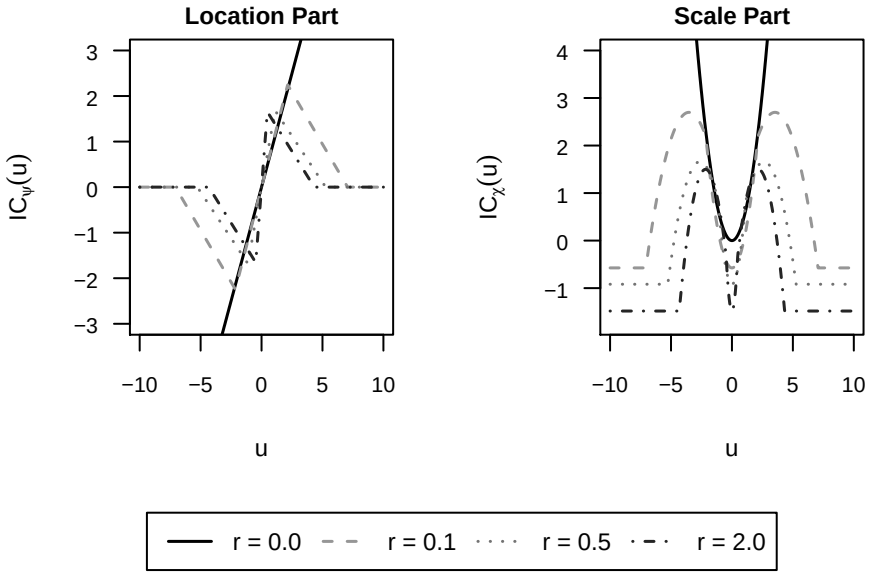

Figure 8.9: Location and scale part of the optimally robust ICs in case of Ha3 estimators and radius $r = 0.1, 0.5, 2.0$.

Since the Ha4 and and HaMad estimator are numerically identical to the Hu2 and HuMad estimator, the plots of the absolute and relative information are also identical. Hence, we refer to Figure 8.8, respectively Subsection 8.5.1 for the interpretation. The optimal influence curve of the Ha3 estimator is redescending in the location and the scale part; confer Figure 8.9. Thus, the absolute information of large observations is considered very small. But, in the inner part the absolute and relative information is divided up similar to the classical optimal estimators; confer Figure 8.10. As a consequence, the Ha3 estimator loses some efficiency compared with the optimal AL estimators. However, this loss is not too large; confer Section 8.7.

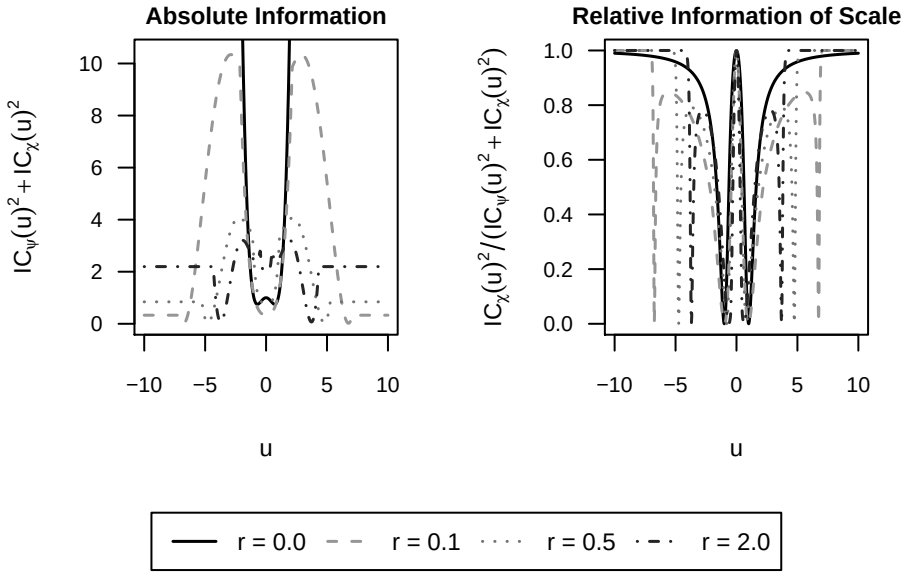

Figure 8.10: Absolute and relative information in case of Ha3 estimators and radius $r = 0, 0.1, 0.5, 2.0$.

8.5.3 Andrews Estimators

We call these estimators Andrews estimators as they are based on the following ψ function

$$\psi_{\mathrm{An1}}(u) = \sin(u/a)\, \mathrm{I}_{[-a\pi, a\pi]}(u) \qquad a \in (0, \infty) \tag{8.5.24}$$

which was first implemented by Andrews; confer Subsections 2C3, 3C3 of Andrews et al. (1972) and also Example 2.6.2 of Hampel et al. (1986). Based on ψ_{An1} we investigate the following estimators:

(a) **An1 estimator:** This estimator is motivated by the choice $\chi = u\psi - 1$. That is,

$$\psi(u) = \psi_{\mathrm{An1}}(u) \tag{8.5.25}$$

$$\chi(u) = \chi_{\mathrm{An1}}(u) = Au\psi_{\mathrm{An1}}(u) - 1 \qquad A^{-1} = \mathrm{E}\, u\psi_{\mathrm{An1}} \tag{8.5.26}$$

(b) **An2 estimator:** We combine Andrews's location estimator with the optimally robust estimator for normal scale; i.e.,

$$\psi(u) = \psi_{\mathrm{An1}}(u) \tag{8.5.27}$$
$$\chi(u) = \chi_{\mathrm{Hu1}}(u) = u^2 \wedge k^2 - \mathrm{E}[u^2 \wedge k^2] \tag{8.5.28}$$

where $k \in [0, \infty]$.

(c) **AnMad estimator:** This estimator was considered in Andrews et al. (1972). It is based on

$$\psi(u) = \psi_{\mathrm{An1}}(u) \tag{8.5.29}$$
$$\chi(u) = \chi_{\mathrm{Mad}}(u) = \mathrm{sign}(|u| - \Phi^{-1}(3/4)) \tag{8.5.30}$$

For a plot of the location and scale part of the optimally robust ICs for different values of r see Figure 8.11.

The results for the absolute and relative information of the Andrews estimators are clearly different from the corresponding results in case of the classical optimal estimator; confer Figure 8.12. This might explain why these estimators have a rather large subefficiency; confer Section 8.7.

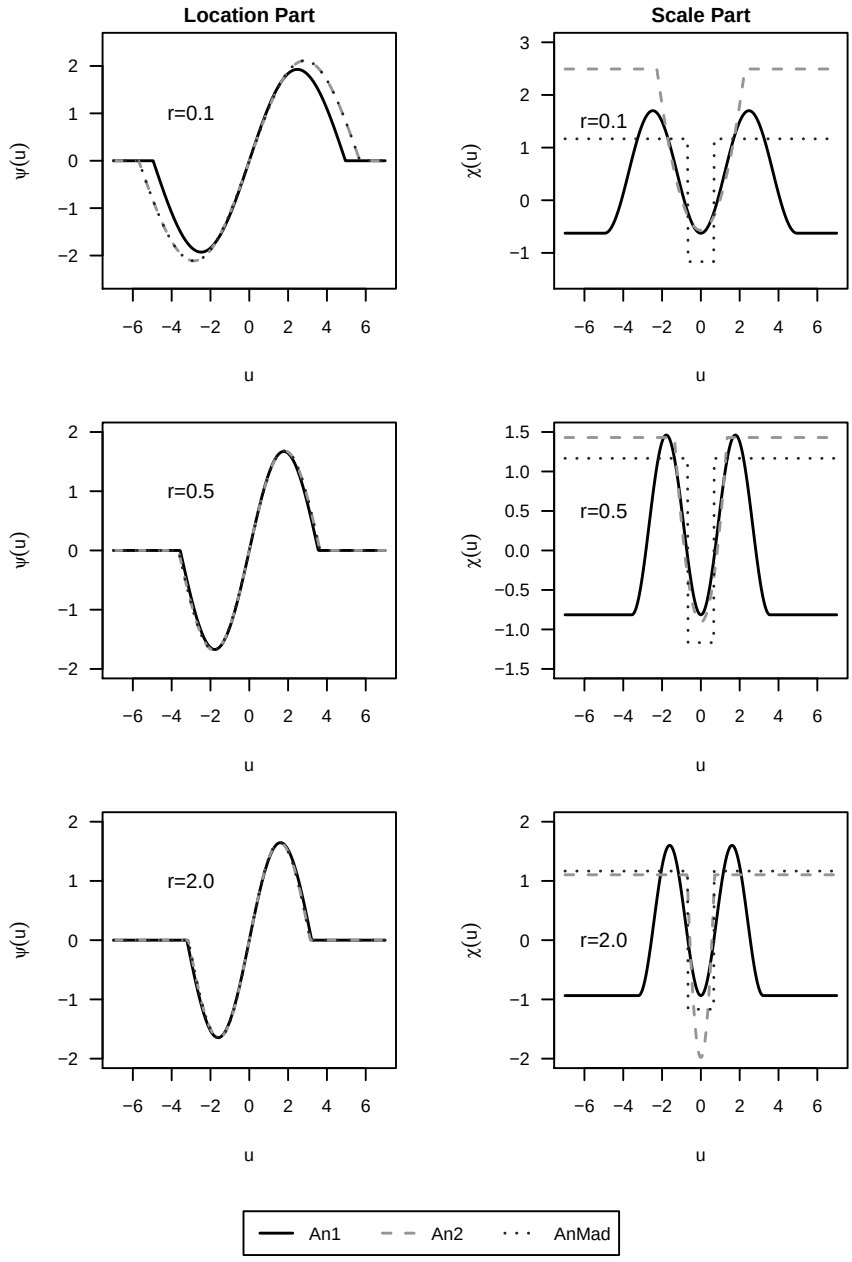

Figure 8.11: Location and scale part of the optimally robust ICs in case of Andrews estimators and radius $r = 0.1, 0.5, 2.0$.

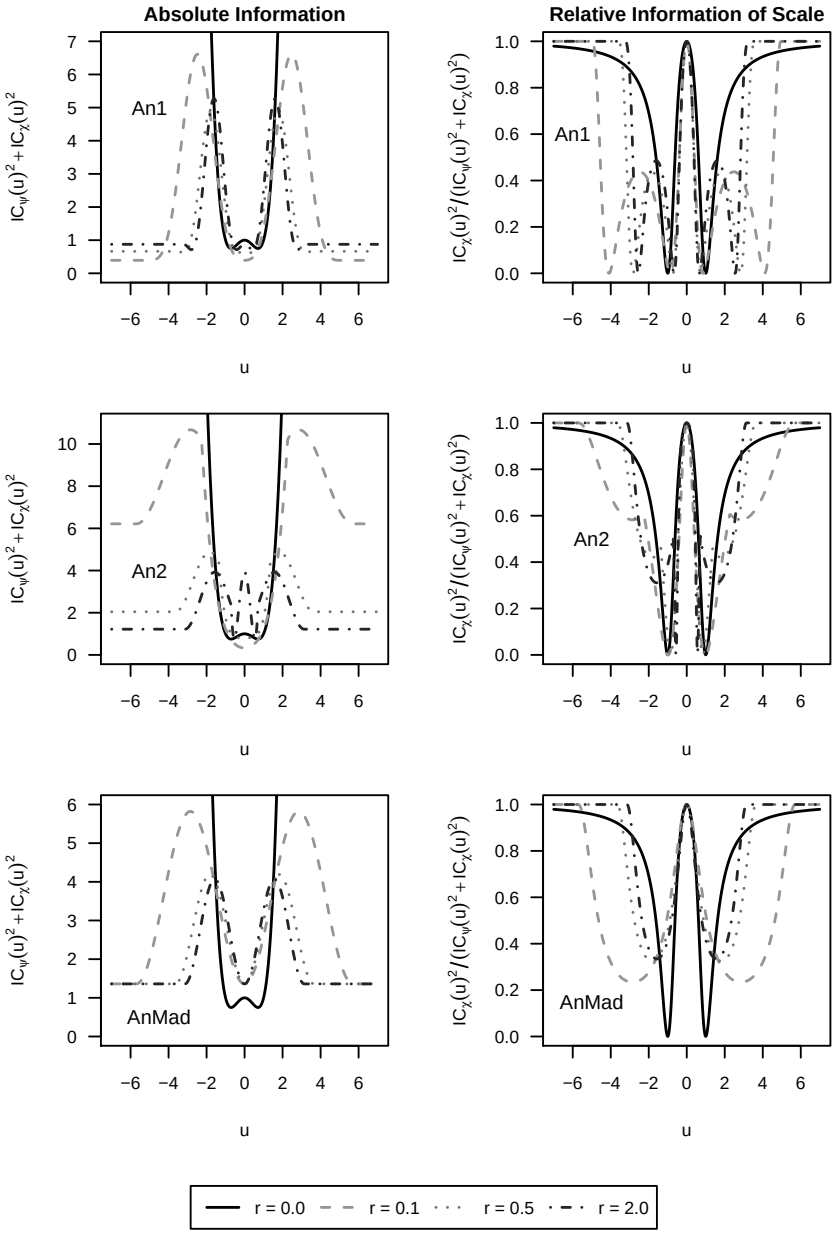

Figure 8.12: Absolute and relative information in case of Andrews estimators and radius $r = 0, 0.1, 0.5, 2.0$ ($r = 0$ corresponds to the classical optimal ICs).

8.5.4 Tukey Estimators

These estimators are based on an idea of Beaton and Tukey (1974) who proposed the following ψ function

$$\psi_{\mathrm{Tu1}}(u) = u(a^2 - u^2)^2 \, \mathrm{I}_{[-a,a]}(u) \qquad a \in (0, \infty) \tag{8.5.31}$$

which is also known as Tukey's biweight (or "bisquare"); confer Example 2.6.3 of Hampel et al. (1986). Based on ψ_{Tu1} we introduce the following estimators:

(a) **Tu1 estimator:** This estimator is motivated by the choice $\chi = u\psi - 1$; i.e.,

$$\psi(u) = \psi_{\mathrm{Tu1}}(u) \tag{8.5.32}$$

$$\chi(u) = \chi_{\mathrm{Tu1}}(u) = Au\psi_{\mathrm{Tu1}}(u) - 1 \qquad A^{-1} = \mathrm{E}\, u\psi_{\mathrm{Tu1}} \tag{8.5.33}$$

(b) **Tu2 estimator:** We combine Tukey's biweight with the optimally robust estimator for normal scale; i.e.,

$$\psi(u) = \psi_{\mathrm{Tu1}}(u) \tag{8.5.34}$$

$$\chi(u) = \chi_{\mathrm{Hu1}}(u) = u^2 \wedge k^2 - \mathrm{E}[u^2 \wedge k^2] \tag{8.5.35}$$

where $k \in [0, \infty]$; confer also Chapter 11 of Marazzi (1993).

(c) **TuMad estimator:** This estimator combines Tukey's biweight with the median of the absolute deviations of the median; i.e.,

$$\psi(u) = \psi_{\mathrm{Tu1}}(u) \tag{8.5.36}$$

$$\chi(u) = \chi_{\mathrm{Mad}}(u) = \mathrm{sign}(|u| - \Phi^{-1}(3/4)) \tag{8.5.37}$$

For a plot of the location and scale part of the optimally robust ICs for different values of r see Figure 8.13.

The plots of the relative and absolute information of the Tukey estimators look similar to those of the Andrews estimators (cf. Figure 8.14, resp. Figure 8.12) and also the results for the maximum asymptotic MSE and the corresponding MSE–inefficiencies are similar; confer Section 8.7.

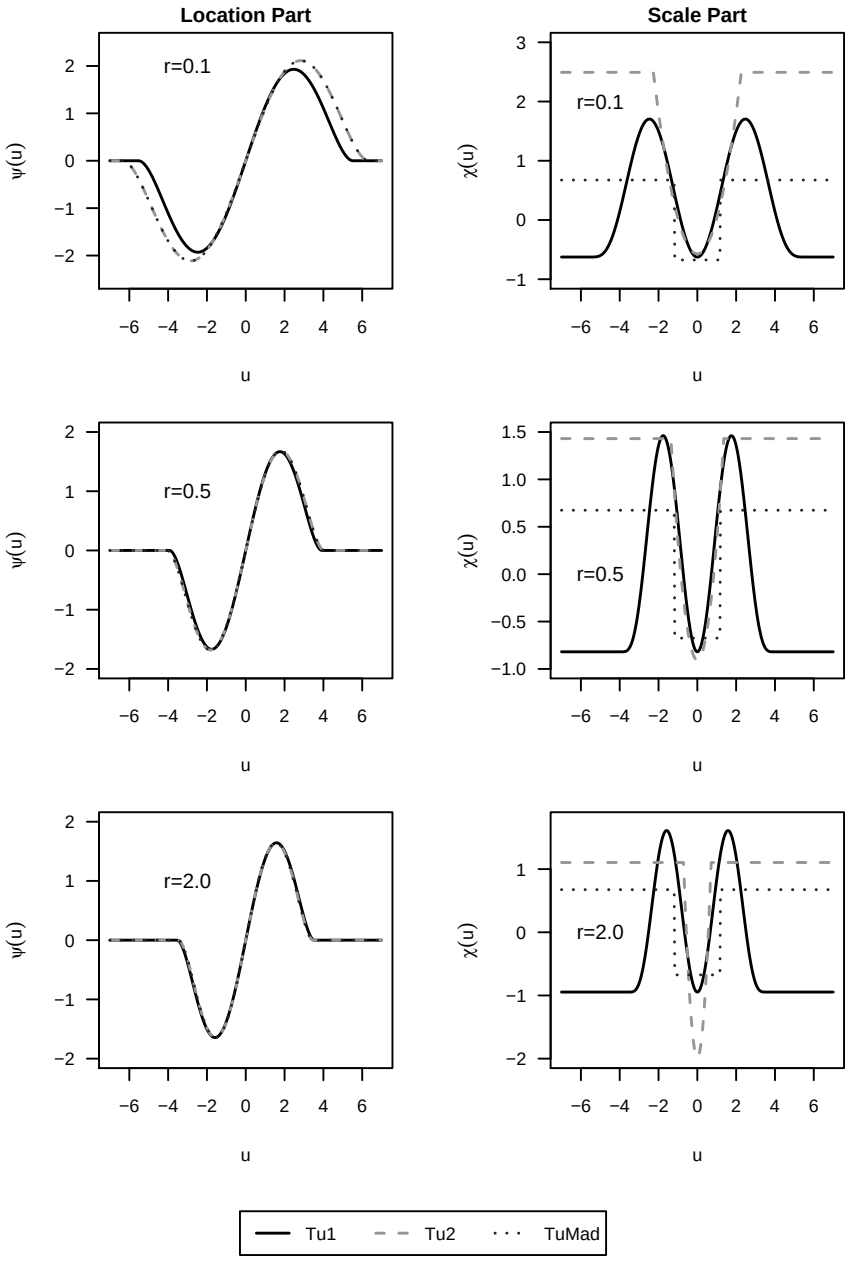

Figure 8.13: Location and scale part of the optimally robust ICs in case of Tukey estimators and radius $r = 0.1, 0.5, 2.0$.

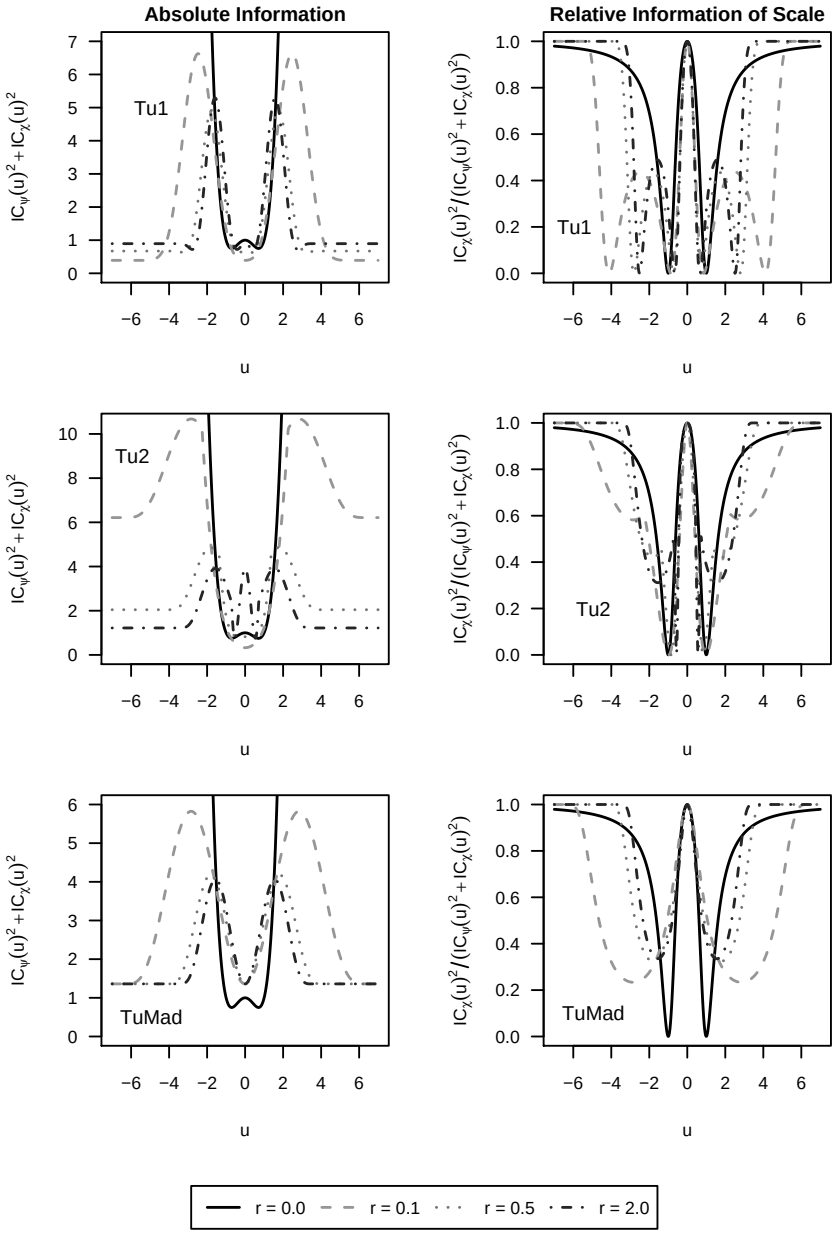

Figure 8.14: Absolute and relative information in case of Tukey estimators and radius $r = 0, 0.1, 0.5, 2.0$ ($r = 0$ corresponds to the classical optimal ICs).

8.6 MM Estimators

The notion of MM estimators was introduced by Yohai (1987). These estimators combine an M estimator T_n of location with an S estimator S_n of scale. In this setup the parameter of interest is the location parameter μ whereas σ is regarded as an unknown nuisance parameter. The M estimator T_n is the solution of an M equation of the form

$$\sum_{i=1}^{n} \psi\Big(\frac{y_i - T_n}{S_n}\Big) = 0 \tag{8.6.1}$$

where $\psi\colon \mathbb{R} \to \mathbb{R}$ is a non-decreasing, odd and continuously differentiable function. The S estimator S_n is a scale estimator of the residuals. It was introduced by Rousseeuw and Yohai (1984) and is defined as

$$S_n = \inf_{t\in\mathbb{R}} s_n(t) \tag{8.6.2}$$

where $s_n(t)$ is the solution of

$$\sum_{i=1}^{n} \Big[\chi\Big(\frac{y_i - t}{s_n(t)}\Big) - \mathrm{E}\,\chi\Big] = 0 \tag{8.6.3}$$

Thus, we also obtain an S location estimator via this procedure, namely

$$\arg\inf_{t\in\mathbb{R}} s_n(t) \tag{8.6.4}$$

Remark 8.6.1 **(a)** For more details on MM estimators of location and scale including uniform consistency and uniform asymptotic normality over fixed contamination neighborhoods we refer to Chapter 2 of Salibian-Barrera (2000) and Salibian-Barrera and Zamar (2004), respectively.

(b) The corresponding ICs in the normal location model are

$$\mathrm{IC}_\psi(u) = A\psi(u) \qquad \text{with} \qquad A^{-1} = \mathrm{E}\,u\psi \tag{8.6.5}$$

and

$$\mathrm{IC}_\chi(u) = \gamma(\chi(u) - \mathrm{E}\,\chi) \qquad \text{with} \qquad \gamma^{-1} = \mathrm{E}\,u^2\chi - \mathrm{E}\,\chi \tag{8.6.6}$$

The asymptotic linearity of the S estimator of scale may for instance be read off from the proof of Theorem 2.6 in Salibian-Barrera (2000). ////

For the purpose of this comparative study we use

$$\psi_c(u) = \mathrm{sign}(u) \begin{cases} |u|/c & \text{if } |u| \le 0.8c \\ p_4(|u|/c) & \text{if } 0.8c < |u| \le c \\ p_4(1) & \text{if } |u| > c \end{cases} \tag{8.6.7}$$

where $c \in (0, \infty)$ and $p_4(u) = 38.4 - 175u + 300u^2 - 225u^3 + 62.5u^4$. This ψ_c is a smoothed version of Huber's ψ function ψ_{Hu1} and we combine it with Tukey's biweight family (cf. Beaton and Tukey (1974))

$$\chi_d(u) = \begin{cases} 3(u/d)^2 - 3(u/d)^4 + (u/d)^6 & \text{if } |u| \le d \\ 1 & \text{if } |u| > d \end{cases} \tag{8.6.8}$$

where $d \in (0, \infty)$.

Remark 8.6.2 This special choices of ψ and χ are based on a proposal of Fraiman et al. (2001), p 206. The values $d = 1.548, 1.988$ which correspond to breakdown points 0.5 and 0.4, respectively in the setup of fixed contamination neighborhoods (cf. Table 1 of Salibian-Barrera and Zamar (2004)) in our setup are the MSE optimal MM estimators for radius $r \approx 0.775$ and $r \approx 0.440$, respectively. We determine the optimal MM estimators by numerical minimization of (8.5.5) and denote these estimators by MM2 estimators. ////

For a plot of the location and scale part of the optimally robust ICs of the MM2 estimators for different values of r see Figure 8.15. As already mentioned, the location

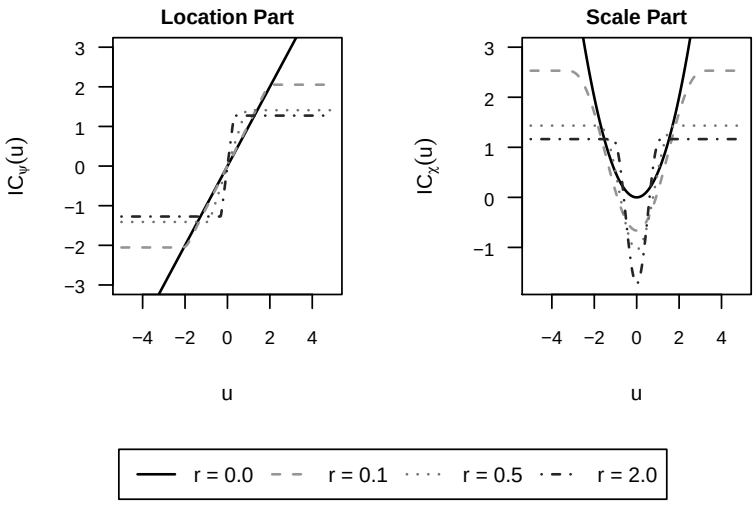

Figure 8.15: Location and scale part of the optimally robust ICs in case of the MM2 estimators and radius $r = 0.1, 0.5, 2.0$ ($r = 0.0$ corresponds to the classical optimal ICs).

part is a smoothed version of Huber's ψ function and as we see, the scale part looks like a smoothed version of χ_{Hu1}. Hence, not surprisingly the results for the maximum asymptotic MSE and the corresponding MSE–inefficiency are also very similar to the results for the Hu2, Hu3 and Ha4 estimator; confer Section 8.7. In particular, the plots of the absolute and relative information look very similar, too; confer Figure 8.16 and Figure 8.8, respectively. Thus, we refer to Subsection 8.5.1 for the interpretation.

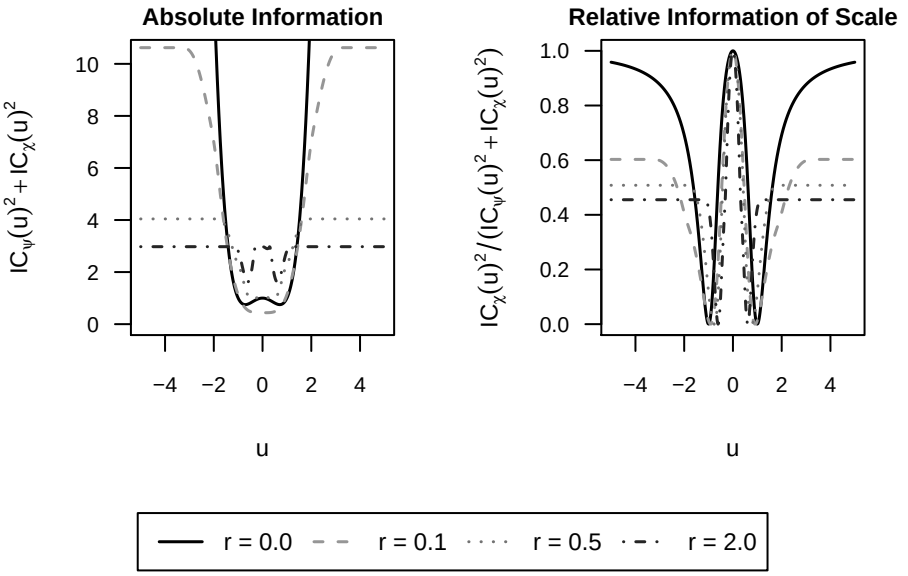

Figure 8.16: Absolute and relative information in case of the MM2 estimators and radius $r = 0.0, 0.1, 0.5, 2.0$.

8.7 Numerical Comparison

We give some results for the maximum asymptotic MSE of the estimators introduced in Sections 8.2–8.6 and specify the MSE–inefficiency of these estimators with respect to the optimal AL estimators. The values for the MSE–inefficiency at $r = \infty$ are numerical approximations for very large radii; i.e., we computed the maximum asymptotic MSE and the corresponding asymptotic bias for increasing values of r until the asymptotic bias was numerically constant.
The M estimators are closest to the optimal AL estimators; they lose almost no

efficiency ($< 0.5\%$). Moreover, all estimators except for those which include the MAD, perform well for small radii ($r < 0.25$). The Hu2, Hu3, Ha4 and MM2 estimators yield good results for all considered radii not losing much more than 10% efficiency. This is also true for the Hu1, Hu2a and Ha3 estimators for radii up to about 1.5 and for the An2, Tu2 and BM estimators for radii smaller than 1.0. In case the neighborhood radius is large (> 1.0), the HuMad and HaMad estimators are also acceptable candidates. In contrast, the An1 and Tu1 estimators for radii $r > 0.25$ and the AnMad and the TuMad estimators for all considered radii have quite large MSE–inefficiencies; i.e., with respect to the maximum asymptotic MSE these estimators cannot be supported.

Remark 8.7.1 In view of the maximum asymptotic MSE and as the computational effort for all these estimators is comparable, we recommend to use the optimal AL estimators for the simultaneous estimation of location and scale. ////

r	AL	M	BM	Hu1/2a	Hu2/3, Ha4	Hu/HaMad	Ha3
0.1	1.648	1.648	1.660	1.655	1.651	2.428	1.650
0.25	2.011	2.013	2.073	2.050	2.030	2.666	2.026
0.5	2.820	2.829	3.056	2.967	2.900	3.337	2.896
1.0	5.137	5.159	6.152	5.692	5.480	5.685	5.580
1.5	8.506	8.508	10.940	9.680	9.246	9.422	9.687
2.0	13.092	13.096	17.602	15.222	14.370	14.589	15.321

Table 8.2: Minimax asymptotic MSE for AL, M, BM, Huber and Hampel estimators.

r	An1	An2	AnMad	Tu1	Tu2	TuMad	MM2
0.1	1.673	1.657	2.434	1.673	1.657	2.434	1.666
0.25	2.116	2.058	2.693	2.111	2.058	2.693	2.049
0.5	3.363	3.064	3.501	3.337	3.064	3.500	2.921
1.0	7.968	6.378	6.583	7.850	6.373	6.578	5.504
1.5	15.532	11.483	11.673	15.259	11.469	11.660	9.279
2.0	26.099	18.405	18.788	25.606	18.381	18.764	14.486

Table 8.3: Minimax asymptotic MSE for Andrews, Tukey and MM2 estimators.

r	M	BM	Hu1/2a	Hu2/3, Ha4	Hu/HaMad	Ha3
0.1	1.000	1.007	1.004	1.002	1.473	1.001
0.25	1.001	1.031	1.019	1.009	1.326	1.007
0.5	1.003	1.084	1.052	1.028	1.183	1.027
1.0	1.004	1.198	1.108	1.067	1.107	1.086
1.5	1.000	1.286	1.138	1.087	1.108	1.139
2.0	1.000	1.344	1.163	1.098	1.114	1.170
∞	1.000	1.450	1.209	1.117	1.120	1.216

Table 8.4: MSE–Inefficiency for AL, M, BM, Huber and Hampel estimators.

r	An1	An2	AnMad	Tu1	Tu2	TuMad	MM2
0.1	1.015	1.005	1.477	1.015	1.005	1.477	1.011
0.25	1.052	1.023	1.339	1.050	1.023	1.339	1.019
0.5	1.193	1.087	1.241	1.183	1.087	1.241	1.036
1.0	1.551	1.242	1.281	1.528	1.241	1.281	1.071
1.5	1.826	1.350	1.372	1.795	1.348	1.371	1.091
2.0	1.994	1.406	1.435	1.956	1.404	1.433	1.106
∞	2.304	1.498	1.552	2.256	1.496	1.549	1.137

Table 8.5: MSE–Inefficiency for Andrews, Tukey and MM2 estimators.

8.8 One-Step Construction

This section treats only the optimal AL estimators. We do not verify if the one-step construction is also applicable for the other estimators introduced throughout this chapter.
Since the normal location and scale model forms an exponential family of full rank with sufficiently smooth parameter functions α and β (cf. Beispiel 1.151 of Witting (1985) or Example 5.1 of Lehmann and Casella (1998), respectively), Lemma 2.3.6 applies. Thus, we can construct the optimally robust AL estimators by means of the one-step method. As starting estimator we provide the Kolmogorov(–Smirnov) minimum distance estimator which has the necessary properties (strict and $\sqrt{n}$ consistent on $\mathcal{U}_\kappa(\theta) \supset \mathcal{U}_v(\theta) \supset \mathcal{U}_c(\theta)$) by Subsection 6.3.2 of Rieder (1994) and is well computable in this model.

Remark 8.8.1 In view of Subsection 2.3.4 we may also use the median together with the MAD as initial estimators in the (normal) location and scale model. ////

8.9 Implementation Using R

The optimal AL estimators can be computed via our R package ROptEst which is in detail described in Appendix D.3. In case of normal location and scale, we provide the generating function NormLocationScaleFamily which is called with some parameters mean $= \mu$ and sd $= \sigma$ ($\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$). The rest of the implementation including the calls to the methods for the computation of the optimal ICs and of the least favorable radii is completely analogous to the models treated in Part II. Thus, we refer to Section 3.5 where the implementation of the binomial model is specified in detail.
We also included the computation of the optimally robust ICs for the AL estimators in our R package RobLox which provides functions for the computation of the optimal ICs for all estimators introduced in this chapter. We implemented functions of the form rlsOptIC.Est where Est has to be replaced by the short form of the corresponding estimators; i.e., Est = AL, M, BM, Hu1, ... Given some neighborhood radius $r \in (0, \infty)$, these functions are called via

```
> IC1 <- rlsOptIC.Est(r = r)
```

One can also specify additional parameters in the calls for these functions but in most cases the default values work fine. For more details about these further parameters we refer to the help pages of package RobLox. The rlsOptIC.Est functions return an object of class ContIC in case of the AL estimators, respectively an object of class IC in case of the other estimators; i.e., package RobLox requires the package ROptEst.

Remark 8.9.1 **(a)** The computation of the optimal robust ICs in case of the AL estimators using the method optIC of package ROptEst (takes about 8 sec. on an AMD Athlon with 2.5 GHz and 512 MB RAM using R 2.0.1; cf. R Development Core Team (2005)) is about 10–15 times slower than via the function rlsOptIC.AL of package RobLox (takes about 0.6 sec.). Hence, we can say, this is the price we have to pay for the greater generality that optIC works for various models; for instance, for all L_2 differentiable models which are based on a univariate distribution.

(b) Since the functions rlsOptIC.Est return an object of class IC, all generic functions of the package ROptEst with IC in its signature can be applied to the corresponding results. In particular, one can call infoPlot which generates some plots of the absolute and relative information for a specified IC. ////

The algorithm for the computation of the optimally robust ICs in case of the AL estimators is analogous to Algorithm D.3.1. In case of the M estimators we use a iteration algorithm which is constructed analogously to the AL case. However, there is an additional outer optimization in γ necessary which is done using the R function optimize; confer R Development Core Team (2005).

Remark 8.9.2 The computation of the optimal IC for the M estimators is numerically instable for radii $r > 1.0$ and one has to try several starting values for

α_1 and α_3 as well as to restrict the interval of γ values to be searched for the minimum, to obtain the corresponding results. ////

In case of the BM estimators α and γ can be calculated by the equations (8.4.2) and (8.4.3). To find the optimal clipping bounds b_{loc} and $b_{\text{sc},0}$, we use the `R` function `optim` with `method = "Nelder-Mead"` (cf. R Development Core Team (2005)). In addition, we introduce the parameter `MAX` to control the constraints on b_{loc} and $b_{\text{sc},0}$ as described below.
If there is only one tuning constant as in case of the Hu1, HuMad, An1, AnMad, Tu1 and TuMad estimators, the computations are based on the `R` function `optimize` (cf. R Development Core Team (2005)) whereas if there are two or more parameters as in case of Hu2a, Hu2, Hu3, Ha3, Ha4, HaMad, An2, Tu2, and MM2 estimators, we use the `R` function `optim` with `method = "Nelder-Mead"`; confer R Development Core Team (2005). To control the constraints on the tuning constants included in the ICs, we introduce an additional parameter `MAX`. That is, if `optim` uses values beyond the admitted ones, the function which is minimized returns `MAX`. Thus, `MAX` should be (clearly) larger than the expected minimax asymptotic MSE to obtain valid results. The default `MAX =` 100 works well in most cases.
In case of the AL, M and BM estimators the maximum asymptotic bias can easily be obtained whereas in case of the other estimators we use different strategies. In case of Hu1, Hu2a, Hu2, Hu3, HuMad, Ha4, HaMad, An2, AnMad, Tu2 and TuMad estimators we can restrict the set of possible maxima to a few points (between 1 and 6 points) by analytic calculations and then take the maximum over these few points to compute the maximum asymptotic bias. In case of the Ha3, An1, Tu1 and MM2 estimators we first evaluate the length of the IC over a grid of width 0.01 where we can restrict this grid to non-negative values by symmetry. Since the maximum asymptotic bias corresponds to the maximum length of the IC, we in a second step take the grid point x which leads to largest length and use the `R` function `optimize` (cf. R Development Core Team (2005)) on the interval $[x - 0.01, x + 0.01]$ to determine the maximum asymptotic bias.

Remark 8.9.3 After installing our `R` bundle `RobASt` one can find the `R` script `NormalLocationScaleModel.R`, which contains some examples for normal location and scale using our `R` package `ROptEst`, in the directory
"`.../RHome/library/ROptEst/scripts/`" where `RHome` stands for the local home directory of `R`. In addition, there are example sections in the help files of the functions `rlsOptIC.Est` which can be executed via

```
> example(rlsOptIC.Est)
```

where `Est = AL, M, BM, Hu1, ....` ////

Part IV

Robust Adaptivity

In his well known paper, Stein (1956) considers estimation and testing of a finite-dimensional Euclidean parameter θ in the presence of an infinite-dimensional nuisance parameter ν. He derives a simple necessary condition for adaptivity, which is the diagonal form of the Fisher information of imbedded finite-dimensional parametric models. Subject to suitable constructions, classical adaptivity means the estimation (testing) of θ with unknown ν is asymptotically no harder than the estimation (testing) of θ with known ν. The necessity result of Stein (1956) is picked up by Bickel (1982) who obtains sufficient conditions under which adaptive estimates exist (cf. Theorems 3.1 and 3.2, ibid.). A very detailed treatment of adaptivity in semiparametric models is given in Bickel et al. (1998).

Since semiparametric models, which rely on strict assumptions like symmetry, may be enlarged to neighborhood models, the issue of adaptation also arises in robust statistics. But, since the classical scores is no longer optimal in that context, one has to think about the meaning of robust adaptivity a new.

In our opinion, the definition of robust adaptivity by means of the same value of two robust optimization problems is most convincing. With this definition, adaptivity is no longer only a dichotomous criterion but, in contrast to previous literature, now has a quantitative meaning, too. Generally speaking, whether we have robust adaptivity or not does not depend on a given starting radius $r \in (0, \infty)$. But, the situation may differ for $r = 0$ (classical adaptivity) or the limiting case $r \to \infty$, respectively.

In Section 9.1 we define adaptation in terms of two asymptotic MSE problems. That is, by considering the MSE–inefficiencies between the corresponding solutions we get an expression for the amount of non-adaptivity.

In this thesis, we restrict our considerations to finite-dimensional parameters, however, the notion of robust adaptivity may easily be extended to neighborhood models with infinite-dimensional parameters; confer Section 6.1 of Rieder (2003) and Section 2 of Shen (1995) (used implicitly), respectively.

In the current chapter various combinations of classical and robust adaptivity arise. First, there are models which are classically as well as robust–adaptive. Second, we give examples where we have classical adaptivity but no robust adaptivity and, third, we treat models which are neither classically nor robust-adaptive. Our study of adaptivity is supplemented by numerical evaluations of the amount of non-adaptivity.

In Section 9.2 we consider the linear model with random regressors where we additionally consider neighborhoods about the ideal model. These are unconditional $(* = c,\ t = 0)$, average conditional $(* = c,\ t = \alpha = 1)$ and average square conditional $(* = c,\ t = \alpha = 2)$ contamination neighborhoods as well as average conditional total variation neighborhoods $(* = v,\ t = \alpha = 1)$.

First, we investigate robust adaptivity in case of linear regression with scale; confer Subsection 9.2.1. Assuming symmetric ideal error distribution F, this model is classically adaptive with respect to scale. If we change our point of view and consider scale as main parameter and the regression parameter as nuisance, we again obtain classical adaptivity. Now, the ideal model is equipped with the neighborhoods mentioned above. Due to the symmetry of the ideal error distribution F,

adaptivity with respect to scale indeed extends to these neighborhoods.

In case of unconditional $(* = c,\ t = 0)$ and average conditional $(* = c,\ t = \alpha = 1)$ contamination neighborhoods the solutions for the simultaneous estimation of regression and scale are derived in Subsection 7.2.1. With these solutions on hand we can also investigate robust adaptivity in case scale is the main parameter and the regression parameter is nuisance. As it turns out, the linear model equipped with these neighborhoods is robust–adaptive with respect to the regression parameter, too.

Second, we treat regression with intercept where we again assume symmetric ideal error distribution F; confer Subsection 9.2.2. In addition, we assume asymmetric ideal regressor distribution K. Under these assumptions the linear model is classically but not necessarily robust–adaptive with respect to intercept. In case of average square conditional contamination neighborhoods $(* = c,\ t = \alpha = 2)$ we indeed have robust adaptivity with respect to intercept. However, this is not true in connection with the larger unconditional $(* = c,\ t = 0)$ and average conditional $(* = c,\ t = \alpha = 1)$ contamination neighborhoods as well as with average conditional total variation neighborhoods $(* = v,\ t = \alpha = 1)$.

In Subsubsection 9.2.2.5 we give some numerical results for the amount of non-adaptivity in case of the linear model with average conditional contamination neighborhoods $(* = c,\ t = \alpha = 1)$. For these computations we assume ideal error distribution $F = \mathcal{N}(0, 1)$ and consider some simple one-dimensional regressor distributions K. As we see, already in these exemplary situations the amount of robust non-adaptivity in terms of MSE–inefficiency may become quite large. We obtain subefficiencies up to 300%. Moreover, the numerical results show that the amount of non-adaptivity is not necessarily isotone in the starting radius r. In one of the examples we even get robust adaptivity in the limiting case $r \to \infty$; confer Example 9.2.26.

These computations as well as the numerical determination of the optimal influence curves in case of linear regression can be done with the help of our `R` package `ROptRegTS` which is part of our `R` bundle `RobASt`; confer Appendix D. In Subsection 9.2.3 we give a short description of this `R` package. The implementation is based on `S4` classes and methods as introduced in Chambers (1998). Further details about the implementation of linear regression models can be found in Subsection 7.6.1 and Appendix D.4, respectively.

In Section 9.3 we investigate adaptivity of time series models. As we see in Appendix A some of these models which have a certain regression structure can be treated analogously to linear models. Important examples are the ARMA(p, q) and the ARCH(p) model which we study in combination with average $(* = c,\ t = \alpha = 1)$ and average square $(* = c,\ t = \alpha = 2)$ contamination neighborhoods of transition probabilities as well as with average total variation neighborhoods $(* = v,\ t = \alpha = 1)$ of transition probabilities.

First, we consider the ARMA(p, q) model with shift; confer Subsection 9.3.1. In fact, this model has a similar structure as the linear model with intercept treated in Subsection 9.2.2.

Assuming the ideal distribution of the innovations F has expectation $\mu_F = 0$, the ARMA(p, q) model is classically adaptive with respect to shift. Adaptivity extends to neighborhoods in case of average square conditional contamination neighborhoods $(* = c,\ t = \alpha = 2)$; confer Subsubsection 9.3.1.1. However, in case of the larger average conditional contamination $(* = c,\ t = \alpha = 1)$ and total variation $(* = v,\ t = \alpha = 1)$ neighborhoods we need additional assumptions to get robust adaptivity with respect to shift. If we additionally assume a symmetric ideal innovation distribution F, robust adaptivity indeed extends; confer Subsubsections 9.3.1.2 and 9.3.1.3, respectively.

We then change our point of view and consider shift as main parameter whereas the ARMA parameters are nuisance. This model may be regarded as a generalization of the classical i.i.d. location model where the error additionally has an ARMA structure. Again, we obtain classical adaptivity as a consequence of $\mu_F = 0$. Moreover, we have robust adaptivity in case of average square conditional contamination neighborhoods $(* = c,\ t = \alpha = 2)$. But, this time, robust adaptivity also holds in case of the larger average conditional contamination $(* = c,\ t = \alpha = 1)$ and total variation $(* = v,\ t = \alpha = 1)$ neighborhoods (without additional assumptions).

The amount of robust non-adaptivity is numerically evaluated in case of AR(1), respectively MA(1) with shift and Gumbel distributed innovations. As we see, the amount in terms of the MSE efficiency loss is only small. For all considered parameter values the subefficiency stays below 3%; confer Subsubsection 9.3.3.1.

Second, we investigate robust adaptivity of the ARCH(p) model which was introduced by Engle (1982); confer Subsection 9.3.2. Again, we consider average $(* = c,\ t = \alpha = 1)$ and average square $(* = c,\ t = \alpha = 2)$ contamination neighborhoods of transition probabilities as well as average total variation neighborhoods $(* = v,\ t = \alpha = 1)$ of transition probabilities. This model is not classically adaptive with respect to innovation scale. Moreover, in connection with the cited neighborhoods it is also not robust-adaptive with respect to innovation scale.

Numerical results in case of the ARCH(1) model with lognormal innovations are given in Subsubsection 9.3.3.2. As it turns out, the amount of non-adaptivity may be very large in this setup. In particular, if we choose parameters close to the boundary of the (strict) stationarity region of the considered ARCH(1) process, we obtain really huge subefficiencies ($> 1000\%$); confer Table 9.8.

In case of times series models, we have, so far, only implemented the special cases considered in Subsection 9.3.3 not using any object orientation. The implementation of AR(1) and ARCH(1) is described in Subsection 9.3.4. However, these special regression-type time series models fit well in the object orientated framework of our `R` package `ROptRegTS` (cf. Appendix D.4) and will be integrated in the near future.

Chapter 9

Robust Adaptivity

First, we introduce the notions of classical and robust adaptivity; confer Section 9.1. In Section 9.2 we investigate classical and robust adaptivity of regression models with unknown error scale (cf. Subsection 9.2.1), respectively unknown intercept (cf. Subsection 9.2.2). Subsequently, we study classical and robust adaptivity of ARMA (p, q) time series models with unknown shift (cf. Subsection 9.3.1) and of ARCH (p) time series models with unknown innovation scale (cf. Subsection 9.3.2).

9.1 Introduction

Assume an L_2 differentiable parametric family

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}) \tag{9.1.1}$$

of probability measures on some sample space $(\Omega, \mathcal{A})$, whose parameter space Θ is an open subset of some finite dimensional $\mathbb{R}^k$ $(k \geq 1)$. The parameter $\theta = (\theta_1^\tau, \theta_2^\tau)^\tau$ decomposes in a main parameter $\theta_1 \in \mathbb{R}^p$ $(p \leq k)$ and a nuisance parameter $\theta_2 \in \mathbb{R}^{k-p}$. Moreover, let $\Lambda_\theta = (\Lambda_{\theta_1}, \Lambda_{\theta_2})^\tau$ be the L_2 derivative of $\mathcal{P}$ in $\theta = (\theta_1, \theta_2)^\tau$ and

$$\mathcal{I}_\theta = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix} \succ 0 \tag{9.1.2}$$

the corresponding Fisher information matrix where

$$\mathcal{I}_{11} = \mathrm{E}\, \Lambda_{\theta_1} \Lambda_{\theta_1}^\tau \qquad \mathcal{I}_{21}^\tau = \mathcal{I}_{12} = \mathrm{E}\, \Lambda_{\theta_1} \Lambda_{\theta_2}^\tau \qquad \mathcal{I}_{22} = \mathrm{E}\, \Lambda_{\theta_2} \Lambda_{\theta_2}^\tau \tag{9.1.3}$$

Hence, $\mathcal{I}_\theta$ is a positive definite and symmetric block matrix. In case the nuisance parameter θ_2 is known, $\mathcal{I}_{11}^{-1}$ (by the Cramér-Rao bound) is the smallest (in the positive definite sense) covariance which can be achieved for the estimation of θ_1. Whereas, in case θ_2 is unknown, the smallest covariance for the estimation of θ_1 is $\mathcal{I}_{11.2}^{-1}$ where

$$\mathcal{I}_{11.2} = \mathcal{I}_{11} - \mathcal{I}_{12} \mathcal{I}_{22}^{-1} \mathcal{I}_{21} \tag{9.1.4}$$

Then, the parametric model $\mathcal{P}$ is called **adaptive** iff $\mathcal{I}_{11}^{-1} = \mathcal{I}_{11.2}^{-1}$; i.e., the estimation of θ_1 with unknown θ_2 is asymptotically no harder than the estimation of θ_1 for known θ_2. This classical notion of adaptivity dates back to Stein (1956) where he derives the necessary condition

$$\mathcal{I}_{12} = 0 \qquad \text{for all } \theta \in \Theta \tag{9.1.5}$$

Remark 9.1.1 For more details on the classical notion of adaptivity we refer to Stein (1956), Bickel (1982) and Section 2.4 of Bickel et al. (1998), respectively. ////

Now, let $\theta_0 = (\theta_{1,0}^\tau, \theta_{2,0}^\tau)^\tau$ be a fixed parameter value in Θ and consider infinitesimal neighborhoods (cf. Sections 1.2 and Appendix A.1) about P_{θ_0} with starting radius $r \in [0, \infty)$ and corresponding bias terms ω as introduced in Subsection 1.3.2 and Appendix A.2, respectively more generally in Subsection 5.3.1 and Section 7.3 of Rieder (1994). We define robust adaptivity by means of the following two MSE-problems where the side conditions guarantee that the results are (partial) ICs (expectation is taken under P_{θ_0}).

(P1) $\mathrm{E}\,|\eta|^2 + r^2\omega^2(\eta) = \min!$ s.t. $\eta \in L_2^p(P_{\theta_0})$, $\mathrm{E}\,\eta = 0$, $\mathrm{E}\,\eta\Lambda_1^\tau = \mathbb{I}_p$
i.e., the nuisance parameter $\theta_2 = \theta_{2,0}$ is known and only θ_1 has to be estimated.

(P2) $\mathrm{E}\,|\eta|^2 + r^2\omega^2(\eta) = \min!$ s.t. $\eta \in L_2^p(P_{\theta_0})$, $\mathrm{E}\,\eta = 0$, $\mathrm{E}\,\eta\Lambda_1^\tau = \mathbb{I}_p$ **and** $\mathrm{E}\,\eta\Lambda_2^\tau = 0$
i.e., the nuisance parameter θ_2 is unknown and has to be estimated together with θ_1.

Definition 9.1.2 *The parametric model $\mathcal{P}$ is called* robust-adaptive, *if the solutions to problem (P1) and problem (P2) coincide for all $r \in [0, \infty)$; i.e.,*

$$\text{minmaxMSE}\,(\text{P1}) = \text{minmaxMSE}\,(\text{P2}) \tag{9.1.6}$$

This is the case iff the solution to problem (P1) is already orthogonal to Λ_2.

Remark 9.1.3 **(a)** Obviously, $\text{minmaxMSE}(P1) \le \text{minmaxMSE}(P2)$ as the side conditions are less restrictive in case of problem (P1). Moreover, the solutions to problem (P1) and (P2) are unique. Hence, the solution to problem (P2) coincides with the solution to problem (P1) iff the solution to problem (P1) is already orthogonal to Λ_2.

(b) In case $r = 0$, the unique solutions to problems (P1) and (P2) are

$$\hat{\eta}_1 = \mathcal{I}_{11}^{-1}\Lambda_1 \qquad \text{and} \qquad \hat{\eta}_2 = (\mathbb{I}_p, 0)\mathcal{I}_\theta^{-1}\Lambda_\theta = \mathcal{I}_{11.2}^{-1}\Lambda_{1.2} \tag{9.1.7}$$

respectively, where $\Lambda_{1.2} = \Lambda_1 - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\Lambda_2$; confer also p 117 of Rieder (1994). Thus, we have classical adaptivity iff $\mathcal{I}_{12} = \mathrm{E}\,\Lambda_1\Lambda_2^\tau = 0$; i.e., $\mathcal{I}_{11}^{-1} = \mathcal{I}_{11.2}^{-1}$.

(c) In case of conditionally centered ICs robust adaptivity is defined by simply replacing $\mathrm{E}\,\eta = 0$ by $\mathrm{E}_\bullet\,\eta = 0$.

(d) Robust adaptivity may also be formulated by the corresponding Hampel type problems (1.3.7) and (A.2.1).

(e) The definition of robust adaptivity can be generalized to semiparametric models, where the nuisance parameter may be infinite dimensional and then generalizes the semiparametric notion of adaptivity by Bickel et al. (1998), p 94.

(f) Whether or not we have robust adaptivity does not depend on the given starting radius $r \in (0,\infty)$, since the forms of the solutions stay the same for all $r \in (0,\infty)$. However, this fact does not extend to $r \in [0,\infty]$ as the examples in this chapter show. ////

We express the amount of non-adaptivity in terms of the MSE–inefficiency, respectively MSE–subefficiency of the solution to problem (P2) with respect to the solution to problem (P1) where

$$\text{MSE–inefficieny} = \frac{\text{minmaxMSE}\,(\text{P2})}{\text{minmaxMSE}\,(\text{P1})} \tag{9.1.8}$$

and MSE–subefficiency = MSE–inefficieny $-\,1$, respectively.

9.2 Regression Models

9.2.1 Regression and Scale

We consider the linear regression model with regression parameter β and unknown scale parameter σ of the form

$$y_i = x_i^\tau \beta + \sigma u_i \qquad (i = 1,\dots,n) \tag{9.2.1}$$

where $x_1,\dots,x_n$ are i.i.d. realizations of the regressor x distributed according to some probability K on $\mathbb{B}^k$, and $u_1,\dots,u_n$ are i.i.d. copies of the error $u \sim F$. It is assumed that $x \sim K$ and $u \sim F$ are stochastically independent. Furthermore we make the following assumptions on F and K in the ideal case,

(F1) F is symmetric.

(F2) The Fisher information of location for F is finite; i.e., F has an absolutely continuous density f and $\mathcal{I}_F^{\mathrm{loc}} = \int (\Lambda_f^{\mathrm{loc}})^2 f\,d\lambda < \infty$ where $\Lambda_f^{\mathrm{loc}} := -f'/f$; confer Proposition 5.1.4.

(F3) The Fisher information of scale for F is finite; i.e., $u \mapsto uf(u) =: uf$ is absolute continuous and $\mathcal{I}_F^{\mathrm{sc}} = \int (\Lambda_f^{\mathrm{sc}})^2 f\,d\lambda < \infty$ where $\Lambda_f^{\mathrm{sc}} := u\Lambda_f^{\mathrm{loc}} - 1$; confer Proposition 5.1.1.

(K) $\mathcal{K} := \int xx^\tau K(dx) \in \mathbb{R}^{k\times k}$ has $\mathrm{rk}(\mathcal{K}) = k$.

Remark 9.2.1 **(a)** This setup is identical to the setup considered in Chapter 7. Thus, we refer to this previous chapter particularly to Section 7.1 for more details about this model.

(b) If we consider scale as known nuisance parameter, the classical scores is $\hat{\eta}_{\mathrm{rg}} = \mathcal{K}^{-1} x \Lambda_f^{\mathrm{loc}}$. Due to (F1), we obtain

$$\mathrm{E}\, \Lambda_f^{\mathrm{loc}} \Lambda_f^{\mathrm{sc}} = \mathrm{E}\, u (\Lambda_f^{\mathrm{loc}})^2 \overset{(\mathrm{F1})}{=} 0 \tag{9.2.2}$$

Thus, the linear regression model is classically adaptive with respect to scale. If we change our point of view and consider σ as main and β as nuisance parameter, the classical scores is $\hat{\eta}_{\mathrm{sc}} = (\mathcal{I}_F^{\mathrm{sc}})^{-1} \Lambda_f^{\mathrm{sc}}$. By (F1), we again get

$$\mathrm{E}\, \Lambda_f^{\mathrm{sc}} x^\tau \Lambda_f^{\mathrm{loc}} = \mathrm{E}\, x^\tau\, \mathrm{E}\, u (\Lambda_f^{\mathrm{loc}})^2 \overset{(\mathrm{F1})}{=} 0 \tag{9.2.3}$$

Therefore, the linear regression model is also classically adaptive with respect to the regression parameter. Without symmetry of F, the linear regression model is not classical adaptive with respect to scale nor with respect to the regression parameter, in general. As we will see throughout this subsection, the symmetry assumption (F1) is also essential for the verification of robust adaptivity. ////

9.2.1.1 Unconditional Contamination Neighborhoods

We consider unconditional (errors-in-variables) infinitesimal contamination neighborhoods (1.2.4) (i.e., $* = c,\ t = 0$) with starting radius $r \in (0, \infty)$.

Linear Regression with main parameter β and nuisance parameter σ

We assume the regression parameter β has to be estimated where the scale parameter $\sigma = \sigma_0 = 1$ is known; i.e., we are interested in problem (P1). The solution to this problem can be read off from Theorem 1.3.11

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}} (x \Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{rg}}) \min \Big\{ 1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}} (x \Lambda_f^{\mathrm{loc}}(u) - z_{\mathrm{rg}})|} \Big\} \tag{9.2.4}$$

where we get $z_{\mathrm{rg}} = 0$ as a possible choice by $(F1)$. Hence, the solution specializes to

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u) \min \Big\{ 1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}(u)|} \Big\} \tag{9.2.5}$$

with

$$A_{\mathrm{rg}} = \Big[\mathrm{E} (\Lambda_f^{\mathrm{loc}})^2 x x^\tau \min \Big\{ 1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}|} \Big\} \Big]^{-1} \tag{9.2.6}$$

and clipping bound $b_{\mathrm{rg}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 b_{\mathrm{rg}} = \mathrm{E} \left(|A_{\mathrm{rg}} x \Lambda_f^{\mathrm{loc}}| - b_{\mathrm{rg}} \right)_+ \tag{9.2.7}$$

Remark 9.2.2 By Theorem 2.4.1 we get the following connection between the Lagrange multipliers contained in the solutions at $\beta_0 = 0$ and β

$$A_{\mathrm{rg},\beta} = A_{\mathrm{rg},\beta_0} \qquad b_{\mathrm{rg},\beta} = b_{\mathrm{rg},\beta_0} \tag{9.2.8}$$

confer also Remark 2.4.2 (f). ////

Proposition 9.2.3 *Assume (F1). Then, the linear model with unconditional contamination neighborhoods ($* = c$, $t = 0$) is robust–adaptive with respect to scale.*

Proof It holds

$$\mathrm{E}\,\tilde{\eta}_{\mathrm{rg}}\Lambda_f^{\mathrm{sc}} = \mathrm{E}\,\tilde{\eta}_{\mathrm{rg}}(u\Lambda_f^{\mathrm{loc}} - 1) = \mathrm{E}\,\tilde{\eta}_{\mathrm{rg}}u\Lambda_f^{\mathrm{loc}} = 0 \tag{9.2.9}$$

since $\tilde{\eta}_{\mathrm{rg}}$ and Λ_f^{loc} are odd functions in u, by condition (F1). ////

Remark 9.2.4 **(a)** Without assumption (F1), the linear model with unconditional contamination neighborhoods ($* = c$, $t = 0$) is not robust–adaptive with respect to scale, in general.

(b) As a consequence of the previous proposition, the solutions to problem (P1) and (P2) coincide and it was no restriction to assume $\sigma = \sigma_0 = 1$ to be known.

(c) Under elliptical, respectively spherical symmetry of the ideal regressor distribution K, this solution can be further specialized analogously to Proposition 7.2.3 and 7.2.4, respectively. ////

Now, we change our point of view.

Linear Regression with main parameter σ and nuisance parameter β

We want to estimate σ and consider β as a nuisance parameter where we assume $\beta = \beta_0 = 0$ to be known. That is, we have to solve a problem of form (P1). The unique solution to this problem is provided by Theorem 1.3.11 and reads

$$\tilde{\eta}_{\mathrm{sc}}(u) = A_{\mathrm{sc}}(\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}})w(u) \qquad w(u) = \min\Big\{1, \frac{c_{\mathrm{sc}}}{|\Lambda_f^{\mathrm{sc}}(u) - z_{\mathrm{sc}}|}\Big\} \tag{9.2.10}$$

with

$$z_{\mathrm{sc}} = \mathrm{E}\,\Lambda_f^{\mathrm{sc}} w_{\mathrm{sc}}\big[\mathrm{E}\,w_{\mathrm{sc}}\big]^{-1} \qquad A_{\mathrm{sc}} = \big[\mathrm{E}(\Lambda_f^{\mathrm{sc}} - z_{\mathrm{sc}})^2 w_{\mathrm{sc}}\big]^{-1} \tag{9.2.11}$$

where the clipping bound $c_{\mathrm{sc}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 c_{\mathrm{sc}} = \mathrm{E}\,\big(|\Lambda_f^{\mathrm{sc}}| - c_{\mathrm{sc}}\big)_+ \tag{9.2.12}$$

confer also Subsection 5.2.1.

Remark 9.2.5 By Theorem 2.4.1 we get the following connection between the Lagrange multipliers contained in the solutions at $\sigma_0 = 1$ and σ

$$A_{\mathrm{sc},\sigma} = \sigma^2 A_{\mathrm{sc},\sigma_0} \qquad z_{\mathrm{sc},\sigma} = \sigma^{-1} z_{\mathrm{sc},\sigma_0} \qquad b_{\mathrm{sc},\sigma} = \sigma b_{\mathrm{sc},\sigma_0} \tag{9.2.13}$$

confer also Remark 2.4.2 (f). ////

Proposition 9.2.6 *Assume (F1). Then, the linear model with unconditional contamination neighborhoods ($* = c,\ t = 0$) is robust–adaptive with respect to the regression parameter.*

PROOF Analogously to the proof of Proposition 9.2.3 we get

$$\mathrm{E}\,\tilde{\eta}_{\mathrm{sc}}\Lambda_f^{\mathrm{loc}}x^\tau = 0 \tag{9.2.14}$$

by condition (F1). ////

Remark 9.2.7 **(a)** Without assumption (F1), the linear model with unconditional contamination neighborhoods ($* = c,\ t = 0$) is not robust–adaptive with respect to the regression parameter, in general.

(b) As a consequence of the previous proposition, the solutions to problem (P1) and (P2) coincide and it was no restriction to assume $\beta = \beta_0 = 0$ to be known. ////

9.2.1.2 Average Square Conditional Contamination Neighborhoods

We consider average square conditional (error-free-variables) contamination neighborhoods (i.e., $* = c,\ t = \alpha = 2$) with starting radius $r \in (0, \infty)$. These are neighborhoods of form (7.1.9) with $\mathrm{E}\,\varepsilon^2 \le 1$.

Linear Regression with main parameter β and nuisance parameter σ

We assume the regression parameter β has to be estimated where the scale parameter $\sigma = \sigma_0 = 1$ is known. That is, we have to solve problem (P1). The solution to this problem is provided by Theorem A.2.5 and reads

$$\tilde{\eta}(x,u) = A\mathcal{K}^{-1}x\big(\Lambda_f^{\mathrm{loc}}(u) - z\big)\min\Big\{1, \frac{c}{|\Lambda_f^{\mathrm{loc}}(u) - z|}\Big\} \tag{9.2.15}$$

By (F1), $z = 0$ is a possible choice; i.e., the solution simplifies to

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}}\mathcal{K}^{-1}x\Lambda_f^{\mathrm{loc}}(u)\min\Big\{1, \frac{c_{\mathrm{rg}}}{|\Lambda_f^{\mathrm{loc}}(u)|}\Big\} \tag{9.2.16}$$

where

$$A_{\mathrm{rg}}^{-1} = \mathrm{E}(\Lambda_f^{\mathrm{loc}})^2\min\Big\{1, \frac{c_{\mathrm{rg}}}{|\Lambda_f^{\mathrm{loc}}(u)|}\Big\} \qquad b_{\mathrm{rg}} = c_{\mathrm{rg}}A_{\mathrm{rg}}\sqrt{\mathrm{tr}\,\mathcal{K}^{-1}} \tag{9.2.17}$$

and clipping bound $c_{\mathrm{rg}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 c_{\mathrm{rg}} = \mathrm{E}(|\Lambda_f^{\mathrm{loc}}| - c_{\mathrm{rg}})_+ \tag{9.2.18}$$

Proposition 9.2.8 *Assume (F1). Then, the linear model with average square conditional contamination neighborhoods ($* = c,\ t = \alpha = 2$) is robust–adaptive with respect to scale.*

PROOF Analogous to the proof of Proposition 9.2.3. ////

Remark 9.2.9 **(a)** Without assumption (F1), the linear model with average square conditional contamination neighborhoods $(* = c,\ t = \alpha = 2)$ is not robust–adaptive with respect to scale, in general.

(b) As a consequence of the previous proposition, the solutions to problem (P1) and (P2) coincide and it was no restriction to assume $\sigma = \sigma_0 = 1$ to be known.

(c) Of course, we could also change the point of view in this context of average square conditional contamination neighborhoods $(* = c,\ t = \alpha = 2)$. That is, consider σ as main and β as nuisance parameter. However, to do so, one first has to derive the corresponding optimal solutions. ////

The next question is whether robust adaptivity also holds in the larger neighborhood model $* = c,\ t = \alpha = 1$.

9.2.1.3 Average Conditional Contamination Neighborhoods

We consider average conditional (error-free-variables) contamination neighborhoods (7.1.9) (i.e., $* = c,\ t = \alpha = 1$) with starting radius $r \in (0, \infty)$.

Linear Regression with main parameter β and nuisance parameter σ

We assume the regression parameter β has to be estimated where scale is an additional nuisance parameter.

Remark 9.2.10 Due to (F1), the solution to problem (P1) specified in Theorem A.2.3 coincides with the solution in case of unconditional contamination neighborhoods $(* = c,\ t = 0)$ given in equation (9.2.5) which is conditionally centered; i.e., $\mathrm{E}_\bullet\, \tilde{\eta}_{\mathrm{rg}} = 0$ and $z_{\mathrm{rg}}(x) \equiv 0$ for all x is a possible choice. As a consequence, the corresponding results about robust adaptivity derived in Subsubsection 9.2.1.1 also apply to the linear model with average conditional contamination neighborhoods $(* = c,\ t = \alpha = 1)$. ////

Linear Regression with main parameter σ and nuisance parameter β

We assume scale σ has to be estimated and the regression parameter β is a nuisance parameter.

Remark 9.2.11 The solution to problem (P1) can be read off from Proposition 7.2.6. In the considered setup $z_{\mathrm{sc}}(x) \equiv z_{\mathrm{sc}}$ for all x is a possible choice. That is, the solution in case of average conditional contamination neighborhoods $(* = c,\ t = \alpha = 1)$ coincides with the solution $\tilde{\eta}_{\mathrm{sc}}$ in case of unconditional contamination neighborhoods given in equation (9.2.10). Thus, the corresponding results about robust adaptivity derived in Subsubsection 9.2.1.1 again apply to the linear model with average conditional contamination neighborhoods $(* = c,\ t = \alpha = 1)$, too. ////

9.2.1.4 Average Conditional Total Variation Neighborhoods

We consider average conditional (error-free-variables) total variation neighborhoods $(* = v,\ t = \alpha = 1)$ of starting radius $r \in (0, \infty)$. These neighborhoods, which are also called error-free-variables neighborhoods, for any contamination curve $\varepsilon\colon (\mathbb{R}^k, \mathbb{B}^k) \to (\bar{\mathbb{R}}, \bar{\mathbb{B}})$ such that $\mathrm{E}\,\varepsilon \leq 1$, consist of all probabilities $Q(dx, dy) = Q(dy \,|\, x)Q(dx)$ on $\mathbb{B}^{k+1}$ such that

$$Q(dx) = K(dx) \tag{9.2.19}$$

and

$$Q(dy \,|\, x) \in B_v(P_\theta(dy \,|\, x), r\varepsilon(x)) \qquad \text{a.e. } K(dx) \tag{9.2.20}$$

confer Subsection 7.2.2 of Rieder (1994).

Linear Regression with main parameter β and nuisance parameter σ

We assume the regression parameter β has to be estimated where scale is an additional nuisance parameter.

Remark 9.2.12 **(a)** Due to (F1), the solution to problem (P1) specified in Theorem A.2.7 coincides with the solution $\tilde{\eta}_{\mathrm{rg}}$ in case of unconditional $(* = c,\ t = 0)$, respectively average conditional contamination neighborhoods $(* = c,\ t = \alpha = 1)$ and radius $2r$; confer also Remark A.2.8. That is, the corresponding results about robust adaptivity derived in Subsubsection 9.2.1.1 also apply in case of the linear model with average conditional total variation neighborhoods $(* = v,\ t = \alpha = 1)$.

(b) Of course, we could also change the point of view in this context of average conditional total variation neighborhoods $(* = v,\ t = \alpha = 1)$. That is, consider scale as main and the regression parameter as nuisance parameter. However, to do so, one first has to derive the corresponding optimal solutions. ////

9.2.2 Regression with Intercept

We consider the linear regression model with regression parameter β and intercept parameter μ of the form

$$y_i = \mu + x_i^\tau \beta + u_i \qquad (i = 1, \ldots, n) \tag{9.2.21}$$

where $x_1, \ldots, x_n$ are i.i.d. realizations of the regressor x distributed according to some probability K on $\mathbb{B}^k$, and $u_1, \ldots, u_n$ are i.i.d. copies of the error $u \sim F$. It is assumed that $x \sim K$ and $u \sim F$ are stochastically independent. Furthermore we make the following assumptions on F and K in the ideal case,

(F1) F is symmetric.

(F2) The Fisher information of location for F is finite; i.e., F has an absolutely continuous density f and $\mathcal{I}_F^{\mathrm{loc}} = \int (\Lambda_f^{\mathrm{loc}})^2 f \, d\lambda < \infty$ where $\Lambda_f^{\mathrm{loc}} := -f'/f$; confer Proposition 5.1.4.

(K1) K is asymmetric.

(K2) $\mu_K = \int x\, K(dx) = 0$

(K3) $\mathcal{K} := \int xx^\tau K(dx) \in \mathbb{R}^{k\times k}$ has $\mathrm{rk}(\mathcal{K}) = k$.

Remark 9.2.13 The assumptions (K2) and (K3) guarantee the positiv definiteness of $\mathrm{Cov}\,(X)$, which by the following lemma is equivalent to the identifiability of model (9.2.21). In case of linear regression without intercept, we have identifiability without condition (K2); confer p 68 of Rieder (1994). ////

Lemma 9.2.14 *(a) Model (9.2.21) is identifiable iff for all* $\theta = (\beta^\tau, \mu)^\tau \in \mathbb{R}^{k+1}$,

$$\tilde{x}^\tau\theta = 0 \qquad \text{for almost all } \tilde{x} \quad \Longrightarrow \quad \theta = 0 \tag{9.2.22}$$

where $\tilde{x} = (x^\tau, 1)^\tau$.

(b) Positiv definiteness of $\mathrm{Cov}\,(X)$ *is equivalent to (9.2.22).*

Proof

(a) Analogous to Lemma 2.4.4 of Rieder (1994).

(b) (9.2.22) is equivalent to

$$\tilde{x}\tilde{x}^\tau\theta = 0 \qquad \text{for almost all } \tilde{x} \quad \Longrightarrow \quad \theta = 0 \tag{9.2.23}$$

which is equivalent to

$$\mathrm{E}\,\tilde{x}\tilde{x}^\tau\theta = 0 \quad \Longrightarrow \quad \theta = 0 \tag{9.2.24}$$

The symmetric matrix

$$\mathrm{E}\,\tilde{x}\tilde{x}^\tau = \begin{pmatrix} \mathrm{E}\,xx^\tau & \mathrm{E}\,x \\ \mathrm{E}\,x^\tau & 1 \end{pmatrix} \tag{9.2.25}$$

is positive definite iff $\mathrm{E}\,xx^\tau - \mathrm{E}\,x\,\mathrm{E}\,x^\tau = \mathrm{Cov}\,(X)$ is positive definite; confer Corollary 14.8.6 of Harville (1997). ////

The corresponding model distributions with parameter $\theta = (\beta^\tau, \mu)^\tau \in \mathbb{R}^{k+1}$ read

$$P_\theta(dx, dy) = f(y - \mu - x^\tau\beta)\lambda(dy)K(dx) \tag{9.2.26}$$

This model stays invariant under the following transformations,

$$g_\theta(x, y) = (x, \mu + x^\tau\beta + y) \tag{9.2.27}$$

in the sense that, for all $\theta \in \mathbb{R}^{k+1}$, P_θ is the image measure

$$P_\theta = g_\theta(P_{\theta_0}) \tag{9.2.28}$$

of P_{θ_0} under the transformation g_θ. We define the scores at $\theta \in \mathbb{R}^{k+1}$,

$$\Lambda_\theta(x, y) := \Lambda_f^{\mathrm{loc}}(u) \begin{pmatrix} x \\ 1 \end{pmatrix} = \Lambda_{\theta_0} \circ g_\theta^{-1}(x, y) \tag{9.2.29}$$

where $u = (y - \mu - x^\tau\beta)$ and $\theta_0 := 0$. Thus, we have for the corresponding Fisher information

$$\mathcal{I}_\theta = \mathrm{E}_\theta\, \Lambda_\theta\Lambda_\theta^\tau = \mathrm{E}_{\theta_0}\, \Lambda_{\theta_0}\Lambda_{\theta_0}^\tau = \mathcal{I}_{\theta_0} \tag{9.2.30}$$

for all $\theta \in \mathbb{R}^{k+1}$ and

$$\mathcal{I}_{\theta_0} = \mathcal{I}_f^{\mathrm{loc}} \begin{pmatrix} \mathcal{K} & \mu_K \\ \mu_K^\tau & 1 \end{pmatrix} \overset{(\mathrm{K2})}{=} \mathcal{I}_f^{\mathrm{loc}} \begin{pmatrix} \mathcal{K} & 0 \\ 0 & 1 \end{pmatrix} \tag{9.2.31}$$

E_θ and E_{θ_0} denote expectation under P_θ and P_{θ_0}, respectively.

Proposition 9.2.15 *The parametric family*

$$\mathcal{P} := \{P_\theta \,|\, \theta = (\beta^\tau, \mu)^\tau \in \mathbb{R}^{k+1}\} \tag{9.2.32}$$

of probability measures on $(\mathbb{R}^{k+1}, \mathbb{B}^{k+1})$ is L_2-differentiable at θ with L_2-derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by (9.2.29) and (9.2.30), respectively.

Proof Analogously to proof of Theorem 2.4.6 in Rieder (1994) using $\tilde{x} = (x^\tau, 1)^\tau$. ////

Remark 9.2.16 (a) If we consider μ as known nuisance parameter, the classical solution is $\hat{\eta}_{\mathrm{rg}} = \mathcal{K}^{-1} x \Lambda_f^{\mathrm{loc}}$. By stochastic independence of x and u and (K2), we obtain

$$\mathrm{E}\, x \Lambda_f^{\mathrm{loc}} \Lambda_f^{\mathrm{loc}} = \mathrm{E}\, x\, \mathrm{E}(\Lambda_f^{\mathrm{loc}})^2 \overset{(\mathrm{K2})}{=} 0 \tag{9.2.33}$$

Thus, the linear model is classically adaptive with respect to intercept μ.

(b) The construction problem in case of ideal error distribution $F = \mathcal{N}(0,1)$ and bounded regressors is solved in Subsection 7.4.1; confer Remark 7.4.1 (b). ////

9.2.2.1 Unconditional Contamination Neighborhoods

We consider unconditional (errors-in-variables) infinitesimal contamination neighborhoods (1.2.4) (i.e., $* = c$, $t = 0$) with starting radius $r \in (0, \infty)$.

Linear Regression with main parameter β and nuisance parameter μ

We assume β is the parameter of interest and the intercept parameter $\mu = \mu_0 = 0$ is a known nuisance parameter; i.e., we have to solve problem (P1). The unique solution provided by Theorem 1.3.11 under (F1) is of form (9.2.5) where clipping bound $b_{\mathrm{rg}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via (9.2.7).

Proposition 9.2.17 *Assume (F1) and (K1). Then, the linear model with unconditional contamination neighborhoods ($* = c$, $t = 0$), is not robust–adaptive with respect to intercept μ.*

Proof It holds

$$\mathrm{E}\, \tilde{\eta}_{\mathrm{rg}} \Lambda_f^{\mathrm{loc}} = A_{\mathrm{rg}}\, \mathrm{E}\, x (\Lambda_f^{\mathrm{loc}})^2 \min\left\{1, \frac{b_{\mathrm{rg}}}{|A_{\mathrm{rg}}(x \Lambda_f^{\mathrm{loc}}(u))|}\right\} \neq 0 \tag{9.2.34}$$

as a consequence of conditions (F1) and (K1). ////

Remark 9.2.18 If K is symmetric, the linear model is robust–adaptive with respect to intercept μ by analogous arguments as in case of the ARMA (p,q) model treated in Subsubsection 9.3.1.2. ////

Second, we specify the solution to problem (P2); i.e., the intercept parameter μ is unknown. Due to (F1), the unique solution given by Theorem 1.3.11 can be rewritten as

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = (A_{\mathrm{rg}}x + A_\mu)\Lambda_f^{\mathrm{loc}}(u)w(x,u) \tag{9.2.35}$$

with

$$w(x,u) = \min\Big\{1, \frac{b_{\mathrm{rg}}}{|(A_{\mathrm{rg}}x + A_\mu)\Lambda_f^{\mathrm{loc}}|}\Big\} \tag{9.2.36}$$

and

$$A_{\mathrm{rg}} = \Big[\mathbb{I}_k - A_\mu \,\mathrm{E}\, x^\tau(\Lambda_f^{\mathrm{loc}})^2 w\Big] \Big/ \mathrm{E}\, xx^\tau(\Lambda_f^{\mathrm{loc}})^2 w \tag{9.2.37}$$

$$A_\mu = -A_{\mathrm{rg}}\,\mathrm{E}\, x(\Lambda_f^{\mathrm{loc}})^2 w \Big/ \mathrm{E}(\Lambda_f^{\mathrm{loc}})^2 w \tag{9.2.38}$$

where the clipping bound $b_{\mathrm{rg}} \in (0,\infty)$ and radius $r \in (0,\infty)$ are related via

$$r^2 b_{\mathrm{rg}} = \mathrm{E}\,\big(|(A_{\mathrm{rg}}x + A_\mu)\Lambda_f^{\mathrm{loc}}| - b_{\mathrm{rg}}\big)_+ \tag{9.2.39}$$

Obviously, the solution to problem (P1) given in equation (9.2.5) is not equal to the solution to problem (P2) given in equation (9.2.35).

Remark 9.2.19 We numerically determined the amount of non-adaptivity in some simple examples. As it turns out, there are situations where the amount of non-adaptivity is very large. In Example 9.2.25 the maximum subefficiency is 300%. For further numerical results we refer to Subsubsection 9.2.2.5. ////

9.2.2.2 Average Square Conditional Contamination Neighborhoods

We consider average square conditional (error-free-variables) contamination neighborhoods (i.e., $* = c$, $t = \alpha = 2$) with starting radius $r \in (0,\infty)$. That is, neighborhoods of form (7.1.9) with $\mathrm{E}\,\varepsilon^2 \le 1$.

Linear Regression with main parameter β and nuisance parameter μ

We assume β is the parameter of interest and the intercept parameter $\mu = \mu_0 = 0$ is a known nuisance parameter; i.e., we have to solve problem (P1). The unique solution provided by Theorem A.2.5 under condition (F1) is of form (9.2.16) where clipping bound $c_{\mathrm{rg}} \in (0,\infty)$ and radius $r \in (0,\infty)$ are related via (9.2.18).

Proposition 9.2.20 *The linear model with average conditional contamination neighborhoods ($* = c$, $t = \alpha = 2$), is robust–adaptive with respect to intercept μ.*

Proof It holds

$$\mathrm{E}\,\tilde{\eta}_{\mathrm{rg}}\Lambda_f^{\mathrm{loc}} = A_{\mathrm{rg}}\mathcal{K}^{-1}\,\mathrm{E}\, x\,\mathrm{E}(\Lambda_f^{\mathrm{loc}})^2 \min\Big\{1, \frac{c_{\mathrm{rg}}}{|\Lambda_f^{\mathrm{loc}}(u)|}\Big\} = 0 \tag{9.2.40}$$

as a consequence of condition (K2). ////

Remark 9.2.21 **(a)** As the proof of the previous proposition indicates, the symmetry assumption (F1) is not needed to obtain robust adaptivity in case $* = c$, $t = \alpha = 2$.

(b) As a consequence of the previous proposition, the solutions to problem (P1) and (P2) coincide and it was no restriction to assume $\mu = \mu_0 = 0$ to be known. ////

The next questions is whether robust adaptivity also holds in the larger neighborhood model $* = c$, $t = \alpha = 1$.

9.2.2.3 Average Conditional Contamination Neighborhoods

We consider average conditional contamination neighborhoods (7.1.9) (i.e., $* = c$, $t = \alpha = 1$) with starting radius $r \in (0, \infty)$.

Linear Regression with main parameter β and nuisance parameter μ

We assume β is the parameter of interest and the intercept parameter $\mu = \mu_0 = 0$ is a further nuisance parameter.

Remark 9.2.22 The solutions to problems (P1) and (P2) in case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) coincide with the solution in case of unconditional contamination neighborhoods ($* = c$, $t = 0$) specified in Subsubsection 9.2.2.1 which are conditionally centered; i.e., $\mathrm{E}_\bullet \tilde{\eta}_{\mathrm{rg}} = 0$. In particular, under (F1) and (K1), the linear model with average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) is not robust–adaptive with respect to intercept μ; confer Proposition 9.2.17. ////

9.2.2.4 Average Conditional Total Variation Neighborhoods

We consider average conditional total variation neighborhoods (9.2.20) (i.e., $* = v$, $t = \alpha = 1$) with starting radius $r \in (0, \infty)$.

Linear Regression with main parameter β and nuisance parameter μ

We assume β is the parameter of interest and the intercept parameter $\mu = \mu_0 = 0$ is a further nuisance parameter.

Remark 9.2.23 Due to (F1), the solutions to problems (P1) and (P2) in case of average conditional total variation neighborhoods ($* = c$, $t = \alpha = 1$) with starting radius $r \in (0, \infty)$ coincide with the solution in case of unconditional ($* = c$, $t = 0$), respectively average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) and radius $2r$; confer also Remarks 9.2.22 and 9.2.2, respectively. The solutions are specified in Subsubsection 9.2.2.1. Consequently, under (F1) and (K1), the linear regression model with average conditional total variation neighborhoods ($* = v$, $t = \alpha = 1$) is not robust–adaptive with respect to intercept μ; confer Proposition 9.2.17. ////

9.2.2.5 Numerical Results

Under (F1), which is the symmetry of the ideal error distribution F, the results in case of unconditional contamination neighborhoods ($* = c,\ t = 0$) with starting radius $r \in (0, \infty)$ coincide with the corresponding results in case average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$); confer Remark 9.2.22. Moreover, these results are also identical to the corresponding results in case of average conditional total variation neighborhoods ($* = v,\ t = \alpha = 1$) with starting radius $2r$; confer Remark 9.2.23. Hence, we only give results in case $* = c,\ t = 0$. In addition, we restrict our considerations to ideal error distribution $F = \mathcal{N}(0,1)$ and simple one-dimensional regressor distributions K. In this setup, the solution to problem (P1) simplifies to

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}} x u \min\left\{1, \frac{c_{\mathrm{rg}}}{|xu|}\right\} \tag{9.2.41}$$

with

$$A_{\mathrm{rg}}^{-1} = \mathrm{E}\,|xu| \min\{|xu|, c_{\mathrm{rg}}\} = 2\,\mathrm{E}\,x^2 \Phi\big(c_{\mathrm{rg}}/|x|\big) - \mathcal{K} \tag{9.2.42}$$

and clipping bound $c_{\mathrm{rg}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 c_{\mathrm{rg}} = \mathrm{E}\left(|xu| - c_{\mathrm{rg}}\right)_+ = 2\,\mathrm{E}\left[|x|\varphi(c_{\mathrm{rg}}/|x|) - c_{\mathrm{rg}}\Phi(-c_{\mathrm{rg}}/|x|)\right] \tag{9.2.43}$$

where Φ and φ are the cumulative distribution function and the density of $\mathcal{N}(0,1)$, respectively. The minimum bias is

$$\omega_{c,t}^{\min} = \sqrt{\tfrac{\pi}{2}} \Big/ \mathrm{E}\,|x|, \qquad t = 0,\ t = \alpha = 1 \tag{9.2.44}$$

The solution to problem (P2) can be rewritten as

$$\tilde{\eta}_{\mathrm{rg}}(x,u) = A_{\mathrm{rg}}(x - M) u\, w(x,u) \qquad w(x,u) = \min\left\{1, \frac{c_{\mathrm{rg}}}{|x - M|\,|u|}\right\} \tag{9.2.45}$$

with

$$A_{\mathrm{rg}}^{-1} = \mathrm{E}\,x(x - M)u^2 w = 2\,\mathrm{E}\,x(x - M)\Phi(c_{\mathrm{rg}}/|x - M|) - \mathcal{K} \tag{9.2.46}$$

$$M = \frac{\mathrm{E}\,x u^2 w}{\mathrm{E}\,u^2 w} = \frac{\mathrm{E}\,x \Phi(c_{\mathrm{rg}}/|x - M|)}{\mathrm{E}\,\Phi(c_{\mathrm{rg}}/|x - M|) - 0.5} \tag{9.2.47}$$

where the clipping bound $c_{\mathrm{rg}} \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 c_{\mathrm{rg}} = \mathrm{E}\left(|x - M|\,|u| - c_{\mathrm{rg}}\right)_+ \tag{9.2.48}$$

$$= 2\,\mathrm{E}\left[|x - M|\varphi(c_{\mathrm{rg}}/|x - M|) - c_{\mathrm{rg}}\Phi(-c_{\mathrm{rg}}/|x - M|)\right] \tag{9.2.49}$$

The minimum bias reads

$$\omega_{c,t}^{\min} = \frac{\sqrt{\frac{\pi}{2}}}{\mathrm{E}\,|x - \mathrm{med}(x)|} \qquad t = 0,\ t = \alpha = 1 \tag{9.2.50}$$

Remark 9.2.24 Since the linear model with intercept is robust–adaptive if K is symmetric, we also specify the skewness γ_1 of K which therefore might be an indicator for the amount of robust non-adaptivity. This is true for K as specified in Example 9.2.25 and also for regressor distribution K used in Example 9.2.27. However, it is not necessarily true as Example 9.2.26 shows. Similarly, the MSE–inefficiency is increasing with increasing radius $r \in (0, \infty)$ at least in the first and third example. But, again Example 9.2.26 shows this is not necessarily true. ////

Example 9.2.25 We first take a closer look at

$$K(dx) = \frac{1+\tau}{2+\tau}\,\mathrm{I}_{\{-1\}}(dx) + \frac{1}{2+\tau}\,\mathrm{I}_{\{1+\tau\}}(dx) \tag{9.2.51}$$

where $\tau \in (0, \infty)$. We obtain

$$\mathrm{E}\,X = 0 \qquad \mathcal{K} = \mathrm{E}\,X^2 = 1+\tau \qquad \gamma_1 = \mathrm{E}\,X^3/\mathcal{K}^{1.5} = \tau/\sqrt{1+\tau} \tag{9.2.52}$$

$$\mathcal{I}^{-1} = \mathcal{K}^{-1} = 1/(1+\tau) \tag{9.2.53}$$

Letting τ tend to infinity, the limiting values in case of problem (P1) can be determined by

$$A_{\mathrm{rg},\infty}^{-1} = 2\big[\Phi(c_{\mathrm{rg},\infty}) + c_{\mathrm{rg},\infty}\varphi(0) - 0.5\big] \tag{9.2.54}$$

and

$$r^2 c_{\mathrm{rg},\infty} = 2\big[\varphi(c_{\mathrm{rg},\infty}) - c_{\mathrm{rg},\infty}\Phi(-c_{\mathrm{rg},\infty}) + \varphi(0)\big] \tag{9.2.55}$$

where

$$\omega_{c,t} = \frac{2+\tau}{2(1+\tau)}\sqrt{\frac{\pi}{2}} \longrightarrow \frac{1}{2}\sqrt{\frac{\pi}{2}} \qquad t=0,\, t=\alpha=1 \tag{9.2.56}$$

In case of problem (P2), we obtain

$$A_{\mathrm{rg},\infty} = \frac{\pi}{2}\,r^2 \qquad M_\infty = -1 \qquad c_{\mathrm{rg},\infty} = r^{-2}\sqrt{\frac{2}{\pi}} \tag{9.2.57}$$

and independent of $\tau \in (0, \infty)$ the minimum bias is

$$\omega_{c,t} = \sqrt{\frac{\pi}{2}} \qquad t=0,\, t=\alpha=1 \tag{9.2.58}$$

Hence, the MSE–inefficiency for $r = \infty$ is $4(1+\tau)^2/(2+\tau)^2$ which tends to 4 as $\tau \to \infty$. This is also the maximum MSE–inefficiency that we obtain in this example. The results for different values of τ and r are given in Table 9.1. As we see, the amount of non-adaptivity is isotone in τ and r in this example. ////

τ	r	$b_{(\mathrm{P1})}$	$b_{(\mathrm{P2})}$	M	$\mathrm{MSE}_{(\mathrm{P1})}$	$\mathrm{MSE}_{(\mathrm{P2})}$	MSE–ineff.
0.25	0.1	1.875	1.898	-0.036	0.846	0.847	1.001
	0.25	1.479	1.520	-0.120	0.983	0.989	1.007
	0.5	1.271	1.339	-0.298	1.322	1.357	1.027
	1.0	1.164	1.268	-0.610	2.399	2.601	1.084
	1.5	1.140	1.257	-0.776	4.049	4.588	1.133
	2.0	1.133	1.255	-0.860	6.307	7.347	1.165
	∞	1.128	1.253	-1.000	∞	∞	1.235
1.0	0.1	1.721	1.790	-0.077	0.540	0.542	1.003
	0.25	1.289	1.448	-0.251	0.649	0.670	1.032
	0.5	1.076	1.309	-0.502	0.897	1.014	1.130
	1.0	0.974	1.262	-0.773	1.656	2.233	1.349
	1.5	0.951	1.256	-0.880	2.806	4.212	1.501
	2.0	0.945	1.254	-0.927	4.376	6.965	1.592
	∞	0.940	1.253	-1.000	∞	∞	1.778
5.0	0.1	1.559	1.612	-0.171	0.200	0.201	1.006
	0.25	1.183	1.35	-0.459	0.291	0.308	1.060
	0.5	0.907	1.273	-0.732	0.484	0.623	1.288
	1.0	0.769	1.255	-0.908	0.976	1.813	1.859
	1.5	0.743	1.254	-0.956	1.683	3.780	2.246
	2.0	0.736	1.253	-0.974	2.638	6.530	2.475
	∞	0.731	1.253	-1.000	∞	∞	2.939
10.0	0.1	1.472	1.516	-0.249	0.120	0.121	1.007
	0.25	1.160	1.309	-0.586	0.205	0.219	1.071
	0.5	0.891	1.262	-0.827	0.394	0.524	1.330
	1.0	0.724	1.254	-0.947	0.842	1.707	2.028
	1.5	0.696	1.253	-0.975	1.465	3.672	2.506
	2.0	0.689	1.253	-0.986	2.303	6.421	2.789
	∞	0.684	1.253	-1.000	∞	∞	3.361
∞	0.1	1.234	1.253	-1.000	0.015	0.016	1.016
	0.25	1.141	1.253	-1.000	0.089	0.098	1.098
	0.5	0.900	1.253	-1.000	0.282	0.393	1.393
	1.0	0.673	1.253	-1.000	0.692	1.571	2.271
	1.5	0.641	1.253	-1.000	1.223	3.534	2.890
	2.0	0.632	1.253	-1.000	1.930	6.283	3.256
	∞	0.627	1.253	-1.000	∞	∞	4.000

Table 9.1: Robust non-adaptivity in terms of MSE–inefficiency for two points regressor in case of contamination neighborhoods ($t = 0$, $t = \alpha = 1$); confer Example 9.2.25.

Example 9.2.26 The second example is

$$K(dx) = \frac{1}{2+\tau}\,\mathrm{I}_{\{-1\}}(dx) + \frac{\tau}{1+\tau}\,\mathrm{I}_{\{0\}}(dx) + \frac{1}{(1+\tau)(2+\tau)}\,\mathrm{I}_{\{1+\tau\}}(dx) \tag{9.2.59}$$

where $\tau \in (0,\infty)$ and

$$\mathrm{E}\,X = 0 \qquad \mathcal{K} = \mathrm{E}\,X^2 = 1 \qquad \gamma_1 = \mathrm{E}\,X^3/\mathcal{K}^{1.5} = \tau \tag{9.2.60}$$

$$\mathcal{I}^{-1} = \mathcal{K}^{-1} = 1 \tag{9.2.61}$$

With this choice of K we can now investigate the dependence between the amount of robust non-adaptivity and the skewness γ_1 of K. As the results in Table 9.2 show, the MSE–inefficiency of the optimal solution of problem (P2) with respect to the optimal solution of problem (P1) is not monotone in τ and it is also not monotone in the radius r. Since

$$\frac{1}{2+\tau} + \frac{\tau}{1+\tau} = \frac{\tau^2+3\tau+1}{\tau^2+3\tau+2} > \frac{1}{2} \qquad \forall\,\tau \in (0,\infty) \tag{9.2.62}$$

and $1/(2+\tau) < 0.5$ on $(0,\infty)$, we obtain $\mathrm{med}(x) = 0$. Consequently, the minimum bias in both cases is

$$\omega_{c,t}^{\min} = \frac{2+\tau}{2}\sqrt{\frac{\pi}{2}} \qquad t = 0,\, t = \alpha = 1 \tag{9.2.63}$$

and hence, the MSE–inefficiency is equal to 1 in the limiting case $r \to \infty$. Moreover, the numerical results for very large values of τ indicate

$$\lim_{\tau\to\infty} \mathrm{relMSE} \to 1 \qquad \forall\, r \in [0,\infty] \tag{9.2.64}$$

This is plausible since $K \xrightarrow{\mathrm{w}} \mathrm{I}_{\{0\}}$ as $\tau \to \infty$. ////

τ	r	$b_{(\mathrm{P1})}$	$b_{(\mathrm{P2})}$	M	$\mathrm{MSE}_{(\mathrm{P1})}$	$\mathrm{MSE}_{(\mathrm{P2})}$	MSE–ineff.
0.25	0.1	2.276	2.299	-0.031	1.068	1.069	1.001
	0.25	1.799	1.827	-0.088	1.270	1.276	1.005
	0.5	1.559	1.580	-0.150	1.776	1.798	1.012
	1.0	1.444	1.448	-0.130	3.419	3.462	1.013
	1.5	1.421	1.421	-0.081	5.972	6.020	1.008
	2.0	1.414	1.414	-0.052	9.485	9.535	1.005
	∞	1.410	1.410	0.000	∞	∞	1.000
1.0	0.1	3.087	3.178	-0.059	1.133	1.138	1.004
	0.25	2.336	2.441	-0.145	1.483	1.518	1.024
	0.5	2.025	2.059	-0.165	2.333	2.420	1.037
	1.0	1.907	1.909	-0.092	5.159	5.276	1.023
	1.5	1.887	1.888	-0.051	9.643	9.766	1.013
	2.0	1.883	1.883	-0.031	15.858	15.982	1.008
	∞	1.880	1.880	0.000	∞	∞	1.000
5.0	0.1	7.143	7.290	-0.071	1.723	1.738	1.009
	0.25	5.091	5.210	-0.105	3.475	3.599	1.036
	0.5	4.517	4.527	-0.062	7.584	7.776	1.025
	1.0	4.401	4.402	-0.022	22.302	22.508	1.009
	1.5	4.390	4.390	-0.011	46.431	46.639	1.004
	2.0	4.388	4.388	-0.006	80.135	80.343	1.003
	∞	4.387	4.387	0.000	∞	∞	1.000
10.0	0.1	11.671	11.856	-0.060	2.893	2.923	1.010
	0.25	8.240	8.298	-0.058	7.405	7.582	1.024
	0.5	7.623	7.626	-0.026	18.772	18.992	1.012
	1.0	7.529	7.529	-0.008	61.535	61.762	1.004
	1.5	7.522	7.522	-0.004	132.303	132.531	1.002
	2.0	7.521	7.521	-0.002	231.294	231.522	1.001
	∞	7.520	7.520	0.000	∞	∞	1.000
100.0	0.1	68.456	68.490	-0.006	71.914	72.127	1.003
	0.25	64.173	64.173	-0.002	293.485	293.731	1.001
	0.5	63.938	63.938	0.000	1061.220	1061.468	1.000
	1.0	63.920	63.920	0.000	4125.928	4126.176	1.000
	1.5	63.919	63.919	0.000	9233.071	9233.319	1.000
	2.0	63.919	63.919	0.000	16382.970	16383.218	1.000
	∞	63.919	63.919	0.000	∞	∞	1.000

Table 9.2: Robust non-adaptivity in terms of MSE–inefficiency for three points regressor in case of contamination neighborhoods ($t = 0$, $t = \alpha = 1$); confer Example 9.2.26.

Example 9.2.27 The last example we consider is a piecewise uniform regressor; i.e.,

$$K(dx) = p\,\mathrm{I}\big([-1,0)\big)(dx) + \frac{(1-p)^2}{p}\,\mathrm{I}\big([0,\tfrac{p}{1-p}]\big)(dx) \tag{9.2.65}$$

where $p \in (0.5,1)$ and

$$\mathrm{E}\,X = 0 \qquad \mathcal{K} = \mathrm{E}\,X^2 = \frac{p}{3(1-p)} \tag{9.2.66}$$

$$\gamma_1 = \frac{\mathrm{E}\,X^3}{\mathcal{K}^{1.5}} = \frac{3\sqrt{3}\,(2p-1)}{4\sqrt{p(1-p)}} \tag{9.2.67}$$

and

$$\mathcal{I}^{-1} = \mathcal{K}^{-1} = 3\frac{1-p}{p} \tag{9.2.68}$$

With this choice of K the amount of robust non-adaptivity is increasing in $p \in (0.5,1)$ and $r \in (0,\infty)$; confer Table 9.3. The minimum bias in case of problem (P1) is

$$\omega_{c,t}^{\min} = p^{-1}\sqrt{\frac{\pi}{2}} \qquad t=0,\, t=\alpha=1 \tag{9.2.69}$$

Since $\mathrm{med}(x) = 1/(2p) - 1$, we obtain in case of problem (P2)

$$\omega_{c,t}^{\min} = \frac{4p}{4p-1}\sqrt{\frac{\pi}{2}} \qquad t=0,\, t=\alpha=1 \tag{9.2.70}$$

Hence, the MSE–inefficiency in the limiting case $r \to \infty$ is $4p^4/(4p-1)^2$ which is strictly increasing on $(0.5,1)$.

p	r	$b_{(\mathrm{P1})}$	$b_{(\mathrm{P2})}$	M	$\mathrm{MSE}_{(\mathrm{P1})}$	$\mathrm{MSE}_{(\mathrm{P2})}$	MSE–ineff.
0.55	0.1	4.084	4.099	-0.014	2.684	2.686	1.000
	0.25	3.114	3.133	-0.033	3.306	3.314	1.002
	0.5	2.631	2.651	-0.056	4.779	4.807	1.006
	1.0	2.380	2.401	-0.078	9.321	9.425	1.011
	1.5	2.317	2.338	-0.086	16.169	16.394	1.014
	2.0	2.296	2.316	-0.089	25.464	25.853	1.015
	∞	2.279	2.298	-0.091	∞	∞	1.017
2/3	0.1	3.599	3.691	-0.044	1.689	1.695	1.003
	0.25	2.635	2.780	-0.115	2.148	2.194	1.021
	0.5	2.190	2.348	-0.188	3.180	3.366	1.059
	1.0	1.968	2.113	-0.237	6.299	6.965	1.106
	1.5	1.914	2.048	-0.246	10.974	12.334	1.124
	2.0	1.895	2.025	-0.248	17.309	19.577	1.131
	∞	1.880	2.005	-0.250	∞	∞	1.138
0.75	0.1	3.430	3.534	-0.062	1.174	1.179	1.004
	0.25	2.433	2.668	-0.167	1.577	1.638	1.038
	0.5	1.973	2.250	-0.270	2.430	2.717	1.118
	1.0	1.755	1.997	-0.322	4.927	5.970	1.212
	1.5	1.703	1.926	-0.330	8.636	10.737	1.243
	2.0	1.686	1.902	-0.332	13.651	17.131	1.255
	∞	1.671	1.880	-0.333	∞	∞	1.266
0.9	0.1	3.065	3.145	-0.124	0.470	0.473	1.007
	0.25	2.259	2.525	-0.281	0.810	0.862	1.065
	0.5	1.727	2.191	-0.398	1.503	1.874	1.247
	1.0	1.479	1.876	-0.439	3.318	4.835	1.457
	1.5	1.425	1.790	-0.443	5.928	8.983	1.515
	2.0	1.408	1.760	-0.444	9.429	14.478	1.535
	∞	1.393	1.735	-0.444	∞	∞	1.553
0.999	0.1	2.457	2.497	-0.471	0.067	0.068	1.015
	0.25	2.215	2.426	-0.495	0.353	0.386	1.095
	0.5	1.679	2.216	-0.499	1.040	1.394	1.340
	1.0	1.349	1.834	-0.499	2.607	4.315	1.655
	1.5	1.289	1.732	-0.499	4.756	8.229	1.730
	2.0	1.270	1.699	-0.499	7.614	13.359	1.755
	∞	1.255	1.672	-0.499	∞	∞	1.775
$p \to 1$	∞	1.253	1.671	-0.500	∞	∞	1.778

Table 9.3: Robust non-adaptivity in terms of MSE–inefficiency for piecewise uniform regressor in case of contamination neighborhoods ($t = 0$, $t = \alpha = 1$); confer Example 9.2.27.

9.2.3 Implementation Using R

The optimal AL estimators can be computed via our R package ROptRegTS which is in detail described in Appendix D.4. In case of normal regression (with unknown scale/intercept), we provide the generating functions NormLinRegFamily, NormLinRegScaleFamily and NormLinRegInterceptFamily which can be used to generate objects of class L2RegTypeFamily. The implementation of normal linear regression, respectively normal linear regression with unknown intercept is analogously to normal linear regression with unknown scale which is presented in Subsection 7.6.1.
As an extension, we provide methods for the computation of optimally robust ICs in case of average square conditional contamination neighborhoods ($t = \alpha = 2$) and average conditional total variation neighborhoods ($t = \alpha = 1$). For this purpose we implemented the S4 classes (and corresponding generating functions) Av2CondContNeighborhood, respectively Av1CondTotalVarNeighborhood. That is, given some regressor distribution K and some neighborhood radius $r \in [0, \infty]$ we can instantiate the following infinitesimal robust models

```
> RobLM1 <-  InfRobRegTypeModel(center = LM,
+                   neighbor=Av2CondContNeighborhood(radius = r))
```

respectively

```
> RobLM2 <-  InfRobRegTypeModel(center = LM,
+               neighbor=Av1CondTotalVarNeighborhood(radius = r))
```

where LM is an object of class L2RegTypeFamily. Afterwards, we can call optIC, radiusMinimaxIC or leastFavorableRadius as described in Subsection 7.6.1.

Remark 9.2.28 **(a)** So far, only models without further nuisance parameters like scale or intercept are implemented for average square conditional contamination neighborhoods ($t = \alpha = 2$) as well as average conditional total variation neighborhoods ($t = \alpha = 1$).

(b) The implementation of normal location and scale is in detail presented in Section 8.9.

(c) After the installation of our R bundle RobASt one can find the R scripts NormLinReg.R, NormLinRegIntercept.R and NormLinRegScale.R, which contain some examples for normal regression (with unknown scale/intercept), in the directory "... /RHome/library/ROptRegTS/scripts/" where RHome stands for the local home directory of R. In addition, we provide the R script NormLinRegAdaption which demonstrates how the results given in Subsubsection 9.2.2.5 may be recomputed. ////

9.3 Time Series Models

9.3.1 ARMA(p, q) with Shift

We consider the ARMA(p, q) model

$$\phi(B)(X_t - \mu) = \xi(B)V_t \qquad t \in \mathbb{Z} \tag{9.3.1}$$

with shift $\mu = \mu_X$ where the innovations $V_t \overset{\text{i.i.d.}}{\sim} F$, the autoregressive (AR) and moving average (MA) polynomials are

$$\phi(z) = 1 + \phi_1 z + \ldots + \phi_p z^p \tag{9.3.2}$$

$$\xi(z) = 1 + \xi_1 z + \ldots + \xi_q z^q \tag{9.3.3}$$

and the backshift operator B is defined by

$$B^j X_t = X_{t-j} \qquad j \in \mathbb{Z} \tag{9.3.4}$$

We make the following assumptions

(F2) The Fisher information of location for F is finite; i.e., F has an absolutely continuous density f and $\mathcal{I}_F^{\text{loc}} = \int (\Lambda_f^{\text{loc}})^2 f \, d\lambda < \infty$ where $\Lambda_f^{\text{loc}} := -f'/f$; confer Proposition 5.1.4.

(F4) $\mu_F = \int v dF(v) = 0$

(F5) $\sigma_F^2 = \int v^2 dF(v) < \infty$ is known.

(H1) ϕ and ξ have no common zeros.

(H2) $\phi(z)\xi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.

Hence, the parameter of interest is $\theta = (\phi, \xi, \mu) = (\phi_1, \ldots, \phi_p, \xi_1, \ldots \xi_q, \mu) \in \Theta$ where

$$\Theta = \big\{\theta \in \mathbb{R}^{p+q+1} \,\big|\, \phi_j \neq 0 \; (j = 1, \ldots, p),\, \xi_k \neq 0 \; (k = 1, \ldots, q) \text{ with (H1),(H2)} \\ \text{and } \mu \in \mathbb{R} \big\} \tag{9.3.5}$$

which is an open subset of $\mathbb{R}^{p+k+1}$.

Remark 9.3.1 **(a)** By (H1) and (H2) the ARMA(p, q) process with shift μ is causal and invertible; confer for instance Theorems 3.1.1 and 3.1.2 of Brockwell and Davis (1991).

(b) As a consequence of $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$, $X_t = \mu + \psi(B)V_t$ where $\psi(z) = \xi(z)/\phi(z)$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. Hence, $(X_t)_{t\in\mathbb{Z}}$ is (strictly) stationary and ergodic since $V_t \overset{\text{i.i.d.}}{\sim} F$.

(c) For simplicity we assume (F4) and (F5). Sometimes the scale of the innovations σ_F is treated as an additional nuisance parameter. This would be possible in our context, too and would lead to a further side condition in the considered optimization problems. ////

Proposition 9.3.2 *Assuming (F2), (F4), (F5), (H1) and (H2), the* ARMA(p, q) *model (9.3.1) is* L_2 *differentiable at every* $\theta \in \Theta$ *with* L_2 *derivative* Λ_θ *and Fisher information* $\mathcal{I}_\theta$ *given by*

$$\Lambda_{\theta,t} = \Lambda_f^{\mathrm{loc}} \tilde{H}_{\theta,t} \qquad \mathcal{I}_\theta = \mathcal{I}_f^{\mathrm{loc}} \begin{pmatrix} \mathcal{K}_{\theta,f} & 0 \\ 0 & \nu^2 \end{pmatrix} \tag{9.3.6}$$

where $\tilde{H}_{\theta,t}^\tau = (H_{\theta,t}^\tau, \nu)$ *with*

$$H_{\theta,t}^\tau = \big(- B\phi^{-1}(B), \ldots, -B^p\phi^{-1}(B), B\xi^{-1}(B), \ldots, B^q\xi^{-1}(B)\big)V_t \tag{9.3.7}$$

$\nu = \phi(1)/\xi(1)$ *and* $\mathcal{K}_{\theta,f} = \mathrm{Cov}_\theta\, H_{\theta,1}$.

Proof Similar to 1.3 Proposition 4 of Staab (1984) who considers double infinite processes. ////

Remark 9.3.3 **(a)** For the verification of the LAN property of the ARMA(p, q) model we also refer to Kreiss (1987) and Drost et al. (1997).

(b) A quick derivation, based on the backshift calculus, of $\Lambda_{\theta,t}$ and $\mathcal{I}_\theta$ is given in Section 1 of Rieder (2003).

(c) $H_{\theta,t}$ depends on $V_{t-1}, V_{t-2}, \ldots$, hence is stochastically independent of V_t. Moreover, it is (strictly) stationary and ergodic.

(d) It holds, $\mathcal{I}_\theta \succ 0$ iff (H1); confer 1.4 Corollary 2 of Staab (1984).

(e) The ARMA(p, q) model is classically adaptive with respect to shift due to the block diagonal form of $\mathcal{I}_\theta$ which is caused by (F4); i.e.,

$$\mathcal{K}_{\theta,f}^{-1} = \mathcal{I}_{11}^{-1} = \mathcal{I}_{11.2}^{-1} = \mathcal{K}_{\theta,f}^{-1} \tag{9.3.8}$$

Now, we change the point of view and consider μ as main and ϕ, ξ as nuisance parameters. This model can be regarded as a generalization of the classical i.i.d. location model where the error additionally has an ARMA structure. The block diagonal form of $\mathcal{I}_\theta$ also implies classical adaptivity of the ARMA(p, q) model with respect to the ARMA parameters ϕ, ξ.

(f) The ARMA(p, q) model with shift has a similar structure as the linear model with intercept considered in Subsection 9.2.2. Hence, not surprisingly, the results about robust adaptivity are similar, too. In fact, the idea of studying robust adaptivity of the linear model with intercept arose during the work on the ARMA(p, q) model with shift. ////

9.3.1.1 Average Square Conditional Contamination Neighborhoods

We consider conditional (error-free-variables) contamination neighborhoods. These are average square transition neighborhoods $(* = c,\ t = \alpha = 2)$ with starting radius $r \in (0, \infty)$; confer Remark A.1.6. The neighborhoods are about the ideal ARMA model with parameters $\phi_1, \ldots, \phi_p, \xi_1, \ldots, \xi_q$ and shift μ.

ARMA(p, q) with main parameter ϕ, ξ and nuisance parameter μ

First, we assume shift μ to be known. Then, the solution to problem (P1) is provided by Theorem A.2.5 and reads

$$\tilde{\eta}(H_{\theta,1}, V_1) = A\mathcal{K}_{\theta,f}^{-1} H_{\theta,1}\big(\Lambda_f^{\mathrm{loc}}(V_1) - z\big)w(V_1) \tag{9.3.9}$$

where

$$w(V_1) = \min\Big\{1, \frac{c}{|\Lambda_f^{\mathrm{loc}}(V_1) - z|}\Big\} \tag{9.3.10}$$

with

$$z = \mathrm{E}\,\Lambda_f^{\mathrm{loc}} w[\mathrm{E}\,w]^{-1} \qquad A^{-1} = \mathrm{E}(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.11}$$

and the clipping bound $c \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 c = \mathrm{E}(|\Lambda_f^{\mathrm{loc}} - z| - c)_+ \tag{9.3.12}$$

The question is whether robust adaptivity holds in case shift μ is unknown.

Proposition 9.3.4 *The* ARMA(p, q) *model with average square conditional contamination neighborhoods ($* = c$, $t = \alpha = 2$) is robust–adaptive with respect to shift μ.*

Proof We obtain

$$\mathrm{E}\,\tilde{\eta}\nu\Lambda_f^{\mathrm{loc}} = A\nu\mathcal{K}_{\theta,f}^{-1}\,\mathrm{E}\,H_{\theta,1}\,\mathrm{E}_\bullet\,\big(\Lambda_f^{\mathrm{loc}} - z\big)^2 w \overset{(9.3.11)}{=} \nu\mathcal{K}_{\theta,f}^{-1}\,\mathrm{E}\,H_{\theta,1} = 0 \tag{9.3.13}$$

since $H_{\theta,1}$ and V_1 are stochastically independent and $\mathrm{E}\,H_{\theta,1} = 0$ by (F4). ////

Remark 9.3.5 By the previous proposition the solution to problem (P2) coincides with the solution to problem (P1). ////

We now change the point of view.

ARMA(p, q) with main parameter μ and nuisance parameters ϕ, ξ

First, we assume ϕ, ξ to be known. Then, the solution to problem (P1) in case $* = c$, $t = \alpha = 2$ is provided by Theorem A.2.5 and reads

$$\tilde{\eta}(H_{\theta,1}, V_1) = A\nu^{-1}\big(\Lambda_f^{\mathrm{loc}}(V_1) - z\big)w(V_1) \tag{9.3.14}$$

where

$$w(V_1) = \min\Big\{1, \frac{c}{|\Lambda_f^{\mathrm{loc}}(V_1) - z|}\Big\} \tag{9.3.15}$$

with

$$z = \mathrm{E}\,\Lambda_f^{\mathrm{loc}} w[\mathrm{E}\,w]^{-1} \qquad A^{-1} = \mathrm{E}(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.16}$$

and the clipping bound $c \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 c = \mathrm{E}(|\Lambda_f^{\mathrm{loc}} - z| - c)_+ \tag{9.3.17}$$

Remark 9.3.6 The Lagrange multipliers A, z and b included in the solution to problem (P1) are identical to those of the corresponding i.i.d. location model with unconditional contamination neighborhoods. Moreover, the minimax asymptotic MSE is just ν^{-2} times the minimax asymptotic MSE in case of i.i.d. location. Generally speaking, we are in a situation which could be interpreted in the light of Section 2.4. ////

Proposition 9.3.7 *The* ARMA(p, q) *model with average square conditional contamination neighborhoods* ($* = c$, $t = \alpha = 2$) *is robust–adaptive with respect to the ARMA parameters* ϕ, ξ.

Proof It holds

$$\mathrm{E}\,\tilde{\eta} H_{\theta,1}^{\tau} \Lambda_f^{\mathrm{loc}} = \mathrm{E}\, H_{\theta,1}^{\tau}\, \mathrm{E}\,\tilde{\eta}\Lambda_f^{\mathrm{loc}} \overset{(9.3.16)}{=} \nu^{-1}\, \mathrm{E}\, H_{\theta,1}^{\tau} = 0 \tag{9.3.18}$$

by condition (F4) which implies $\mathrm{E}\, H_{\theta,1} = 0$. ////

Remark 9.3.8 By the previous proposition the solution to problem (P2) coincides with the solution to problem (P1). ////

The next questions is whether robust adaptivity also holds in the larger neighborhood model $* = c$, $t = \alpha = 1$.

9.3.1.2 Average Conditional Contamination Neighborhoods

We consider conditional (error-free-variables) contamination neighborhoods. These are average ($* = c$, $t = \alpha = 1$) transition neighborhoods with starting radius $r \in (0, \infty)$; confer Remark A.1.6. The neighborhoods are about the ideal ARMA model with parameters $\phi_1, \ldots, \phi_p, \xi_1, \ldots, \xi_q$ and shift μ.

ARMA(p, q) with main parameter ϕ, ξ and nuisance parameter μ

First, we assume shift μ to be known; i.e., we have to solve problem (P1). The solution is given in Theorem A.2.3 and can be rewritten as

$$\tilde{\eta}(H_{\theta,1}, V) = A H_{\theta,1}\big(\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1})\big) w(H_{\theta,1}, V_1) \tag{9.3.19}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{b}{|A H_{\theta,1}(\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1}))|}\Big\} \tag{9.3.20}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{loc}} w \big[\mathrm{E}_\bullet\, w\big]^{-1} \qquad A^{-1} = \mathrm{E}\, H_{\theta,1} H_{\theta,1}^{\tau}\, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.21}$$

The clipping bound $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 b = \mathrm{E}(|A H_{\theta,1}(\Lambda_f^{\mathrm{loc}} - z)| - b)_+ \tag{9.3.22}$$

Indeed, under the price of symmetry, robust adaptivity extends.

Proposition 9.3.9 *Assume the innovation distribution F is symmetric. Then, the* ARMA(p, q) *model with average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) is robust–adaptive with respect to shift μ.*

Proof It holds

$$\mathrm{E}\,\tilde{\eta}\nu\Lambda_f^{\mathrm{loc}} = \nu\,\mathrm{E}\,H_{\theta,1}\,\mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.23}$$

where $z(-H_{\theta,1}) = z(H_{\theta,1})$. By symmetry of F, $\mathcal{L}(H_{\theta,1})$ is symmetric, too. As a consequence, $H_{\theta,1}\,\mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w$ is an odd function in $H_{\theta,1}$, hence, $\mathrm{E}\,\tilde{\eta}\nu\Lambda_f^{\mathrm{loc}} = 0$. ////

If we do not insist on symmetry of the innovation distribution F, we may have $\mathrm{E}\,\tilde{\eta}\nu\Lambda_f^{\mathrm{loc}} \neq 0$. In this general case the solution to problem (P2) differs from the solution to problem (P1). The solution to problem (P2) can again be obtained by Theorem A.2.3 and reads

$$\tilde{\eta}(H_{\theta,1}, V) = \big[(A_H H_{\theta,1} + \nu A_\nu)\Lambda_f^{\mathrm{loc}}(V_1) - a(H_{\theta,1})\big]\, w(H_{\theta,1}, V_1) \tag{9.3.24}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{b}{|(A_H H_{\theta,1} + \nu A_\nu)\Lambda_f^{\mathrm{loc}}(V_1) - a(H_{\theta,1})|}\Big\} \tag{9.3.25}$$

with

$$a(H_{\theta,1}) = (A_H H_{\theta,1} + \nu A_\nu)\,\mathrm{E}_\bullet\,\Lambda_f^{\mathrm{loc}} w\big[\mathrm{E}_\bullet\, w\big]^{-1} \tag{9.3.26}$$

$$A_H = \big[\mathbb{I}_{p+q} - \nu A_\nu\,\mathrm{E}\,H_{\theta,1}^\tau(\Lambda_f^{\mathrm{loc}})^2 w + \mathrm{E}\,a H_{\theta,1}^\tau \Lambda_f^{\mathrm{loc}} w\big]\big[\mathrm{E}\,H_{\theta,1} H_{\theta,1}^\tau(\Lambda_f^{\mathrm{loc}})^2 w\big]^{-1} \tag{9.3.27}$$

$$A_\nu = \big[\mathrm{E}\,a\Lambda_f^{\mathrm{loc}} w - A_H\,\mathrm{E}\,H(\Lambda_f^{\mathrm{loc}})^2 w\big]\big[\nu\,\mathrm{E}(\Lambda_f^{\mathrm{loc}})^2 w\big]^{-1} \tag{9.3.28}$$

and the clipping bound $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 b = \mathrm{E}(|(A_H H_{\theta,1} + \nu A_\nu)\Lambda_f^{\mathrm{loc}} - a)| - b)_+ \tag{9.3.29}$$

It is formally visible, the solution to problem (P1) given in (9.3.19) is not equal to the solution to problem (P2) given in (9.3.24).

Remark 9.3.10 We numerically determine the amount of non-adaptivity in case of AR(1), respectively MA(1) with shift and Gumbel distributed innovations in Example 9.3.25. As it turns out, the amount of non-adaptivity is very small ($\leq$ 3%); confer Table 9.6. ////

We now change the point of view.

ARMA(p, q) with main parameter μ and nuisance parameters ϕ, ξ

We assume μ is the parameter of interest and the ARMA parameters are additional nuisance parameters.

Remark 9.3.11 The solution to problem (P1) in case of average conditional neighborhoods ($* = c$, $t = \alpha = 1$) is provided by Theorem A.2.3. It can be rewritten such that it attains the form (9.3.14); i.e., the solution is identical to the solution

in case of average square conditional neighborhoods ($* = c,\ t = \alpha = 2$) which is specified in equation (9.3.14). As a consequence, the ARMA(p, q) with average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$) is robust–adaptive with respect to the ARMA parameters ϕ, ξ; confer Proposition 9.3.7. ////

9.3.1.3 Average Conditional Total Variation Neighborhoods

We consider average transition neighborhoods of total variation type (i.e., $* = v$, $t = \alpha = 1$) with starting radius $r \in (0, \infty)$. The neighborhoods are about the ideal ARMA model with parameters $\phi_1, \ldots, \phi_p, \xi_1, \ldots, \xi_q$ and shift μ.

ARMA(p, q) with main parameter ϕ, ξ and nuisance parameter μ

First, we assume shift μ to be known. Then, the solution to problem (P1) in case $* = v,\ t = \alpha = 1$ can be read off from Theorem A.2.7

$$\tilde{\eta}(H_{\theta,1}, V_1) = AH_{\theta,1}\Lambda_f^{\mathrm{loc}}(V_1) w(H_{\theta,1}, V_1) \tag{9.3.30}$$

$$w(H_{\theta,1}, V_1) = 1 \wedge \max\left\{ \frac{c(H_{\theta,1})}{|AH_{\theta,1}|\Lambda_f^{\mathrm{loc}}(V_1)},\ \frac{c(H_{\theta,1}) + b}{|AH_{\theta,1}|\Lambda_f^{\mathrm{loc}}(V_1)} \right\} \tag{9.3.31}$$

where

$$\mathrm{E}_\bullet \left(c - |AH_{\theta,1}|\Lambda_f^{\mathrm{loc}}\right)_+ = \mathrm{E}_\bullet \left(|AH_{\theta,1}|\Lambda_f^{\mathrm{loc}} - (c+b)\right)_+ \tag{9.3.32}$$

$$A^{-1} = \mathrm{E}\, H_{\theta,1} H_{\theta,1}^\tau\, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w \tag{9.3.33}$$

and $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 b = \mathrm{E}_\bullet \left(|AH_{\theta,1}|\Lambda_f^{\mathrm{loc}} - (c+b)\right)_+ \tag{9.3.34}$$

The question is whether robust adaptivity holds in case shift μ is unknown. Analogously to the cases $* = c,\ t = \alpha = 1$ (cf. Proposition 9.3.9) we have robust adaptivity under the price of symmetry.

Proposition 9.3.12 *Assume the innovation distribution F is symmetric. Then, the* ARMA(p, q) *model with average conditional total variation neighborhoods ($* = v,\ t = \alpha = 1$) is robust–adaptive with respect to shift μ.*

Proof We have

$$\mathrm{E}\, \tilde{\eta} \nu \Lambda_f^{\mathrm{loc}} = \nu\, \mathrm{E}\, H_{\theta,1}\, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w \tag{9.3.35}$$

where $c(-H_{\theta,1}) = c(H_{\theta,1})$. Since the symmetry of F implies the symmetry of $\mathcal{L}(H_{\theta,1})$, $H_{\theta,1}\, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w$ is an odd function in $H_{\theta,1}$; i.e., $\mathrm{E}\, \tilde{\eta} \nu \Lambda_f^{\mathrm{loc}} = 0$. ////

However, if we do not insist on symmetry of the innovation distribution F, we may have $\mathrm{E}\, \tilde{\eta} \nu \Lambda_f^{\mathrm{loc}} \neq 0$. In this general case the solution to problem (P2) differs

from the solution to problem (P1). The solution to problem (P2) can again be obtained by Theorem A.2.7 and reads

$$\tilde{\eta}(H_{\theta,1}, V_1) = (A_H H_{\theta,1} + A_\nu \nu)\Lambda_f^{\mathrm{loc}}(V_1) w(H_{\theta,1}, V_1) \tag{9.3.36}$$

$$w(H_{\theta,1}, V_1) = 1 \wedge \max\left\{ \frac{c(H_{\theta,1})}{|A_H H_{\theta,1} + A_\nu \nu| \Lambda_f^{\mathrm{loc}}(V_1)}, \frac{c(H_{\theta,1}) + b}{|A_H H_{\theta,1} + A_\nu \nu| \Lambda_f^{\mathrm{loc}}(V_1)} \right\} \tag{9.3.37}$$

where

$$\mathrm{E}_\bullet \left(c - |A_H H_{\theta,1} + A_\nu \nu| \Lambda_f^{\mathrm{loc}} \right)_+ = \mathrm{E}_\bullet \left(|A_H H_{\theta,1} + A_\nu \nu| \Lambda_f^{\mathrm{loc}} - (c+b) \right)_+ \tag{9.3.38}$$

$$A_H = \left[\mathbb{I}_{p+q} - A_\nu \nu \, \mathrm{E}\, H_{\theta,1}^\tau \, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w \right] \left[\mathrm{E}\, H_{\theta,1} H_{\theta,1}^\tau \, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w \right]^{-1} \tag{9.3.39}$$

$$A_\nu = \left[- A_H \, \mathrm{E}\, H_{\theta,1} \, \mathrm{E}_\bullet (\Lambda_f^{\mathrm{loc}})^2 w \right] \left[\nu \, \mathrm{E} (\Lambda_f^{\mathrm{loc}})^2 w \right]^{-1} \tag{9.3.40}$$

and $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 b = \mathrm{E}_\bullet \left(|A_H H_{\theta,1} + A_\nu \nu| \Lambda_f^{\mathrm{loc}} - (c+b) \right)_+ \tag{9.3.41}$$

Obviously, the solution to problem (P1) given in (9.3.30) is not equal to the solution to problem (P2) given in (9.3.36).

ARMA(p, q) with main parameter μ and nuisance parameters ϕ, ξ

First, we assume the ARMA parameters ϕ, ξ to be known. Then, the solution to problem (P1) in case $* = v$, $t = \alpha = 1$ is provided by Theorem A.2.7 and can be rewritten as

$$\tilde{\eta}(H_{\theta,1}, V) = A\nu^{-1} \left[g \vee \Lambda_f^{\mathrm{loc}}(V) \wedge (g+c) \right] \tag{9.3.42}$$

where

$$\mathrm{E}(g - \Lambda_f^{\mathrm{loc}})_+ = \mathrm{E} \left(\Lambda_f^{\mathrm{loc}} - (g+c) \right)_+ \tag{9.3.43}$$

$$A^{-1} = \mathrm{E} \left[g \vee \Lambda_f^{\mathrm{loc}} \wedge (g+c) \right] \Lambda_f^{\mathrm{loc}} \tag{9.3.44}$$

and the clipping bounds $c \in (0, \infty)$, $g \in (-\infty, c)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 c = \mathrm{E}(g - \Lambda_f^{\mathrm{loc}})_+ \tag{9.3.45}$$

Remark 9.3.13 The Lagrange multipliers A, g and c included in the solution to problem (P1) are identical to those of the corresponding i.i.d. location model with unconditional total variation neighborhoods. Moreover, the minimax asymptotic MSE is just ν^{-2} times the minimax asymptotic MSE in case of i.i.d. location; confer also Remark 9.3.6. Generally speaking, we are in a situation which could be interpreted in the light of Section 2.4. ////

Proposition 9.3.14 *The* ARMA(p, q) *model with average conditional total variation neighborhoods* $(* = v,\ t = \alpha = 1)$ *is robust–adaptive with respect to the ARMA parameters* ϕ, ξ.

PROOF Analogously to the proof of Proposition 9.3.7. ////

Remark 9.3.15 By the previous proposition the solution to problem (P2) coincides with the solution to problem (P1). ////

9.3.2 ARCH(p) with Scale

We consider the ARCH(p) model

$$X_t = \sigma(1 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2)^{1/2} V_t \qquad t \in \mathbb{Z} \tag{9.3.46}$$

with scale σ where the innovations $V_t \overset{\text{i.i.d.}}{\sim} F$. This model was introduced by Engle (1982). We make the following assumptions

(F3) The Fisher information of scale for F is finite; i.e., $u \mapsto uf(u) =: uf$ is absolute continuous and $\mathcal{I}_F^{\text{sc}} = \int (\Lambda_f^{\text{sc}})^2 f \, d\lambda < \infty$ where $\Lambda_f^{\text{sc}} := u\Lambda_f^{\text{loc}} - 1$; confer Proposition 5.1.1.

(F4) $\mu_F = \int v dF(v) = 0$

(F5) $\sigma_F^2 = \int v^2 dF(v) = 1$

(H3) $\mathrm{E} \log V_t^2 + \log \sigma^2 + \log \max\{\alpha_j \,|\, j = 1, \ldots, p\} < 0$

Hence, the parameter of interest is $\theta = (\alpha_1, \ldots, \alpha_p, \sigma) \in \Theta$ with

$$\Theta = \{\theta \in \mathbb{R}^{p+1} \,|\, \sigma, \alpha_j \in (0, \infty),\ j = 1, \ldots, p \text{ with (H3)}\} \tag{9.3.47}$$

an open subset of $\mathbb{R}^{p+1}$.

Remark 9.3.16 By (H3) the ARCH(p) process $(X_t)_{t \in \mathbb{Z}}$ is (strictly) stationary and ergodic; confer Nelson (1990). ////

Proposition 9.3.17 *Assuming (F3)–(F5) and (H3), the* ARCH(p) *model (9.3.46) is* L_2 *differentiable at every* $\theta \in \Theta$ *with* L_2 *derivative* Λ_θ *and Fisher information* $\mathcal{I}_\theta$ *given by*

$$\Lambda_{\theta,t} = \Lambda_f^{\text{sc}} \tilde{H}_{\theta,t} \qquad \mathcal{I}_\theta = \mathcal{I}_f^{\text{sc}} \begin{pmatrix} \mathcal{K}_{\theta,f} & \mathrm{E}\, H_{\theta,1} \\ \mathrm{E}\, H_{\theta,1}^\tau & \sigma^{-2} \end{pmatrix} \tag{9.3.48}$$

where $\tilde{H}_{\theta,t}^\tau = (H_{\theta,t}^\tau, \sigma^{-1})$ *with*

$$H_{\theta,t}^\tau = \left(\frac{X_{t-1}^2}{2(1 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2)}, \ldots, \frac{X_{t-p}^2}{2(1 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2)} \right) \tag{9.3.49}$$

and $\mathcal{K}_{\theta,f} = \mathrm{E}_\theta\, H_{\theta,1} H_{\theta,1}^\tau$.

PROOF Consequence of Drost et al. (1997) (cf. Example 4.4). ////

Remark 9.3.18 (a) By Proposition 6.5 of Breiman (1968) (strictly) stationary processes are extendable to an infinite past. In case of the GARCH$(1,1)$ model and fixed parameter θ this is verified in Theorem 2 of Nelson (1990). If such an extension is achievable for stationary ARCH(p) processes, where the special structure of these processes has to be preserved, one probably can dispense assumption (A′) of Drost et al. (1997), which takes care about the starting conditions.

(b) $H_{\theta,t}$ is stochastically independent of V_t and (strictly) stationary and ergodic.

(c) Since $\mathrm{E}\, H_{\theta,1} \neq 0$, the Fisher information is not of diagonal form. Thus, the ARCH(p) model is not classically adaptive with respect to scale. ////

9.3.2.1 Average Square Conditional Contamination Neighborhoods

We consider average square transition neighborhoods ($t = \alpha = 2$) of contamination type ($* = c$) with starting radius $r \in (0, \infty)$.

ARCH(p) with main parameters $\alpha_1, \ldots, \alpha_p$ and nuisance parameter σ

We assume the ARCH parameters $\alpha_1, \ldots, \alpha_p$ have to be estimated where the scale σ is known. The solution to problem (P1) in case $* = c$, $t = \alpha = 2$ is provided by Theorem A.2.5 and reads

$$\tilde{\eta}(H_{\theta,1}, V_1) = A\mathcal{K}_{\theta,f}^{-1} H_{\theta,1}\big(\Lambda_f^{\mathrm{sc}}(V_1) - z\big) w(V_1) \tag{9.3.50}$$

where

$$w(V_1) = \min\left\{1, \frac{c}{|\Lambda_f^{\mathrm{sc}}(V_1) - z|}\right\} \tag{9.3.51}$$

with

$$z = \mathrm{E}\, \Lambda_f^{\mathrm{sc}} w [\mathrm{E}\, w]^{-1} \qquad A^{-1} = \mathrm{E}(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.52}$$

and the clipping bound $c \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 c = \mathrm{E}(|\Lambda_f^{\mathrm{sc}} - z| - c)_+ \tag{9.3.53}$$

Proposition 9.3.19 *The* ARCH(p) *model with average square conditional contamination neighborhoods* ($* = c$, $t = \alpha = 2$) *is not robust–adaptive with respect to scale.*

PROOF We obtain

$$\mathrm{E}\, \tilde{\eta} \sigma^{-1} \Lambda_f^{\mathrm{sc}} = A\sigma^{-1} \mathcal{K}_{\theta,f}^{-1}\, \mathrm{E}\, H_{\theta,1}\, \mathrm{E}_\bullet \left(\Lambda_f^{\mathrm{sc}} - z\right)^2 w \overset{(9.3.52)}{=} \sigma^{-1} \mathcal{K}_{\theta,f}^{-1}\, \mathrm{E}\, H_{\theta,1} \neq 0 \tag{9.3.54}$$

since $\mathrm{E}\, H_{\theta,1} \neq 0$ and $H_{\theta,1}$ and V_1 are stochastically independent. ////

Now, let σ be unknown. Then, the solution to problem (P2) is also given by Theorem A.2.5 and reads

$$\tilde{\eta}(\tilde{H}_{\theta,1}, V_1) = A\tilde{\mathcal{K}}_{\theta,f}^{-1}\tilde{H}_{\theta,1}\big(\Lambda_f^{\mathrm{sc}}(V_1) - z\big)w(V_1) \qquad \tilde{\mathcal{K}}_{\theta,f} = \begin{pmatrix} \mathcal{K}_{\theta,f} & \mathrm{E}\, H_{\theta,1} \\ \mathrm{E}\, H_{\theta,1}^{\tau} & \sigma^{-2} \end{pmatrix} \tag{9.3.55}$$

where

$$w(V_1) = \min\Big\{1, \frac{c}{|\Lambda_f^{\mathrm{sc}}(V_1) - z|}\Big\} \tag{9.3.56}$$

with

$$z = \mathrm{E}\,\Lambda_f^{\mathrm{sc}} w [\mathrm{E}\, w]^{-1} \qquad A^{-1} = \mathrm{E}(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.57}$$

and the clipping bound $c \in (0, \infty)$ and radius $r \in (0, \infty)$ are connected via

$$r^2 c = \mathrm{E}(|\Lambda_f^{\mathrm{sc}} - z| - c)_+ \tag{9.3.58}$$

Remark 9.3.20 Since A, z, c do not depend on $H_{\theta,1}$, they are identical in both cases and can be taken from the corresponding one-dimensional i.i.d. scale model. As the factor $\mathrm{tr}\,\mathcal{K}_{\theta,f}^{-1}$, respectively $\mathrm{tr}\,\tilde{\mathcal{K}}_{\theta,f}^{-1}$ cancels out, the least favorable radii and the MSE–inefficiencies coincide with the values of one-dimensional i.i.d. scale. Moreover, the amount of non-adaptivity is the same for all $r \in [0, \infty]$, namely $\mathrm{tr}\,\tilde{\mathcal{K}}_{\theta,f}^{-1} / \mathrm{tr}\,\mathcal{K}_{\theta,f}^{-1} > 1$. ////

9.3.2.2 Average Conditional Contamination Neighborhoods

We consider average transition neighborhoods ($t = \alpha = 1$) of contamination type ($* = c$) with starting radius $r \in (0, \infty)$.

ARCH(p) with main parameters $\alpha_1, \ldots, \alpha_p$ and nuisance parameter σ

We assume the ARCH parameters $\alpha_1, \ldots, \alpha_p$ have to be estimated where the scale σ is known. The solution to problem (P1) in case $* = c$, $t = \alpha = 2$ is provided by Theorem A.2.3 and reads

$$\tilde{\eta}(H_{\theta,1}, V) = AH_{\theta,1}\big(\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1})\big)w(H_{\theta,1}, V_1) \tag{9.3.59}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{b}{|AH_{\theta,1}(\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1}))|}\Big\} \tag{9.3.60}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} w \big[\mathrm{E}_\bullet\, w\big]^{-1} \qquad A^{-1} = \mathrm{E}\, H_{\theta,1} H_{\theta,1}^{\tau}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.61}$$

and the clipping bound $b \in (0, \infty)$ and radius $r \in (0, \infty)$ are related via

$$r^2 b = \mathrm{E}(|AH_{\theta,1}(\Lambda_f^{\mathrm{sc}} - z)| - b)_+ \tag{9.3.62}$$

Proposition 9.3.21 *The* ARCH(p) *model with average conditional contamination neighborhoods* ($* = c$, $t = \alpha = 1$) *is not robust–adaptive with respect to scale.*

Proof It holds,

$$\mathrm{E}\,\tilde{\eta}\sigma^{-1}\Lambda_f^{\mathrm{sc}} = \sigma^{-1}\,\mathrm{E}\,H_{\theta,1}\,\mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \neq 0 \tag{9.3.63}$$

since $H_{\theta,1}\,\mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \geq 0$ and > 0 on some set of positive measure. ////

Now, assume σ is unknown. Then, the solution to problem (P2) provided by Theorem A.2.3 is

$$\tilde{\eta}(H_{\theta,1}, V) = \big[(A_H H_{\theta,1} + A_\sigma\sigma^{-1})\Lambda_f^{\mathrm{sc}}(V_1) - a(H_{\theta,1})\big]w(H_{\theta,1}, V_1) \tag{9.3.64}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{b}{|(A_H H_{\theta,1} + A_\sigma\sigma^{-1})\Lambda_f^{\mathrm{sc}}(V_1) - a(H_{\theta,1})|}\Big\} \tag{9.3.65}$$

with

$$a(H_{\theta,1}) = (A_H H_{\theta,1} + A_\sigma\sigma^{-1})\,\mathrm{E}_\bullet\,\Lambda_f^{\mathrm{sc}} w\big[\mathrm{E}_\bullet\, w\big]^{-1} \tag{9.3.66}$$

$$A_H = \big[\mathbb{I}_{p+q} - \sigma^{-1}A_\sigma\,\mathrm{E}\,H_{\theta,1}^\tau(\Lambda_f^{\mathrm{sc}})^2 w + \mathrm{E}\,aH_{\theta,1}^\tau\Lambda_f^{\mathrm{sc}} w\big]\big[\mathrm{E}\,H_{\theta,1}H_{\theta,1}^\tau(\Lambda_f^{\mathrm{sc}})^2 w\big]^{-1} \tag{9.3.67}$$

$$A_\sigma = \big[\mathrm{E}\,a\Lambda_f^{\mathrm{sc}} w - A_H\,\mathrm{E}\,H(\Lambda_f^{\mathrm{sc}})^2 w\big]\big[\sigma^{-1}\,\mathrm{E}(\Lambda_f^{\mathrm{sc}})^2 w\big]^{-1} \tag{9.3.68}$$

and the clipping bound $b \in (0,\infty)$ and radius $r \in (0,\infty)$ are related via

$$r^2 b = \mathrm{E}(|(A_H H_{\theta,1} + A_\sigma\sigma^{-1})\Lambda_f^{\mathrm{sc}} - a)| - b)_+ \tag{9.3.69}$$

Remark 9.3.22 We numerically determine the amount of non-adaptivity in case of ARCH(1) with scale and lognormal distributed innovations in Example 9.3.27. As it turns out, the amount of non-adaptivity may become very large at the boundaries of the corresponding stationarity region; confer Tables 9.7 and 9.8. ////

9.3.2.3 Average Conditional Total Variation Neighborhoods

We consider average transition neighborhoods ($t = \alpha = 1$) of total variation type ($* = v$) with starting radius $r \in (0,\infty)$.

ARCH(p) with main parameters $\alpha_1, \ldots, \alpha_p$ and nuisance parameter σ

We assume the ARCH parameters $\alpha_1, \ldots, \alpha_p$ have to be estimated where the scale σ is known. Then, the solution to problem (P1), which can be read off from Theorem A.2.7, is

$$\tilde{\eta}(H_{\theta,1}, V_1) = AH_{\theta,1}\Lambda_f^{\mathrm{sc}}(V_1)w(H_{\theta,1}, V_1) \tag{9.3.70}$$

$$w(H_{\theta,1}, V_1) = 1 \wedge \max\Big\{\frac{c(H_{\theta,1})}{|AH_{\theta,1}|\Lambda_f^{\mathrm{sc}}(V_1)}, \frac{c(H_{\theta,1}) + b}{|AH_{\theta,1}|\Lambda_f^{\mathrm{sc}}(V_1)}\Big\} \tag{9.3.71}$$

where

$$\mathrm{E}_\bullet \left(c - |AH_{\theta,1}|\Lambda_f^{\mathrm{sc}}\right)_+ = \mathrm{E}_\bullet \left(|AH_{\theta,1}|\Lambda_f^{\mathrm{sc}} - (c+b)\right)_+ \tag{9.3.72}$$

$$A^{-1} = \mathrm{E}\, H_{\theta,1} H_{\theta,1}^\tau\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}})^2 w \tag{9.3.73}$$

and $b \in (0,\infty)$ and radius $r \in (0,\infty)$ are connected via

$$r^2 b = \mathrm{E}_\bullet \left(|AH_{\theta,1}|\Lambda_f^{\mathrm{sc}} - (c+b)\right)_+ \tag{9.3.74}$$

Proposition 9.3.23 *The* ARCH(p) *model with average conditional total variation neighborhoods* ($* = v$, $t = \alpha = 1$) *is not robust–adaptive with respect to scale.*

Proof Analogously to the proof of Proposition 9.3.21. ////

Now, assume σ is unknown. Then, the solution to problem (P2) given by Theorem A.2.7 is

$$\tilde{\eta}(H_{\theta,1}, V_1) = (A_H H_{\theta,1} + A_\sigma \sigma^{-1})\Lambda_f^{\mathrm{sc}}(V_1) w(H_{\theta,1}, V_1) \tag{9.3.75}$$

$$w(H_{\theta,1}, V_1) = 1 \wedge \max\left\{ \frac{c(H_{\theta,1})}{|A_H H_{\theta,1} + A_\sigma \sigma^{-1}|\Lambda_f^{\mathrm{sc}}(V_1)}, \frac{c(H_{\theta,1}) + b}{|A_H H_{\theta,1} + A_\sigma \sigma^{-1}|\Lambda_f^{\mathrm{sc}}(V_1)} \right\} \tag{9.3.76}$$

where

$$\mathrm{E}_\bullet \left(c - |A_H H_{\theta,1} + A_\sigma \sigma^{-1}|\Lambda_f^{\mathrm{sc}}\right)_+ = \mathrm{E}_\bullet \left(|A_H H_{\theta,1} + A_\sigma \sigma^{-1}|\Lambda_f^{\mathrm{sc}} - (c+b)\right)_+ \tag{9.3.77}$$

$$A_H = \left[\mathbb{I}_{p+q} - A_\sigma \sigma^{-1}\, \mathrm{E}\, H_{\theta,1}^\tau\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}})^2 w\right] \left[\mathrm{E}\, H_{\theta,1} H_{\theta,1}^\tau\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}})^2 w\right]^{-1} \tag{9.3.78}$$

$$A_\sigma = \left[- A_H\, \mathrm{E}\, H_{\theta,1}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}})^2 w\right] \left[\sigma^{-1}\, \mathrm{E}(\Lambda_f^{\mathrm{sc}})^2 w\right]^{-1} \tag{9.3.79}$$

and $b \in (0,\infty)$ and radius $r \in (0,\infty)$ are connected via

$$r^2 b = \mathrm{E}_\bullet \left(|A_H H_{\theta,1} + A_\sigma \sigma^{-1}|\Lambda_f^{\mathrm{sc}} - (c+b)\right)_+ \tag{9.3.80}$$

9.3.3 Numerical Results

9.3.3.1 AR(1) and MA(1) with Shift

In case of AR(1), respectively MA(1) with shift we obtain

$$H_{\theta,t} = -\sum_{j=1}^{\infty} (-\phi_1)^{j-1} V_{t-j} \qquad \text{resp.} \qquad H_{\theta,t} = \sum_{j=1}^{\infty} (-\xi_1)^{j-1} V_{t-j} \tag{9.3.81}$$

where $\phi \in (-1,1)$, respectively $\xi \in (-1,1)$ to ensure $(H2)$. Hence,

$$\mathcal{K}_{\theta,1} = \frac{\sigma_F^2}{1-\phi_1^2} \qquad \text{resp.} \qquad \mathcal{K}_{\theta,1} = \frac{\sigma_F^2}{1-\xi_1^2} \tag{9.3.82}$$

and $\nu^2 = (1+\phi)^2$, respectively $\nu^2 = (1+\xi)^{-2}$.

AR(1) and MA(1) with main parameter ϕ_1, respectively ξ_1 and nuisance parameter μ

The solution to problem (P1) in case of average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$) reads

$$\tilde{\eta}(H_{\theta,1}, V) = A H_{\theta,1}\big(\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1})\big) w(H_{\theta,1}, V_1) \tag{9.3.83}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{c(H_{\theta,1})}{|\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1})|}\Big\} \tag{9.3.84}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{loc}} w \big[\mathrm{E}_\bullet\, w\big]^{-1} \qquad A^{-1} = \mathrm{E}\, H_{\theta,1}^2\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.85}$$

and $c(H_{\theta,1})$ and $r \in (0, \infty)$ are related via

$$r^2 = \mathrm{E}\, c^{-1}\, \mathrm{E}_\bullet\, \big(|\Lambda_f^{\mathrm{loc}} - z| - c\big)_+ \tag{9.3.86}$$

In case of average conditional contamination neighborhoods ($* = c,\ t = \alpha = 1$), the solution to problem (P2) can be rewritten as

$$\tilde{\eta}(H_{\theta,1}, V) = A_H (H_{\theta,1} - M)(\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1})) w(H_{\theta,1}, V_1) \tag{9.3.87}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{c(H_{\theta,1})}{|\Lambda_f^{\mathrm{loc}}(V_1) - z(H_{\theta,1})|}\Big\} \tag{9.3.88}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{loc}} w \big[\mathrm{E}_\bullet\, w\big]^{-1} \tag{9.3.89}$$

$$A_H^{-1} = \mathrm{E}(H_{\theta,1} - M) H_{\theta,1}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.90}$$

$$M = \mathrm{E}\, H\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w \big/ \mathrm{E}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{loc}} - z)^2 w \tag{9.3.91}$$

and $c(H_{\theta,1})$ and $r \in (0, \infty)$ are related via

$$r^2 = \mathrm{E}\, c^{-1}\, \mathrm{E}_\bullet(|\Lambda_f^{\mathrm{loc}} - z)| - c)_+ \tag{9.3.92}$$

Remark 9.3.24 (a) Thus, in both problems, the inner (conditional) integrals are identical to the integrals which occur in one-dimensional i.i.d. location. Moreover, in case $\phi_1 = \xi_1$, we have

$$\mathcal{L}\left(H_{\theta,1}^{\mathrm{AR}(1)}\right) = \mathcal{L}\left(-H_{\theta,1}^{\mathrm{MA}(1)}\right) \tag{9.3.93}$$

Hence, we are in an analogous situation as in Theorem 2.4.1 and obtain in case of problem (P1)

$$A^{\mathrm{AR}(1)} = A^{\mathrm{MA}(1)} \qquad z^{\mathrm{AR}(1)} = -z^{\mathrm{MA}(1)} \qquad b^{\mathrm{AR}(1)} = b^{\mathrm{MA}(1)} \tag{9.3.94}$$

Similarly, we get in case of problem (P2)

$$A^{\mathrm{AR}(1)} = A^{\mathrm{MA}(1)} \qquad M^{\mathrm{AR}(1)} = -M^{\mathrm{MA}(1)} \tag{9.3.95}$$

$$z^{\mathrm{AR}(1)} = -z^{\mathrm{MA}(1)} \qquad b^{\mathrm{AR}(1)} = b^{\mathrm{MA}(1)} \tag{9.3.96}$$

In particular, the MSE–inefficiency of problem (P2) with respect to problem (P1) is the same in case of AR(1) and MA(1), respectively. Thus, we only specify the results for AR(1).

(b) In view of the central limit theorem, convolution has a symmetrizing effect. Thus, for values of ϕ_1 close to -1, $\mathcal{L}(H_{\theta,1})$ is nearly symmetric and this effect is even stronger for ϕ_1 close to 1. Since we have robust adaptivity under symmetry, it is not surprising that the amount of robust non-adaptivity is very small in the AR(1) and hence, also in the MA(1) model.

(c) For the computation of the optimally robust ICs we use an approximation of $\mathcal{L}(H_{\theta,1})$ which we compute using FFT as described in Appendix C; i.e., we work with

$$\mathcal{L}(H^{\natural}_{\theta,1}) = \mathcal{L}\left(-\sum_{j=1}^{N}(-\phi_1)^{j-1}V_{t-j}\right) \tag{9.3.97}$$

where $N \in \mathbb{N}$ such that $|\phi_1|^{N-1} < \delta$. As the numerical results in Example 9.3.25 show, this approximations work really well for values of ϕ_1 not too close to ± 1. ////

Example 9.3.25 We consider Gumbel$(\mu, 1)$ distributed innovations. Hence, the inner (conditional) integrals correspond to the integrals of the exponential scale model; confer Subsection 5.2.1. By (F4), which is $\mu_F = \int v dF(v) = 0$, we have to choose $\mu = \gamma$ where $\gamma \approx 0.577216$ $(= -\mathrm{di}\Gamma(1))$ is the Euler-Mascheroni constant. Moreover, $\sigma_F^2 = \pi^2/6$ $(= \mathrm{tri}\Gamma(1))$ and $\int v^3 dF(v) = -2\zeta(3)$ where $\zeta(3) \approx 1.202057$ is Apéry's constant. As a first check of the numerical approximation (9.3.97), we compare the mean, variance and skewness of $H_{\theta,1}$, which are

$$\mu_H = 0 \qquad \sigma_H^2 = \frac{\pi^2}{6(1-\phi_1^2)} \qquad \gamma_{1,H} = \frac{12\zeta(3)(1-\phi_1^2)}{\pi^2(1+\phi_1^3)} \tag{9.3.98}$$

with the corresponding values of $H^{\natural}_{\theta,1}$ (obtained via numerical integration); confer Table 9.4. We additionally checked the results via Monte Carlo simulations. That is, we simulated an AR(1) process of length 110000 and threw away the first 10000 results due to stationary. We then computed the Lagrange multipliers contained in the optimally robust IC via crude Monte Carlo integration. We repeated this procedure 10 times. One example is given in Table 9.5 where we additionally provide 95% confidence intervals based on the central limit theorem. For $\phi_1 = -0.5$ mean, variance and skewness of 10 simulated $H_{\theta,1}$ are

$$-0.006 \pm 0.018 \qquad 2.192 \pm 0.027 \qquad 1.262 \pm 0.072 \tag{9.3.99}$$

respectively. As we see, the results of the numerical and the Monte-Carlo approximations are in good agreement. In any case the amount of non-adaptivity is

quite small and in view of Remark 9.3.24 (b), it is not surprising that the largest MSE–inefficiencies occur for small values of ϕ_1. ////

ϕ	μ_H	σ_H^2	$\gamma_{1,H}$	$\mu_H^\natural$	$(\sigma_H^\natural)^2$	$\gamma_{1,H}^\natural$
-0.75	0	3.760	1.106	0.048	3.561	1.139
-0.50	0	2.193	1.253	0.005	2.179	1.284
-0.10	0	1.662	1.448	0.008	1.648	1.495
0.10	0	1.662	1.445	0.007	1.648	1.487
0.50	0	2.193	0.974	0.001	2.193	0.975
0.75	0	3.760	0.450	0.001	3.759	0.452

Table 9.4: Mean, variance and skewness of $H_{\theta,1}$ and the numerical approximation $H_{\theta,1}^\natural$ in case of AR(1) with Gumbel innovations; confer Example 9.3.25. [We use $\delta = 1e-17$ in approximation (9.3.97) and $\varepsilon = 1e{-}08$ and $q = 14$ ($\phi = \mp 0.1, \mp 0.5$), respectively $q = 13$ ($\phi = \mp 0.75$) in the FFT algorithm.]

r	$b_{(\mathrm{P1})}$	$b_{(\mathrm{P2})}$	M
0.1	2.743 ± 0.014	2.758 ± 0.019	-0.059 ± 0.018
0.25	1.867 ± 0.010	1.880 ± 0.012	-0.100 ± 0.018
0.5	1.494 ± 0.008	1.507 ± 0.010	-0.135 ± 0.017
1.0	1.321 ± 0.007	1.332 ± 0.008	-0.163 ± 0.017
1.5	1.280 ± 0.007	1.290 ± 0.008	-0.172 ± 0.018
2.0	1.266 ± 0.007	1.276 ± 0.008	-0.175 ± 0.018
∞	1.254 ± 0.006	1.264 ± 0.008	-0.177 ± 0.019

r	$\mathrm{MSE}_{(\mathrm{P1})}$	$\mathrm{MSE}_{(\mathrm{P2})}$	MSE–ineff.
0.1	0.593 ± 0.007	0.594 ± 0.007	1.002 ± 0.001
0.25	0.836 ± 0.009	0.840 ± 0.010	1.005 ± 0.002
0.5	1.330 ± 0.014	1.342 ± 0.017	1.009 ± 0.002
1.0	2.750 ± 0.028	2.787 ± 0.035	1.013 ± 0.003
1.5	4.846 ± 0.050	4.918 ± 0.061	1.015 ± 0.003
2.0	7.674 ± 0.079	7.792 ± 0.097	1.015 ± 0.003
∞	∞	∞	1.016 ± 0.003

Table 9.5: Results of 10 Monte Carlo Simulations in case of AR(1) with Gumbel innovations and $\phi_1 = -0.5$; confer Example 9.3.25. [We additionally give 95% confidence intervals based on the central limit theorem.]

ϕ	r	$b_{(\mathrm{P1})}$	$b_{(\mathrm{P2})}$	M	$\mathrm{MSE}_{(\mathrm{P1})}$	$\mathrm{MSE}_{(\mathrm{P2})}$	MSE–ineff.
-0.75	0.1	2.096	2.098	-0.013	0.358	0.358	1.000
	0.25	1.433	1.435	-0.051	0.500	0.501	1.002
	0.5	1.149	1.152	-0.085	0.792	0.794	1.003
	1.0	1.017	1.020	-0.111	1.634	1.640	1.004
	1.5	0.986	0.988	-0.119	2.877	2.889	1.004
	2.0	0.975	0.977	-0.122	4.556	4.576	1.004
	∞	0.966	0.968	-0.125	∞	∞	1.004
-0.5	0.1	2.752	2.767	-0.061	0.596	0.597	1.002
	0.25	1.872	1.886	-0.102	0.841	0.845	1.005
	0.5	1.498	1.512	-0.138	1.338	1.351	1.010
	1.0	1.324	1.336	-0.167	2.765	2.804	1.014
	1.5	1.283	1.294	-0.175	4.872	4.948	1.016
	2.0	1.269	1.280	-0.179	7.716	7.840	1.016
	∞	1.257	1.268	-0.182	∞	∞	1.018
-0.1	0.1	3.234	3.265	-0.065	0.802	0.804	1.002
	0.25	2.192	2.220	-0.112	1.138	1.148	1.009
	0.5	1.750	1.778	-0.153	1.818	1.847	1.016
	1.0	1.545	1.569	-0.186	3.761	3.853	1.024
	1.5	1.496	1.519	-0.196	6.628	6.806	1.027
	2.0	1.480	1.502	-0.200	10.495	10.788	1.028
	∞	1.466	1.487	-0.203	∞	∞	1.029
0.1	0.1	3.234	3.265	-0.067	0.801	0.804	1.004
	0.25	2.192	2.220	-0.114	1.138	1.148	1.009
	0.5	1.750	1.778	-0.155	1.817	1.848	1.017
	1.0	1.545	1.569	-0.187	3.761	3.854	1.025
	1.5	1.496	1.519	-0.197	6.627	6.808	1.027
	2.0	1.480	1.502	-0.201	10.494	10.792	1.028
	∞	1.466	1.488	-0.205	∞	∞	1.030
0.5	0.1	2.768	2.777	-0.048	0.594	0.595	1.002
	0.25	1.884	1.892	-0.078	0.842	0.844	1.002
	0.5	1.510	1.518	-0.105	1.346	1.353	1.005
	1.0	1.334	1.341	-0.125	2.795	2.818	1.008
	1.5	1.292	1.299	-0.132	4.933	4.978	1.009
	2.0	1.278	1.284	-0.134	7.817	7.890	1.009
	∞	1.266	1.272	-0.136	∞	∞	1.010
0.75	0.1	2.073	2.074	-0.023	0.342	0.342	1.000
	0.25	1.415	1.416	-0.037	0.481	0.481	1.000
	0.5	1.136	1.137	-0.050	0.766	0.766	1.000
	1.0	1.005	1.006	-0.060	1.588	1.589	1.001
	1.5	0.974	0.974	-0.063	2.801	2.804	1.001
	2.0	0.963	0.964	-0.064	4.438	4.443	1.001
	∞	0.9539	0.9544	-0.065	∞	∞	1.001

Table 9.6: Robust non-adaptivity in terms of MSE–inefficiency for AR(1) with Gumbel distributed innovations in case of average transition neighborhoods ($t = \alpha = 1$) of contamination type; confer Example 9.3.25.

9.3.3.2 ARCH(1) with Scale

In case of ARCH(1) we have

$$H_{\theta,t}^{\tau} = \frac{X_{t-1}^2}{2(1+\alpha_1 X_{t-1}^2)} \quad (\geq 0) \tag{9.3.100}$$

where $\alpha_1 > 0$ such that

$$\alpha_1 < \exp\{-2(\log\sigma + \mathrm{E}\log|V_t|)\} \tag{9.3.101}$$

ARCH(1) with main parameters α_1 and nuisance parameter σ

In case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) the solution to problem (P1) reads

$$\tilde{\eta}(H_{\theta,1}, V) = A H_{\theta,1}\big(\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1})\big) w(H_{\theta,1}, V_1) \tag{9.3.102}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{c(H_{\theta,1})}{|\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1})|}\Big\} \tag{9.3.103}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} w\big[\mathrm{E}_\bullet\, w\big]^{-1} \qquad A^{-1} = \mathrm{E}\, H_{\theta,1} H_{\theta,1}^{\tau}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.104}$$

and $c(H_{\theta,1})$ and $r \in (0,\infty)$ are related via

$$r^2 = \mathrm{E}\, c^{-1}\, \mathrm{E}_\bullet\, \big(|\Lambda_f^{\mathrm{sc}} - z| - c\big)_+ \tag{9.3.105}$$

The solution to problem (P2) can be rewritten as

$$\tilde{\eta}(H_{\theta,1}, V) = A_H (H_{\theta,1} - M)(\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1})) w(H_{\theta,1}, V_1) \tag{9.3.106}$$

where

$$w(H_{\theta,1}, V_1) = \min\Big\{1, \frac{c(H_{\theta,1})}{|\Lambda_f^{\mathrm{sc}}(V_1) - z(H_{\theta,1})|}\Big\} \tag{9.3.107}$$

with

$$z(H_{\theta,1}) = \mathrm{E}_\bullet\, \Lambda_f^{\mathrm{sc}} w\big[\mathrm{E}_\bullet\, w\big]^{-1} \tag{9.3.108}$$

$$A_H^{-1} = \mathrm{E}(H_{\theta,1} - M) H_{\theta,1}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.109}$$

$$M = \mathrm{E}\, H\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \big/ \mathrm{E}\, \mathrm{E}_\bullet(\Lambda_f^{\mathrm{sc}} - z)^2 w \tag{9.3.110}$$

and $c(H_{\theta,1})$ and $r \in (0,\infty)$ are related via

$$r^2 = \mathrm{E}\, c^{-1}\, \mathrm{E}_\bullet(|\Lambda_f^{\mathrm{sc}} - z)| - c)_+ \tag{9.3.111}$$

Remark 9.3.26 Thus, in both problems, the inner (conditional) integrals are identical to the integrals which occur in one-dimensional i.i.d. scale. ////

Example 9.3.27 We consider lognormal distributed innovations with density

$$f_{\nu,\tau,\beta}(x) = \frac{1}{\sqrt{2\pi}\,\beta(x-\nu)} \exp\Big\{ -\tfrac{1}{2}\big[\log((x-\nu)/\tau)\big]^2/\beta^2 \Big\}\, \mathrm{I}(x \geq \nu) \qquad (9.3.112)$$

where we only consider $\tau = 1$ for simplicity. As noted in Example 5.1.9 (b) the lognormal scale family is closely related to the normal location family. Consequently, the inner (conditional) integrals correspond to the integrals of the corresponding normal location model; i.e., we get $z(H_{\theta,1}) \equiv 0$ as a possible choice. Moreover, the solution in case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) with starting radius $r \in (0, \infty)$ coincide with the solution in case of average conditional total variation neighborhoods ($* = v$, $t = \alpha = 1$) of radius $2r$. Due to (F4), which is $\mu_F = \int v dF(v) = 0$, we have to choose $\nu = -e^{\beta^2/2}$ and obtain

$$\sigma_F^2 = e^{\beta^2}(e^{\beta^2} - 1) \qquad \text{and} \qquad \int v^3 dF(v) = e^{1.5\beta^2}\big(e^{3\beta^2} - 3e^{\beta^2} + 2\big) \qquad (9.3.113)$$

In the considered situation the ARCH(1) process is (strictly) stationary if $\alpha_1 > 0$ and $\alpha_1 < \exp\{-2\,\mathrm{E}\log|V_t|\}$ where

$$\mathrm{E}\log|V_t| = \int_{-\infty}^{+\infty} \log\big|e^{\beta z} - e^{\beta^2/2}\big|\,\varphi(z)\,dz \qquad (9.3.114)$$

with φ the density of $\mathcal{N}(0,1)$. Using numerical integration we obtain the stationarity region plotted in Figure 9.1. To restrict the amount of data, we choose some typical cases for β and α_1; confer Table 9.7. We computed the results via Monte Carlo simulations. That is, we simulated an ARCH(1) process of length 110000 and threw away the first 10000 results due to stationary. We then computed the Lagrange multipliers contained in the optimally robust IC via crude Monte Carlo integration. We repeated this procedure 10, respectively 100 times. Using these replications we determined 95% confidence intervals based on the central limit theorem. In contrast to AR(1) and MA(1) (cf. Subsubsection 9.3.3.1), the amount of non-adaptivity may become really large, in particular, if we choose β and α_1 close to the boundaries of the stationarity region of the considered ARCH(1) process. If β and α_1 are beyond this region, this is denoted by "—" in Table 9.7. In case $\beta = 1.0$ we have $\exp\{-2\,\mathrm{E}\log|V_t|\} \approx 1.536$; i.e., $\alpha_1 = 1.5$, $\beta = 1.0$ is very close to the boundaries of the stationarity region. The very large MSE–inefficiencies (cf. Table 9.8), are mainly caused by very large minimax asymptotic MSEs in case of problem (P2). At $r = 0$ and $r = \infty$ the corresponding inefficiencies are

$$\frac{\mathrm{E}\,H_{\theta,1}^2}{\mathrm{Var}\,H_{\theta,1}} = \frac{\mathrm{E}\,H_{\theta,1}^2}{\mathrm{E}\,H_{\theta,1}^2 - (\mathrm{E}\,H_{\theta,1})^2} \qquad (9.3.115)$$

and

$$\frac{(\mathrm{E}\,|H_{\theta,1}|)^2}{(\mathrm{E}\,|H_{\theta,1} - \mathrm{med}(H_{\theta,1})|)^2} = \frac{(\mathrm{E}\,H_{\theta,1})^2}{(\mathrm{E}\,|H_{\theta,1} - \mathrm{med}(H_{\theta,1})|)^2} \qquad (9.3.116)$$

respectively. In case $\alpha_1 = 1.5$, $\beta = 1.0$ the numerators attain moderate values, however, the denominators are very small.

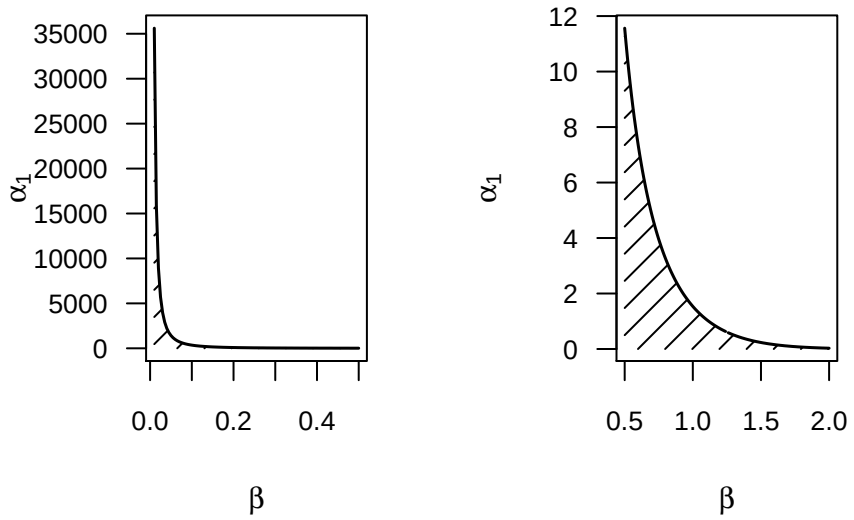

Figure 9.1: Stationarity region of ARCH(1) with lognormal innovations.

We again see that the MSE–inefficiency does not necessarily increase with increasing radius; confer also Example 9.2.26. Moreover, there seem to be situations where we get the largest non-adaptivity for very small radii maybe even for $r = 0$ (cf. $\alpha_1 = 1.5$, $\beta = 0.5$). ////

α_1	r	$\beta = 0.5$	$\beta = 1.0$	$\beta = 1.5$
0.1	0.0	1.295 ± 0.006	1.673 ± 0.007	5.036 ± 0.050
	0.1	1.369 ± 0.006	1.713 ± 0.009	5.138 ± 0.054
	0.5	1.485 ± 0.007	1.857 ± 0.013	5.537 ± 0.067
	1.0	1.458 ± 0.007	1.876 ± 0.015	5.765 ± 0.075
	1.5	1.429 ± 0.008	1.864 ± 0.016	5.845 ± 0.078
	2.0	1.412 ± 0.008	1.854 ± 0.016	5.873 ± 0.080
	∞	1.382 ± 0.008	1.834 ± 0.017	5.889 ± 0.082
0.5	0.0	1.676 ± 0.009	3.840 ± 0.040	—
	0.1	1.726 ± 0.009	3.847 ± 0.043	—
	0.5	1.738 ± 0.009	3.917 ± 0.054	—
	1.0	1.670 ± 0.009	3.955 ± 0.061	—
	1.5	1.630 ± 0.009	3.960 ± 0.063	—
	2.0	1.610 ± 0.009	3.956 ± 0.064	—
	∞	1.578 ± 0.009	3.944 ± 0.065	—
1.0	0.0	2.026 ± 0.007	9.333 ± 0.207	—
	0.1	2.046 ± 0.008	10.368 ± 0.259	—
	0.5	1.975 ± 0.010	14.391 ± 0.489	—
	1.0	1.886 ± 0.012	16.649 ± 0.651	—
	1.5	1.841 ± 0.013	17.488 ± 0.724	—
	2.0	1.819 ± 0.013	17.813 ± 0.757	—
	∞	1.785 ± 0.013	18.083 ± 0.790	—
1.5	0.0	2.332 ± 0.013	cf. Table 9.8	—
	0.1	2.327 ± 0.012		—
	0.5	2.212 ± 0.013		—
	1.0	2.115 ± 0.015		—
	1.5	2.068 ± 0.015		—
	2.0	2.045 ± 0.016		—
	∞	2.011 ± 0.016		—
5.0	0.0	4.900 ± 0.070	—	—
	0.1	5.035 ± 0.080	—	—
	0.5	5.441 ± 0.114	—	—
	1.0	5.598 ± 0.130	—	—
	1.5	5.631 ± 0.135	—	—
	2.0	5.635 ± 0.137	—	—
	∞	5.621 ± 0.138	—	—
10.0	0.0	26.043 ± 2.361	—	—
	0.1	35.078 ± 3.908	—	—
	0.5	86.709 ± 13.698	—	—
	1.0	134.788 ± 23.764	—	—
	1.5	160.984 ± 29.541	—	—
	2.0	174.922 ± 32.719	—	—
	∞	198.367 ± 38.533	—	—

Table 9.7: Robust non-adaptivity in terms of MSE–inefficiency for ARCH(1) with lognormal innovations in case of average transition neighborhoods ($t = \alpha = 1$) of contamination type; confer Example 9.3.27. [In case $\alpha_1 = 10.0$, $\beta = 0.5$, results are based on 100 Monte-Carlo simulations.]

r	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
0.0	92.0	130.8	153.8	163.6	194.7	320.8
0.1	177.3	289.3	361.8	409.7	505.4	1057.0
0.5	787.3	1491.0	2038.0	2422.0	3089.0	7703.0
1.0	1471	2875	4011	4798	6157	15800
1.5	1875	3700	5194	6223	8003	20690
2.0	2101	4163	5829	7025	9041	23450
∞	2526	5040	7123	8551	11020	28750

Table 9.8: Robust non-adaptivity in terms of MSE–inefficiency for ARCH(1) with lognormal innovations and parameters $\alpha_1 = 1.5$, $\beta = 1.0$ in case of average transition neighborhoods ($t = \alpha = 1$) of contamination type; confer Example 9.3.27. [Results are based on 100 Monte-Carlo simulations.]

9.3.4 Implementation Using `R`

We provide some functions for the computation of the standardized bias b and the minimax asymptotic MSE A, respectively A_H. These functions are included in the `R` scripts `AR1Gumbel.R`, `AR1SGumbel.R`, `ARCH1Lnorm.R` and `ARCH1SLnorm.R` which can be found in the directory "`.../RHome/library/ROptRegTS/scripts/tseries`" after installing our `R` bundle `RobASt` where `RHome` stands for the local home directory of `R`.

Remark 9.3.28 The computation of optimally robust ICs in case of regression-type time series models (like ARMA(p, q), ARCH(p), ...) fits well in the framework of our `R` package `ROptRegTS` (cf. Appendix D.4) and will be included in the near future. ////

Part V

Finite-Sample versus Asymptotic Estimator

In this final part asymptotics is checked against finite-sample results.

In case of one-dimensional location and simple regression Huber (1968) and Rieder (1989), respectively, have an exact finite-sample minimax theory based on robust tests for special capacities. The approach requires a particular pseudo-loss function in terms of under-/overshoot probabilities and provides minimax optimality among arbitrary estimators. However, it seems to be restricted to one real location, respectively regression parameter. The corresponding asymptotic minimax estimator for this special pseudo-loss function is derived in Rieder (1980) for the class of asymptotically linear estimators and in Rieder (1981a) for the class of arbitrary estimators. Hence, the asymptotics may be checked against the finite-sample result for contamination and total variation neighborhoods of fixed size. The comparison asymptotic versus finite-sample results, however, requires the computation of the exact finite-sample risks. The analytic evaluation of the finite-sample risks turns out very difficult or even impossible for sample size $n \geq 3$. Thus, we have invented an algorithm based on the fast Fourier transform (FFT) to determine the exact finite-sample distribution (and corresponding finite-sample risks) of these differently robust estimators.

Two interesting findings are: The (first order) asymptotics is too optimistic and the convergence towards the asymptotic values is better in case of total variation than in case of contamination neighborhoods.

We now briefly survey the finite-sample and asymptotic minimax theory.

Huber's (1968) Treatment

Huber (1968) considers one-dimensional location and employs contamination/total variation neighborhoods of fixed size. Based on a minimax test theory for special capacities which has been developed in Huber (1965), he derives a finite-sample minimax solution which minimizes the maximum probability that the estimate exceeds, or falls below, the true value of the parameter by more than some fixed amount. We call this special pseudo-loss function *under-/overshoot confidence risk*. A summary of these results is also given in Chapter 10 of Huber (1981).
In case of one-dimensional normal location and sample size $n = 2$ Huber (1964) has been able to determine the finite-sample under-/overshoot confidence risk analytically. His results for normal location are confirmed and extended by our analytical and numerical evaluations given in Sections 11.3 and 11.4.

Rieder's (1989) Treatment

Rieder (1989) extends the results of Huber (1968) to simple linear regression through the origin with random regressor where he employs fixed sized neighborhoods of two types: unconditional (errors-in-variables) and conditional (error-free-variables) contamination/total variation neighborhoods. For the case of finitely many deterministic regressors and conditional neighborhoods a preliminary extension has been indicated by Huber (1983).
In case of normal errors and unconditional neighborhoods the finite-sample minimax estimator corresponds to the Hampel-Krasker estimator. This contradicts the conjecture of Huber (1983) (cf. Sections 1, 7 and Rejoinder) that the Hampel-Krasker estimator has no optimality in case of errors-in-variables. Moreover, in case

of normal errors and conditional neighborhoods both Hampel-Krasker and Huber estimators arise as minimax solutions for suitable contamination curves.

Rieder's (1980) and (1981) Treatment

The asymptotic counterpart to the finite-sample results given above is derived by Rieder (1980) using results of robust asymptotic testing (cf. Rieder (1978)). To obtain these asymptotic results, the size of the contamination/total variation neighborhoods as well as the width of the considered confidence interval are assumed to shrink at a rate of $\sqrt{n}$.
The results given in Rieder (1980), in contrast to the finite-sample results, apply to an arbitrary parameter. In particular, they apply to linear regression with random regressor and unconditional contamination/total variation neighborhoods.
By evaluating the asymptotic under-/overshoot confidence risk uniformly, supereffi-cient estimators are cut out not be assumption but by a high maximum risk. Doing so, Rieder (1981a) establishes a local asymptotic minimax bound for arbitrary estimators. Thus, the asymptotic minimax estimator derived in Rieder (1980) is, in this sense, minimax for all arbitrary estimators.
In case of conditional contamination neighborhoods Rieder (1994) (cf. Subsection 7.5.2) verifies that the "Huber estimator, when compared with the Hampel-Krasker estimator, is minimax in a similar though slightly smaller model: square, as opposed to mean, conditional contamination" (cf. Remark 7.5.17 (a), ibid.). Since minimizing the asymptotic under-/overshoot confidence risk is equivalent to minimizing the asymptotic variance subject to a bias bound (cf. Lemma 10.3.5), this result also applies to our setup.

Our Treatment

In Chapter 10 we first introduce the finite-sample and asymptotic setup; confer Section 10.1. Second, we present the derivation of the finite-sample minimax regression estimator which for the most part is based on Rieder (1989) and Section 1 of Rieder (1995); confer Section 10.2. In addition to Rieder (1989), we prove the admissibility of the finite-sample minimax estimator (cf. Theorem 10.2.10 (b)) and give some formal arguments for the derivation of estimates in boundary cases (cf. Subsection 10.2.4).

Third, we apply the asymptotic results of Rieder (1980) to linear regression and give an extension of the asymptotic theory to conditional regression neighborhoods; confer Section 10.3. We derive the asymptotic minimax estimator without any connection to the corresponding robust testing problem of Rieder (1978). We instead use optimizations arguments as in Ruckdeschel and Rieder (2004); confer Subsection 10.3.3.

As mentioned above, these results are the asymptotic analogues to the finite-sample case. Thus, we are able to check the asymptotics against finite-sample results obtained for fixed neighborhoods at least for this special pseudo-loss function and a one-dimensional location, respectively regression parameter provided the finite-sample risks can be calculated numerically with high accuracy.

In Chapter 11 we specialize the results of Chapter 10 from simple regression to one-dimensional location and from error distributions with finite Fisher informa-

tion to normal errors. We consider contamination ($* = c$) as well as total variation neighborhoods ($* = v$); confer Sections 11.1 and 11.2. In this setup the finite-sample and the asymptotic minimax estimators are of the same form and may be identified by the corresponding optimal clipping bounds. Hence, we first compare finite-sample against asymptotic optimal clipping bounds by means of Taylor expansions. It turns out, that the convergence towards the asymptotic values is from below; i.e., the (first order) asymptotics is too optimistic. Moreover, in case of contamination neighborhoods ($* = c$) the speed of convergence towards the asymptotic values is of order $n^{-1/2}$ whereas in case of total variation neighborhoods ($* = v$) it is of order n^{-1}. Thus, as for the clipping bounds, there is a clear difference between these two types of neighborhoods which, we conjecture, is caused by the higher symmetry of total variation neighborhoods. In the course of these investigations, we also obtain $\mathrm{O}(n^{-1/2})$-corrected ($* = c$) and $\mathrm{O}(n^{-1})$-corrected ($* = v$) asymptotic optimal clipping bounds, respectively.

In Section 11.3 we investigate the exact finite-sample risk (cf. Subsection 11.3.1) directly and introduce two algorithms based on the fast Fourier transform (cf. Subsection 11.3.2). These provide numerical approximations that are very accurate. We check the precision of these algorithms by analytic calculations, which are available for sample size $n = 2$ (the simpler case $n = 1$ is not considered!), as well as by numerical simulations; confer Subsubsection 11.3.2.3. We also use these algorithms to check the results obtained by means of other approximations: Edgeworth expansions up to first and second order (cf. Ibragimov (1967)) and saddlepoint approximations (cf. Field and Ronchetti (1990)); confer Subsection 11.3.3. As we see, the Edgeworth expansions up to second order as well as the saddlepoint approximations yield good approximations of the finite-sample risk for sample sizes down to about 5 in this confidence setup for robust normal location.

The main application of our algorithms is the numerical comparison between finite-sample and asymptotic results; confer Section 11.4. First, we compare the finite-sample, the asymptotic and the O-corrected asymptotic optimal clipping bounds; confer Subsubsections 11.4.1.1 and 11.4.2.1. It turns out, that there are clear differences between the finite-sample and the asymptotic optimal clipping bounds. However, for moderate sample sizes n of about 10 ($* = v$), respectively 20 ($* = c$) the O-corrected asymptotic optimal clipping bounds are already very close to the finite-sample ones.

Second, we compare the finite-sample risks of the finite-sample and the asymptotic minimax estimators as well as of the estimator which is based on the O-corrected asymptotic optimal clipping bound; confer Subsubsections 11.4.1.2 and 11.4.2.2. Although there are clear differences between the optimal clipping bounds, the differences (in absolute values) between the corresponding finite-sample risks are only small. In addition, the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002) indicates that the speed of convergence of the finite-sample risk towards the asymptotic risk is of the same order as that of the optimal clipping bounds. That is, in case of contamination neighborhoods ($* = c$) the speed of convergence seems to be of order $n^{-1/2}$ whereas in case of total variation neighborhoods ($* = v$) it is seems to be of order n^{-1}.

Since there are (a priori) only small differences in absolute values, we also investigate relative values; i.e., how much efficiency do we lose if we take the asymptotic optimal clipping bound, respectively the O-corrected asymptotic optimal clipping bound instead of the finite-sample optimal bound. The numerical results show that the maximum relative risks of the estimator which is based on the O-corrected asymptotic optimal clipping bound are very small. Moreover, already for moderate sample sizes ($* = c$: $20-50$, $* = v$: 10) there is no big difference between all three estimators (maximum efficiency loss $< 5\%$).

As the size of contamination in most applications is unknown, respectively unknown except to belong to some interval, we also determine the least favorable sizes (radii) and the corresponding inefficiencies. These are defined analogously to Section 2.2. The efficiency losses stay well below 40% ($\rho = 0$), 20% ($\rho = 1/2$) and 10% ($\rho = 1/3$), respectively in all cases considered where the largest inefficiencies occur in the asymptotic case. Since the normal distribution is symmetric to 0, the subefficiencies for contamination ($* = c$) and total variation neighborhoods ($* = v$) are identical in the asymptotic setup. However, in the finite-sample setup the inefficiencies are slightly larger in case of total variation neighborhoods which is probably caused by the faster convergence towards the asymptotic values.

Third, we compute the finite-sample distribution of the finite-sample minimax estimator and compare it to the normal distribution that is closest in Kolmogorov distance d_κ; confer Subsubsections 11.4.1.3 and 11.4.2.3. Our numerical results indicate that the speed of convergence in case of total variation neighborhoods ($* = v$) is not only faster at 0 (corresponds to the finite-sample risk) but uniformly over the whole support of the finite-sample distribution of the finite-sample minimax estimator. But, the differences occur for very small sample sizes and already for $n \geq 10$ we obtain $d_\kappa < 0.02$ in both cases.

In Chapter 12 we specialize the results of Chapter 10 from arbitrary errors with finite Fisher information to normal errors. We consider unconditional (errors-in-variables) neighborhoods of contamination ($* = c, t = 0$) and total variation ($* = v, t = 0$) type as well as conditional (error-free-variables) neighborhoods of contamination ($* = c, t = \varepsilon$) and total variation ($* = v, t = \delta$) type; confer Sections 12.1 and 12.2.

Similar to normal location, the finite-sample and the asymptotic minimax estimators are of the same form and may be identified by the corresponding optimal clipping bounds ($* = c, v; t = 0$) and functions ($* = c, v; t = \varepsilon, \delta$), respectively. Therefore, we first compare finite-sample against asymptotic optimal clipping bounds and functions, respectively. As it turns out, the results are analogously to normal location. That is, we obtain by means of Taylor expansions that the speed of convergence is of order $n^{-1/2}$ in case of contamination neighborhoods ($* = c$; $t = 0, \varepsilon$) whereas it is of order n^{-1} in case of total variation neighborhoods ($* = v$; $t = 0, \delta$). In all cases the convergence towards the asymptotic values is from below; i.e., the (first order) asymptotics is too optimistic. In addition, these Taylor expansions yield $\mathrm{O}(n^{-1/2})$-corrected ($* = c$; $t = 0, \varepsilon$) and $\mathrm{O}(n^{-1})$-corrected ($* = v$; $t = 0, \delta$) asymptotic optimal clipping bounds and functions, respectively.

Moreover, we again derive algorithms for the computation of the finite-sample risk which are based on FFT and are similar to the algorithms used in one-dimensional normal location; confer Subsections 12.1.3 and 12.2.3. Because of the dependence on the regressor we now, contrary to normal location, do not have analytic results for sample size $n = 2$ to check the precision of our algorithms. However, various cross checks and results obtained via numerical simulations indicate that our algorithms are again very accurate.

In Section 12.3 we make several numerical comparisons. In the regression context, contrary to location, the problem of choosing ideal regressor distributions K arises. For the purpose of this section, we consider $K = \frac{1}{3}\left(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\right)$ and $K = \mathrm{Unif}\left([-1, 2]\right)$.

First, we treat unconditional neighborhoods; confer Subsections 12.3.1 and 12.3.2. We begin with a comparison of the finite-sample, the asymptotic and the O-corrected asymptotic optimal clipping bounds; confer Subsubsections 12.3.1.1 and 12.3.2.1. Similar to normal location, there are clear differences between the finite-sample and the asymptotic optimal clipping bounds. However, for moderate sample sizes n of about 20 the O-corrected asymptotic optimal clipping bounds are already very close to the finite-sample ones.

Next, we go for the finite-sample risks. By means of our FFT algorithms we compare the finite-sample risk of the finite-sample and the asymptotic minimax estimators as well as of the estimator which is based on the O-corrected asymptotic optimal clipping; confer Subsubsections 12.3.1.2 and 12.3.2.2. Again, it turns out that in most cases the differences (in absolute values) between the corresponding finite-sample risks are only small.

As in case of normal location, numerical investigations of the speed of convergence of the finite-sample risks towards the asymptotic risks indicate that it is of order $n^{-1/2}$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$ whereas it is of order n^{-1} in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

Subsequently, we consider relative risks. As before the results are similar to normal location and only the numerical values are different. In the cases considered, the maximum efficiency loss of the asymptotic minimax estimator stays well below 10% for sample size $n \geq 50$ $(* = c,\, t = 0)$, respectively $n \geq 20$ $(* = v,\, t = 0)$. In case of the estimator which is based on the O-corrected asymptotic optimal clipping bound small values of n $(* = c,\, t = 0\colon n < 10\,,\ * = v,\, t = 0\colon n < 20)$ may lead to very large subefficiencies. This is caused by the fact that the O-correction is too large for small sample sizes and may even lead to irregular (negative) clipping bounds. Hence, to obtain valid bounds we replace these values by 0. However, for sample sizes $n \geq 10$ $(* = c,\, t = 0)$, respectively $n \geq 20$ $(* = v,\, t = 0)$ the efficiency losses stay well below 5% $(* = c,\, t = 0)$ and 10% $(* = v,\, t = 0)$, respectively.

In many applications the size of contamination is unknown, respectively unknown except to belong to some interval. Thus, we again determine the least favorable sizes (radii) and the corresponding inefficiencies which are defined analogously to Section 2.2. In the considered cases, the efficiency losses are slightly larger than in normal location but, stay well below 60% $(\rho = 0)$, 30% $(\rho = 1/2)$

and 15% ($\rho = 1/3$), respectively. By the symmetry of the normal distribution, the subefficiencies for unconditional contamination ($* = c$, $t = 0$) and total variation ($* = v$, $t = 0$) neighborhoods are identical in the asymptotic setup where the largest inefficiencies occur. Hence, it is not surprising, that the efficiency losses are slightly larger in case of unconditional total variation neighborhoods where the convergence towards the asymptotic values is faster.

Finally, we compute the finite-sample distribution of the finite-sample minimax estimator. It is compared with the normal distribution which is closest in Kolmogorov distance d_κ; confer Subsubsections 12.3.1.3 and 12.3.2.3. Analogously to normal location, our numerical results indicate that the speed of convergence in case of unconditional total variation neighborhoods ($* = v$, $t = 0$) is not only faster at 0 (corresponds to the finite-sample risk) but uniformly over the whole support of the finite-sample distribution of the finite-sample minimax estimator. But, again the differences occur for very small sample sizes and already for $n \geq 10$ we have $d_k < 0.02$ in both cases.

Second, we consider conditional (error-free-variables) neighborhoods; confer Subsections 12.3.3 and 12.3.4. In Subsubsections 12.3.3.1 and 12.3.4.1 we determine the contamination curves for which the Hampel-Krasker estimator is the finite-sample, respectively asymptotic minimax estimator. As it turns out, the corresponding contamination curves are very similar for small values of the regressor. However, for large values the amount of contamination suggested by the asymptotic approach is unrealistically large and can not be realized for finite samples. In contrast, the amount of contamination is bounded in the finite-sample setup. This indicates, that (first order) asymptotics provides only a bad approximation for large regressors.

But, one may also obtain the Huber estimator as finite-sample, respectively asymptotic minimax estimator for suitable contamination curves. The computations in Subsubsections 12.3.3.2 and 12.3.4.2 show that the corresponding contamination curves are again very similar for small values of the regressor. However, for large values the amount of contamination has to tend to 0 very fast in the finite-sample setup whereas in the asymptotic setup again an unrealistic amount of contamination is needed. That is, (first order) asymptotics again seems to provide only bad approximations for large regressors.

For the remaining numerical comparisons we choose constant contamination curves and ideal regressor distributions $K = \frac{1}{3}\left(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\right)$ and $K = \mathrm{Unif}([-1, 2])$.

As we see in Subsubsections 12.3.3.3 and 12.3.4.3, the optimal clipping function is equal to 0 for small values of the regressor in case of constant contamination curves. This is true for the finite-sample as well as the asymptotic setup. However, for large values the clipping function is unbounded in the asymptotic setup whereas it is tending to 0 in the finite-sample setup. Moreover, the O-corrected asymptotic optimal clipping functions are negative for (very) small as well as for (very) large values of the regressor. But, for moderate values the approximations work quite well down to sample size $n = 10$ ($* = c$, $t = \varepsilon$) and $n = 5$ ($* = v$, $t = \delta$), respectively. These results once more indicate that (first order) asymptotics provides a bad approximation for large regressors.

In Subsubsections 12.3.3.4 and 12.3.4.4 we use our FFT algorithm to compare the finite-sample risks of the finite-sample and asymptotic minimax estimator as well as of the estimator which is based on the O-corrected asymptotic optimal clipping functions (negative values are set to 0). Like in normal location, respectively unconditional neighborhoods the differences (in absolute values) between the corresponding finite-sample risks are only small in all cases considered. Moreover, as before, the speed of convergence towards the asymptotic risk seems to be of order $n^{-1/2}$ in case of the contamination type neighborhoods $(* = c, t = \varepsilon)$ whereas it seems to be of order n^{-1} in case of the total variation type neighborhoods $(* = v, t = \delta)$.

We conclude this chapter with the computation of the finite-sample distribution of the finite-sample minimax estimator. As before, we compare it with the normal distribution which is closest in Kolmogorov distance d_κ; confer Subsubsections 12.3.3.5 and 12.3.4.5. This time, the differences between contamination and total variation type neighborhoods are smaller than in the situations before. More precisely, in both cases we obtain $d_\kappa < 0.02$ already for $n \geq 5$.

Chapter 10

Finite-Sample and Asymptotic Minimax Regression Estimator

First, we introduce the finite-sample and asymptotic setup; confer Section 10.1. We then present the derivation of the finite-sample minimax estimator (cf. Subsection 10.2) which for the most part is based on Rieder (1989) and Section 1 of Rieder (1995). In Subsection 10.3 we derive the asymptotic minimax estimator by means of optimizations arguments as in Ruckdeschel and Rieder (2004); i.e., without any connection to the robust testing problem of Rieder (1978).

10.1 Setup

We consider simple linear regression through the origin,

$$y = x\theta + u \tag{10.1.1}$$

with y the observed variable, x the random regressor and u the statistical error where the regressor x and the error u are assumed to be stochastically independent. Given a fixed number of i.i.d. observations (x_i, y_i), $i = 1, \ldots, n$ $(n \in \mathbb{N})$, the unknown slope parameter $\theta \in \mathbb{R}$ has to be estimated. We make the following assumptions about the ideal error law F: The Fisher information of location for F is finite; i.e., F has an absolutely continuous density f and

$$\mathcal{I}_f^{\mathrm{loc}} = \int \left(\Lambda_f^{\mathrm{loc}}(u)\right)^2 f \, \lambda(du) < \infty \tag{10.1.2}$$

with

$$\Lambda_f^{\mathrm{loc}}(u) := -f'(u)/f(u) \tag{10.1.3}$$

where $\lambda(du)$ denotes the Lebesgue measure on $(\mathbb{R}, \mathbb{B})$. Furthermore, we assume,

$$\log f \quad \text{concave} \tag{10.1.4}$$

On the ideal regressor law K we impose the conditions

$$K(x=0)=0 \tag{10.1.5}$$

$$\mathcal{K} := \int x^2 K(dx) < \infty \tag{10.1.6}$$

Remark 10.1.1 **(a)** The log-concavity of f is needed for the construction of the finite-sample minimax-estimator (monotony of the loglikelihood); confer Rieder (1989), proof of Theorem 4.1. Condition (10.1.5) is used to prove the absolute continuity Lemma 10.2.4, which immediately implies a lower risk bound for equivariant estimators in the finite-sample case.

(b) The finiteness of the Fisher information of location and (10.1.6) imply L_2 differentiability of the linear model (10.1.1); confer Theorem 10.1.2 below. This smoothness of the considered ideal center model implies local asymptotic normality and therefore is fundamental for the application of the asymptotic theory of robustness; confer Chapter 1. ////

By the stochastic independence of x and u, the ideal model distributions read

$$P_\theta(dx,dy) = f(y-x\theta)\lambda(dy)K(dx) \tag{10.1.7}$$

Theorem 10.1.2 *Assume $\mathcal{I}_f^{\mathrm{loc}} < \infty$ and (10.1.6). Then, the linear model (10.1.1) is L_2 differentiable at every $\theta \in \mathbb{R}$, with L_2 derivative Λ_θ and Fisher information $\mathcal{I}_\theta$ given by*

$$\Lambda_\theta(x,y) = x\Lambda_f^{\mathrm{loc}}(y-x\theta) \qquad \mathcal{I}_\theta = \mathcal{K}\,\mathcal{I}_f^{\mathrm{loc}} \tag{10.1.8}$$

Proof Special case of Theorem 2.4.6 of Rieder (1994). ////

In addition, we allow that the law of the n i.i.d. observations (x_i, y_i) is not necessary P_θ exactly. Therefore, we introduce two types of neighborhoods of P_θ.

Definition 10.1.3 **(a)** Unconditional (errors-in-variables) neighborhoods $\mathcal{U}_{cv,0}(\theta)$ *for given numbers $\varepsilon, \delta \in [0,1)$ with $\varepsilon + \delta \in (0,1)$ are,*

$$\mathcal{U}_{cv,0}(\theta) = \big\{ Q \in \mathcal{M}_1(\mathbb{B}^2) \,\big|\, Q(dx,dy) \geq (1-\varepsilon)P_\theta(dx,dy) - \delta \big\} \tag{10.1.9}$$

(b) Conditional (error-free-variables) neighborhoods $\mathcal{U}_{cv,\varepsilon\delta}(\theta)$ *consist of all probability measures $Q(dx,dy) = Q(dy\,|\,x)\,Q(dx)$ on $(\mathbb{R}^2, \mathbb{B}^2)$ whose regressor marginals coincide with the ideal $K(dx)$ while the conditional distributions $Q(dy\,|\,x)$ of y given x may be distorted, such that*

$$Q(dy\,|\,x) \geq \big(1-\varepsilon(x)\big) f(dy-x\theta)\lambda(dy) - \delta(x) \qquad \text{a.e. } K(dx) \tag{10.1.10}$$

with Borel measurable contamination curves $\varepsilon, \delta \colon \mathbb{R} \to [0,1]$ such that $\varepsilon(x)+\delta(x) \in (0,1)$ a.e. $K(dx)$.

Remark 10.1.4 (a) These neighborhoods include contamination ($\delta \equiv 0$; $t = 0, \varepsilon\delta$) as well as total variation neighborhoods ($\varepsilon \equiv 0$; $t = 0, \varepsilon\delta$). Moreover, also deviations from the ideal model that may be caused by an actual nonlinearity of the regression are to some extent covered by these neighborhoods.

(b) In the asymptotic setup (cf. Section 10.3) we have to replace ε and δ by $\varepsilon/\sqrt{n}$ and $\delta/\sqrt{n}$ ($t = 0, \varepsilon\delta$), respectively. Moreover, if we want to check the asymptotics against exact finite-sample results for sample size n, we have to identify ε and δ with $\varepsilon/\sqrt{n}$ and $\delta/\sqrt{n}$, respectively in the finite-sample setup. ////

The simple linear regression model (10.1.1) and the corresponding neighborhoods $\mathcal{U}_{cv,0}(\theta)$ and $\mathcal{U}_{cv,\varepsilon\delta}(\theta)$ are invariant under the group of regression transforms

$$g_\theta(x, y) = (x, y + x\theta) \tag{10.1.11}$$

because the image measure of any probability $Q(dx, dy) = Q(dy \,|\, x)\, Q(dx)$ under g_θ is given by $\big(g_\theta(Q)\big)(dx, dy) = Q(dy - x\theta \,|\, x)Q(dx)$; confer Subsection 7.2.3 of Rieder (1994). That is,

$$\mathcal{U}_{cv,t}(\theta) = g_\theta\big(\mathcal{U}_{cv,t}(0)\big) \qquad t = 0,\ \varepsilon\delta \tag{10.1.12}$$

10.2 Finite-Sample Minimax Regression Estimator

In this section we present the finite-sample results mentioned above, following the lines of Rieder (1989) and Section 1 of Rieder (1995), respectively.

10.2.1 Lower Risk Bound for Equivariant Estimators

We consider any Borel measurable function $S : \mathbb{R}^{2n} \to \mathbb{R}$ as estimator for the unknown slope parameter $\theta \in \mathbb{R}$ at fixed sample size n. Moreover, also under deviations from the ideal model distribution P_θ the n observations are assumed to be i.i.d. and Q_θ^n stands for the n-fold product measure of $Q_\theta \in \mathcal{U}_{cv,t}(\theta)$ ($t = 0, \varepsilon\delta$). Since we want to apply results of robust testing, we arrive at the following risk, which is closely related to a confidence interval around the estimator.

Definition 10.2.1 *The risk of an estimator* $S : \mathbb{R}^{2n} \to \mathbb{R}$ *is*

$$\mathrm{Risk}\,(S; t) = \sup_{\theta\in\mathbb{R}} \mathrm{Risk}_\theta(S; t) \qquad (t = 0,\ \varepsilon\delta) \tag{10.2.1}$$

with

$$\mathrm{Risk}_\theta(S; t) = \sup_{Q_\theta \in \mathcal{U}_{cv,t}(\theta)} \max\big\{Q_\theta^n(S > \theta + \tau),\ Q_\theta^n(S < \theta - \tau)\big\} \tag{10.2.2}$$

for a given constant $\tau \in (0, \infty)$.

Thus, the risk of an estimator is always not larger than 50% since we may randomize between $\pm\infty$ with equal probability, without looking at the observations. Although

the result will hold for arbitrary estimators, it is best visible for estimators that are equivariant in the following sense.

Definition 10.2.2 *An estimator* $S : \mathbb{R}^{2n} \to \mathbb{R}$ *is called* equivariant *if*

$$S \circ G_\theta = S + \theta \qquad \theta \in \mathbb{R} \tag{10.2.3}$$

for the transformations

$$G_\theta\big((x_1, y_1)^\tau, \ldots, (x_n, y_n)^\tau\big) = \big(g_\theta(x_1, y_1)^\tau, \ldots, g_\theta(x_n, y_n)^\tau\big)^\tau \tag{10.2.4}$$

Remark 10.2.3 For estimators S that are equivariant in the sense of this definition we get for $t = 0$, $\varepsilon\delta$

$$\mathrm{Risk}_\theta(S; t) = \max\left\{ \sup_{Q_{-\tau} \in \mathcal{U}_{cv,t}(-\tau)} Q^n_{-\tau}(S > 0), \sup_{Q_\tau \in \mathcal{U}_{cv,t}(\tau)} Q^n_\tau(S < 0) \right\} \tag{10.2.5}$$

where $\mathcal{U}_{cv,t}(-\tau) = g_{-\theta-\tau}\big(\mathcal{U}_{cv,t}(\theta)\big)$ and $\mathcal{U}_{cv,t}(\tau) = g_{-\theta+\tau}\big(\mathcal{U}_{cv,t}(\theta)\big)$ by (10.1.12); confer proof of Proposition 2.2 in Rieder (1989). That is, $\mathrm{Risk}_\theta(S; t) = \mathrm{Risk}\,(S; t)$ for all $\theta \in \mathbb{R}$. ////

Under the assumptions of the following lemma the conditional distribution of an equivariant estimator S given $x_1, \ldots, x_n$ is absolutely continuous.

Lemma 10.2.4 *Consider* $S \colon \mathbb{R}^{2n} \to R$ *equivariant and* Q *a probability measure on* $(\mathbb{R}^2, \mathbb{B}^2)$ *such that*

$$Q(x = 0) = 0 \tag{10.2.6}$$

$$Q(dy \,|\, x) \ll \lambda(dy) \qquad \text{a.e. } Q(dx) \tag{10.2.7}$$

Then, under Q^n, *the conditional distribution of* S *given* $x_1, \ldots, x_n$ *satisfies,*

$$Q^n(S \in ds \,|\, x_1, \ldots, x_n) \ll \lambda(ds) \qquad \text{a.e. } Q^n(dx_1, \ldots, dx_n) \tag{10.2.8}$$

Proof Rieder (1989), Lemma 2.1. ////

Remark 10.2.5 As noted on page 281 (Note 1) of Huber (1981) for the location case and on page 176 of Rieder (1995) for the regression case, the lemma would not be true with *absolute continuity* replaced by *continuity*. ////

If there exists a least favorable pair $(R'_{-\tau,t}, R''_{\tau,t})$ in the sense of Huber (1968), respectively Rieder (1977) for testing $\mathcal{U}_{cv,t}(-\tau)$ versus $\mathcal{U}_{cv,t}(\tau)$ $(t = 0, \varepsilon\delta)$ that satisfies the assumptions of the preceding lemma, the following lower risk bound for equivariant estimators may be obtained as the equal error probabilities of the minimax test $\tilde{\phi}$ between the product measures $(R'_{-\tau,t})^n$ and $(R''_{\tau,t})^n$; i.e., $\int \tilde{\phi}\, d(R'_{-\tau,t})^n = \int (1 - \tilde{\phi})\, d(R''_{\tau,t})^n = \alpha$.

Proposition 10.2.6 *Assume* $(R'_{-\tau,t}, R''_{\tau,t})$ *is a least favorable pair for* $\mathcal{U}_{cv,t}(-\tau)$ *versus* $\mathcal{U}_{cv,t}(\tau)$ $(t = 0, \varepsilon\delta)$ *and that conditions (10.2.6) and (10.2.7) are fulfilled by* $R'_{-\tau,t}$ *or by* $R''_{\tau,t}$. *Then, every equivariant estimator* $S \colon \mathbb{R}^{2n} \to \mathbb{R}$ *satisfies* $\mathrm{Risk}_\theta(S; t) = \mathrm{Risk}\,(S; t) \geq \alpha$ *for all* $\theta \in \mathbb{R}$.

Proof Rieder (1989), Proposition 2.2. ////

10.2.2 Least Favorable Pairs

10.2.2.1 Unconditional neighborhoods

In case of unconditional (errors-in-variables) neighborhoods a least favorable pair $(R'_{-\tau,0}, R''_{\tau,0})$ for $\mathcal{U}_{cv,0}(-\tau)$ versus $\mathcal{U}_{cv,0}(\tau)$ may be read off from Huber (1968) and Rieder (1977),

$$R'_{-\tau,0}(dx,dy) = r_0(x,y)\lambda(dy)K(dx) \tag{10.2.9}$$
$$R''_{\tau,0}(dx,dy) = \pi_0(x,y)r_0(x,y)\lambda(dy)K(dx) \tag{10.2.10}$$

with density

$$r_0(x,y) = \begin{cases} \frac{1-\varepsilon}{\varepsilon+(1+d')\delta}\big[(\varepsilon+\delta)f(y+\tau x)+\delta f(y-\tau x)\big] & \text{if } \Delta < d' \\ (1-\varepsilon)f(y+\tau x) & \text{if } d' \le \Delta \le d'' \\ \frac{1-\varepsilon}{d''(\varepsilon+\delta)+\delta}\big[\delta f(y+\tau x)+(\varepsilon+\delta)f(y-\tau x)\big] & \text{if } \Delta > d'' \end{cases} \tag{10.2.11}$$

and likelihood ratio

$$\pi_0(x,y) = \max\big\{d', \min\{\Delta, d''\}\big\} \tag{10.2.12}$$

where $\Delta = \Delta(x,y) = \frac{f(y-\tau x)}{f(y+\tau x)}$ and d', $d'' \in [0,\infty]$ are the unique solutions of the equations

$$\frac{\varepsilon+(1+d')\delta}{1-\varepsilon} = \int \big[d'-\Delta(x,y)\big]_+ f(y+\tau x)\lambda(dy)K(dx) \tag{10.2.13}$$
$$\frac{(\varepsilon+\delta)d''+\delta}{1-\varepsilon} = \int \big[\Delta(x,y)-d''\big]_+ f(y+\tau x)\lambda(dy)K(dx) \tag{10.2.14}$$

By Lemma 4.3 of Rieder (1977) we get the restriction

$$\frac{\varepsilon+2\delta}{1-\varepsilon} < \int \big(f(y+\tau x)-f(y-\tau x)\big)_+ \lambda(dy)K(dx) \tag{10.2.15}$$

which is equivalent to $\mathcal{U}_{cv,0}(-\tau)\cap\mathcal{U}_{cv,0}(\tau) = \emptyset$.

10.2.2.2 Conditional neighborhoods

In case of conditional (error-free-variables) regression neighborhoods we get a least favorable pair for $\mathcal{U}_{cv,\varepsilon\delta}(-\tau)$ versus $\mathcal{U}_{cv,\varepsilon\delta}(\tau)$ by pasting together x-sections that may be obtained according to Huber (1968) and Rieder (1977); i.e.,

$$R'_{-\tau,\varepsilon\delta}(dx,dy) = r_{\varepsilon\delta}(x,y)\lambda(dy)K(dx) \tag{10.2.16}$$
$$R''_{\tau,\varepsilon\delta}(dx,dy) = \pi_{\varepsilon\delta}(x,y)r_{\varepsilon\delta}(x,y)\lambda(dy)K(dx) \tag{10.2.17}$$

where the density $r_{\varepsilon\delta}(x,y)$ and the likelihood ratio $\pi_{\varepsilon\delta}(x,y)$ are defined analogously to (10.2.11) and (10.2.12) but with ε, δ, d' and d'' replaced by $\bar{\varepsilon}(x), \bar{\delta}(x), d'(x)$

and $d''(x)$, respectively. The Borel measurable functions $\bar{\varepsilon}, \bar{\delta}\colon \mathbb{R} \to [0,1]$ satisfy $\bar{\varepsilon}(x) \le \varepsilon(x)$, $\bar{\delta}(x) \le \delta(x)$ as well as,

$$\frac{\bar{\varepsilon}(x) + 2\bar{\delta}(x)}{1 - \bar{\varepsilon}(x)} = \min\left\{\frac{\varepsilon(x) + 2\delta(x)}{1 - \varepsilon(x)}, \int \big(f(y + \tau x) - f(y - \tau x)\big)_+ \lambda(dy)\right\} \quad (10.2.18)$$

and $d'(x), d''(x) \in [0, \infty]$ are the unique solutions of the equations

$$\frac{\bar{\varepsilon}(x) + (1 + d'(x))\bar{\delta}(x)}{1 - \bar{\varepsilon}(x)} = \int \big[d'(x) - \Delta(x,y)\big]_+ f(y + \tau x)\lambda(dy) \quad (10.2.19)$$

$$\frac{(\bar{\varepsilon}(x) + \bar{\delta}(x))d''(x) + \bar{\delta}(x)}{1 - \bar{\varepsilon}(x)} = \int \big[\Delta(x,y) - d''(x)\big]_+ f(y + \tau x)\lambda(dy) \quad (10.2.20)$$

This procedure is justified by the following proposition.

Proposition 10.2.7 **(a)** *The functions $r_{\varepsilon\delta}(x,y)$ and $\pi_{\varepsilon\delta}(x,y)$ are product measurable.*

(b) *The pair $(R'_{-\tau,\varepsilon\delta}, R''_{\tau,\varepsilon\delta})$ is least favorable for $\mathcal{U}_{cv,\varepsilon\delta}(-\tau)$ versus $\mathcal{U}_{cv,\varepsilon\delta}(\tau)$.*

(c) *The probabilities $R'_{-\tau,0}, R''_{\tau,0}, R'_{-\tau,\varepsilon\delta}, R''_{\tau,\varepsilon\delta}$ satisfy conditions (10.2.6) and (10.2.7).*

Proof Rieder (1989), Proposition 3.1. ////

Remark 10.2.8 We additionally assume

$$K\left(\frac{\varepsilon(x) + 2\delta(x)}{1 - \varepsilon(x)} < \int \big(f(y + \tau x) - f(y - \tau x)\big)_+ \lambda(dy)\right) > 0 \quad (10.2.21)$$

i.e., $K(d'(x) = d''(x) = 1) < 1$. ////

10.2.3 Minimax Estimator

The minimax robust test $\tilde{\phi}_t$ for $\mathcal{U}_{cv,t}(-\tau)$ versus $\mathcal{U}_{cv,t}(\tau)$ $(t = 0, \varepsilon\delta)$ is of the explicit form

$$\tilde{\phi}_t = (1 - \gamma)\,\mathrm{I}(h_t > a) + \gamma\,\mathrm{I}(h_t \ge a) \quad (10.2.22)$$

where $2\tau h_t$ denotes the loglikelihood of $(R'_{-\tau,t})^n$ with respect to $(R''_{\tau,t})^n$,

$$h_t(\mathbf{x}, \mathbf{y}) = \frac{1}{2\tau}\sum_{i=1}^{n} \log \pi_t(x_i, y_i) \quad (10.2.23)$$

$\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ and $\gamma \in [0,1]$ and $a \in \mathbb{R}$ are such that

$$\int \tilde{\phi}_t\, d(R'_{-\tau,t})^n = \int (1 - \tilde{\phi}_t)\, d(R''_{\tau,t})^n = \alpha \quad (10.2.24)$$

The estimator given by

$$\tilde{S}_t(\mathbf{x},\mathbf{y}) = \begin{cases} S'_t(\mathbf{x},\mathbf{y}) & \text{with probability } 1-\gamma \\ S''_t(\mathbf{x},\mathbf{y}) & \text{with probability } \gamma \end{cases} \tag{10.2.25}$$

with

$$S'_t(\mathbf{x},\mathbf{y}) = \sup_{u\in\mathbb{R}}\{h_t(\mathbf{x},\mathbf{y}-\mathbf{x}u) > a\} \tag{10.2.26}$$

$$S''_t(\mathbf{x},\mathbf{y}) = \inf_{u\in\mathbb{R}}\{h_t(\mathbf{x},\mathbf{y}-\mathbf{x}u) < a\} \tag{10.2.27}$$

is minimax by the subsequent theorems.

Theorem 10.2.9 **(a)** *S'_t and S''_t are equivariant and Borel measurable.*
(b) $\text{Risk}_\theta(\tilde{S};t) = \text{Risk}\,(\tilde{S};t) = \alpha$ *for all* $\theta\in\mathbb{R}$.

PROOF Rieder (1989), Theorem 4.1. ////

The minimax estimator $\tilde{S}$ is not only optimal for equivariant estimators but for all, arbitrary, estimators which may be shown by a Hunt-Stein type argument. Moreover, the minimax estimator $\tilde{S}$ is also admissible.

Theorem 10.2.10 **(a)** *Every estimator $S:\mathbb{R}^{2n}\to\mathbb{R}$ has* $\text{Risk}(S;t)\geq\alpha$.
(b) *Let $S:\mathbb{R}^{2n}\to\mathbb{R}$ be some estimator such that* $\text{Risk}_\theta(S)<\alpha$ *for some* $\theta\in\mathbb{R}$. *Then, exists some* $\theta_1\in\mathbb{R}$ *such that* $\text{Risk}_{\theta_1}(S)>\alpha$.

PROOF
(a) Rieder (1989), Theorem 4.2.
(b) Assume there exist some estimator $S:\mathbb{R}^{2n}\to\mathbb{R}$ and some $\theta\in\mathbb{R}$ such that

$$\alpha > \text{Risk}_\theta(S) = \sup_{Q_\theta\in\mathcal{U}_{cv,t}(\theta)}\left\{Q^n_\theta(S>\theta+\tau)\vee Q^n_\theta(S<\theta-\tau)\right\} \tag{10.2.28}$$

$$\overset{(10.1.12)}{=} \max\left\{\sup_{Q_{-\tau}\in\mathcal{U}_{cv,t}(-\tau)}\left[g_{\theta+\tau}(Q_{-\tau})^n\right](S>\theta+\tau),\right.$$
$$\left.\sup_{Q_\tau\in\mathcal{U}_{cv,t}(\tau)}\left[g_{\theta-\tau}(Q_\tau)^n\right](S<\theta-\tau)\right\} \tag{10.2.29}$$

$$\overset{(10.2.4)}{=} \max\left\{\sup_{Q_{-\tau}\in\mathcal{U}_{cv,t}(-\tau)}Q^n_{-\tau}(S\circ G_{\theta+\tau}>\theta+\tau),\right.$$
$$\left.\sup_{Q_\tau\in\mathcal{U}_{cv,t}(\tau)}Q^n_\tau(S\circ G_{\theta-\tau}<\theta-\tau)\right\} \tag{10.2.30}$$

$$\geq (R'_{-\tau,t})^n(S\circ G_{\theta+\tau}>\theta+\tau)\vee(R''_{\tau,t})^n(S\circ G_{\theta-\tau}<\theta-\tau) \tag{10.2.31}$$

where $(R'_{-\tau,t},R''_{\tau,t})$ is a least favorable pair for $\mathcal{U}_{cv,t}(-\tau)$ versus $\mathcal{U}_{cv,t}(\tau)$. Since f is continuous in θ by assumption (10.1.4) and also d' and d'' are continuous in θ by translation invariance of Lebesgue measure, we obtain by the Lemma of Scheffé that $\theta\mapsto g_{\theta+\tau}(R'_{-\tau,t})$ and $\theta\mapsto g_{\theta-\tau}(R''_{\tau,t})$ are continuous in total variation. That is, if

$$(R'_{-\tau,t})^n(S\circ G_{\theta+\tau}=\theta+\tau)\neq 0 \qquad \text{and} \qquad (R''_{\tau,t})^n(S\circ G_{\theta-\tau}=\theta-\tau)\neq 0 \tag{10.2.32}$$

there exists some θ' close to θ such that

$$(R'_{-\tau,t})^n(S \circ G_{\theta'+\tau} = \theta' + \tau) = 0 \qquad \text{resp.} \qquad (R''_{\tau,t})^n(S \circ G_{\theta'-\tau} = \theta' - \tau) = 0 \tag{10.2.33}$$

and

$$(R'_{-\tau,t})^n(S \circ G_{\theta'+\tau} > \theta' + \tau) < \alpha \qquad \text{resp.} \qquad (R''_{\tau,t})^n(S \circ G_{\theta'-\tau} < \theta' - \tau) < \alpha \tag{10.2.34}$$

Thus, we may assume some $\theta' > \theta$ without loss of generality such that

$$(R'_{-\tau,t})^n(S \circ G_{\theta'+\tau} = \theta' + \tau) = 0 \tag{10.2.35}$$

and

$$(R'_{-\tau,t})^n(S \circ G_{\theta'+\tau} > \theta' + \tau) \leq (R'_{-\tau,t})^n(S \circ G_{\theta+\tau} > \theta + \tau) < \alpha \tag{10.2.36}$$

Defining $\theta_1 := \theta' + \tau$ and $\phi := \mathrm{I}_{\{S \circ G_{\theta_1} > \theta_1\}}$, we get $\int \phi \, d(R'_{-\tau,t})^n < \alpha$. Furthermore, let $\tilde{\phi}$ be the Neyman-Pearson test for $(R'_{-\tau,t})^n$ versus $(R''_{\tau,t})^n$ such that $\int \tilde{\phi} \, d(R'_{-\tau,t})^n = \int (1 - \tilde{\phi}) \, d(R''_{\tau,t})^n = \alpha$. Then, $\int \phi \, d(R'_{-\tau,t})^n < \int \tilde{\phi} \, d(R'_{-\tau,t})^n = \alpha$ and $\int \phi \, d(R''_{\tau,t})^n < \int \tilde{\phi} \, d(R''_{\tau,t})^n = 1 - \alpha$; i.e., $(R''_{\tau,t})^n(S \circ G_{\theta_1} < \theta_1) > \alpha$ and consequentially, $\mathrm{Risk}_{\theta_1}(S) > \alpha$. ////

A closer look at the proof of Theorem 4.2 of Rieder (1989) shows, we need not take the supremum over the whole real line to prove that every estimator S has $\mathrm{Risk}\,(S) \geq \alpha$.

Proposition 10.2.11 Let $\theta_0 \in \mathbb{R}$ and $\delta > 0$ be fixed. Then, it holds for every estimator $S : \mathbb{R}^{2n} \to \mathbb{R}$

$$\sup_{|\theta - \theta_0| < \tau + \delta} \mathrm{Risk}_\theta(S) \geq \alpha \tag{10.2.37}$$

respectively

$$\sup_{\{|\theta_1 - \theta_0| < \delta\}} \sup_{\{\theta_2 = \theta_1 \pm \tau\}} \mathrm{Risk}_{\theta_2}(S) \geq \alpha \tag{10.2.38}$$

Proof Similar to the proof of Theorem 4.2 of Rieder (1989). ////

10.2.4 Limiting Estimates

10.2.4.1 Unconditional neighborhoods

We now additionally assume that the ideal error law F is symmetric around zero. Thus, the minimax robust test $\tilde{\phi}_0$ for $\mathcal{U}_{cv,0}(-\tau)$ versus $\mathcal{U}_{cv,0}(\tau)$ is of the explicit form

$$\tilde{\phi}_0 = \tfrac{1}{2}\,\mathrm{I}(h_0 > 0) + \tfrac{1}{2}\,\mathrm{I}(h_0 \geq 0) \tag{10.2.39}$$

with

$$h_0(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \Big[(-b) \vee \frac{1}{2\tau} \log \Delta(x_i, y_i) \wedge b\Big] \tag{10.2.40}$$

where $b = \frac{1}{2\tau} \log d''$ and the minimax estimator $\tilde{S}_0$ is an M estimator satisfying

$$\sum_{i=1}^{n} \Big[(-b) \vee \frac{1}{2\tau} \log \Delta\big(x_i, y_i - \tilde{S}_0 x_i\big) \wedge b \Big] = 0 \tag{10.2.41}$$

with equal randomization between the smallest and the largest solutions.
Now, in the limiting case we have equality in condition (10.2.15); i.e., $b = 0$. Thus, by dividing (10.2.41) by b and formally letting b tend to zero, the limiting estimate $\bar{S}_0$ is an M estimator satisfying

$$\sum_{i=1}^{n} \operatorname{sign} \Big[\frac{1}{2\tau} \log \Delta\big(x_i, y_i - \bar{S}_0 x_i\big) \Big] = 0 \tag{10.2.42}$$

with equal randomization between the smallest and the largest solutions. Moreover, we have

$$\operatorname{sign} \Big[\frac{1}{2\tau} \log \Delta(x, y) \Big] = \operatorname{sign}(x) \operatorname{sign} \Big[\frac{\log f(y - \tau x) - \log f(y + \tau x)}{2\tau x} \Big] \tag{10.2.43}$$

where $\frac{\log f(y-\tau x) - \log f(y+\tau x)}{2\tau x}$ is increasing in y for fixed x and τ by assumption (10.1.4), which is $\log f$ concave. In addition, we get $\Delta(x, -y) = \Delta(x, y)^{-1}$; i.e., $\log \Delta(x, -y) = -\log \Delta(x, y)$. Therefore, we can simplify (10.2.42) to

$$0 = \sum_{i=1}^{n} \operatorname{sign}(x_i) \operatorname{sign}(y_i - \bar{S}_0 x_i) = \sum_{i=1}^{n} \operatorname{sign}(y_i / x_i - \bar{S}_0) \tag{10.2.44}$$

That is, $\bar{S}_0$ is the sample median of the slopes with equal randomization between the minimal and maximal median.

Remark 10.2.12 **(a)** We derive the limiting estimate by formally letting b tend to zero like in Huber (1981), p 97. But, actually one additionally has to compare the limiting risk of the corresponding finite-sample minimax estimator with the risk of the limiting estimate. In case of the finite-sample risk (10.2.5) this seems to be very difficult or even impossible analytically. For the setup of Huber (1964) this is done analytically in Section 3.2 of Rieder et al. (2001).

(b) For non-symmetric F we get some constant c instead of $(-b)$ in (10.2.40) and it is not clear, in general, what limiting value c/b attains. Consequentially, we do not specify the corresponding limiting estimate in general. ////

10.2.4.2 Conditional neighborhoods

If we additionally assume F symmetric, we obtain $\tilde{\phi}_{\varepsilon\delta}$, $h_{\varepsilon\delta}$ and $\tilde{S}_{\varepsilon\delta}$, respectively, analogously to (10.2.39)–(10.2.41) where we have to replace b by $b(x) = \frac{1}{2\tau} \log d''(x)$. However, to derive the limiting estimate we need to know more about the dependence of the loglikelihood $\log \Delta(x, y)$ and the clipping function $b(x)$ from x which is not clear in general. For instance, if $F = \mathcal{N}(0, 1)$ and $b(x) = b/|x|$ (Hampel-Krasker estimator) we obtain the sample median of the slopes. Whereas, if we have $F = \mathcal{N}(0, 1)$ and $b(x) \equiv b$ (Huber estimator), this leads to the minimum L_1-estimator; confer Section 12.2.

10.3 Asymptotic Minimax Regression Estimator

In this section we apply the results of Rieder (1980) to simple linear regression. We first introduce the infinitesimal robust setup and the corresponding asymptotic over-/undershooting risk of an asymptotically linear estimator; confer Subsections 10.3.1 and 10.3.2. We then derive the asymptotic minimax estimator without any connection to the corresponding robust testing problem of Rieder (1978) by using optimization arguments as in Ruckdeschel and Rieder (2004); confer Subsection 10.3.3.

10.3.1 Infinitesimal Robust Setup

We consider infinitesimal neighborhoods that are shrinking at a rate of $\sqrt{n}$; i.e., we replace ε and δ by $\varepsilon/\sqrt{n}$ and $\delta/\sqrt{n}$ $(t=0,\, \varepsilon\delta)$, respectively in Definition 10.1.3, where $\varepsilon, \delta \in [0,\infty)$ $(t=0)$ and $\varepsilon, \delta\colon \mathbb{R} \to [0,\infty)$ $(t=\varepsilon\delta)$, respectively. As already noted in Section 10.1 the simple linear regression model (10.1.1) and the neighborhoods $\mathcal{U}_{cv,0}(\theta)$ and $\mathcal{U}_{cv,\varepsilon\delta}(\theta)$ are invariant under the group of regression transforms (10.1.11). Moreover, there is also an invariance of the set of ICs under this transformation, which also holds for the optimal ICs derived under P_θ and P_0, respectively; confer Subsection 7.2.3 of Rieder (1994) and Subsubsection 7.1.4. That is, we may restrict $\theta \in \mathbb{R}$ to $\theta = 0$ and omit it from notation.

As submodel of the shrinking neighborhoods we employ simple perturbations which in case of the unconditional contamination/total variation neighborhoods are of the form (cf. also Rieder (1994), p 176)

$$dQ_{0,n} = \left(1 + \frac{1}{\sqrt{n}}(\varepsilon q_1 + \delta q_2)\right) dP \qquad q_1 \in \mathcal{G}_{c,0},\, q_2 \in \mathcal{G}_{v,0} \tag{10.3.1}$$

where the classes $\mathcal{G}_{*,0}$ $(*=c,v)$ of one-dimensional bounded tangents are defined as in (1.2.10) and (1.2.11); i.e.,

$$\mathcal{G}_{c,0} = \left\{ q \in Z_\infty \,\middle|\, \inf_P q \geq -1 \right\} \tag{10.3.2}$$

and

$$\mathcal{G}_{v,0} = \left\{ q \in Z_\infty \,\middle|\, \mathrm{E}\,|q| \leq 2 \right\} \tag{10.3.3}$$

with

$$Z_\infty = \left\{ q \in L_\infty(P) \,\middle|\, \mathrm{E}\, q = 0 \right\} \tag{10.3.4}$$

and E denotes expectation under P.

Analogously, we obtain the simple perturbations for the conditional contamination/total variation neighborhoods

$$Q_{\varepsilon,n}(dy \,|\, x) = \left(1 + \frac{1}{\sqrt{n}}\left(\varepsilon(x) q_1(y,x) + \delta(x) q_2(y,x)\right)\right) F(dy) \qquad \text{a.e. } K(dx) \tag{10.3.5}$$

where $q_1 \in \mathcal{G}_{c,\varepsilon}$, $q_2 \in \mathcal{G}_{v,\delta}$ and the classes $\mathcal{G}_{*,t}$ $(* = c, v; t = \varepsilon, \delta)$ of one-dimensional, conditionally centered, and bounded tangents are defined as

$$\mathcal{G}_{c,\varepsilon} = \left\{ q \in Z_{\infty\bullet} \,\middle|\, \inf_\bullet q \geq -1 \right\} \tag{10.3.6}$$

and

$$\mathcal{G}_{v,\delta} = \left\{ q \in Z_{\infty\bullet} \,\middle|\, \mathrm{E}_\bullet |q| \leq 2 \right\} \tag{10.3.7}$$

with

$$Z_{\infty\bullet} = \left\{ q \in L_\infty(P) \,\middle|\, \mathrm{E}_\bullet q = 0 \right\} \tag{10.3.8}$$

and $\mathrm{E}_\bullet$ denotes expectation under F.

Remark 10.3.1 In order to apply the results for conditional neighborhoods contained in Chapter 7 of Rieder (1994), we have to identify ε and δ with $r\varepsilon$ and $r\delta$ $(r \in [0, \infty))$, respectively, where our $t = \varepsilon\delta$ corresponds to his $t = \varepsilon$. ////

The following Lemma is the analogue to Lemmata 5.3.1 and 7.2.7 of Rieder (1994).

Lemma 10.3.2 **(a)** *Given $q_1, q_2 \in Z_\infty$ and $\varepsilon, \delta \in (0, \infty)$. Then, for every $n \in \mathbb{N}$ such that $\sqrt{n} \geq -\varepsilon \inf_P q_1 - \delta \inf_P q_2$,*

$$Q_{0,n}(q_1, q_2, \varepsilon, \delta) \in \mathcal{U}_{cv,0}(P) \quad \Longleftrightarrow \quad q_1 \in \mathcal{G}_{c,0},\ q_2 \in \mathcal{G}_{v,0} \tag{10.3.9}$$

(b) *Given $q_1, q_2 \in Z_{\infty\bullet}$ and $\varepsilon, \delta \colon \mathbb{R} \to (0, \infty)$. Then, for every $n \in \mathbb{N}$ such that $\sqrt{n} \geq -\inf_P \varepsilon q_1 - \inf_P \delta q_2$,*

$$Q_{\varepsilon,n}(q_1, q_2, \varepsilon, \delta) \in \mathcal{U}_{cv,\varepsilon\delta}(P) \quad \Longleftrightarrow \quad q_1 \in \mathcal{G}_{c,\varepsilon},\ q_2 \in \mathcal{G}_{v,\delta} \tag{10.3.10}$$

Proof Analogously to Rieder (1994), Lemma 5.3.1. ////

These tangent classes generate the following (infinitesimal) bias terms

$$\omega_0(\eta, \varepsilon, \delta) = \varepsilon \sup \left\{ |\mathrm{E}\, \eta q| \,\middle|\, q \in \mathcal{G}_{c,0} \right\} + \delta \sup \left\{ |\mathrm{E}\, \eta q| \,\middle|\, q \in \mathcal{G}_{v,0} \right\} \tag{10.3.11}$$

and

$$\omega_{\varepsilon\delta}(\eta, \varepsilon, \delta) = \sup \left\{ |\mathrm{E}\, \varepsilon\eta q| \,\middle|\, q \in \mathcal{G}_{c,\varepsilon} \right\} + \sup \left\{ |\mathrm{E}\, \delta\eta q| \,\middle|\, q \in \mathcal{G}_{v,\delta} \right\} \tag{10.3.12}$$

for every $\eta \in L_1(P)$. These terms are positively homogeneous, subadditive, hence convex on $L_1(P)$, and weakly lower semicontinuous on $L_2(P)$ by Lemma 5.3.2 and Lemma 7.3.1 of Rieder (1994). In view of Proposition 5.3.3 and Proposition 7.3.3 of Rieder (1994) we obtain the following explicit terms

$$\omega_{cv,0}(\eta, \varepsilon, \delta) = \varepsilon \sup_P |\eta| + \delta \left(\sup_P \eta - \inf_P \eta \right) \tag{10.3.13}$$

$$= \left[(\varepsilon + \delta) \sup_P \eta - \delta \inf_P \eta \right] \vee \left[\delta \sup_P \eta - (\varepsilon + \delta) \inf_P \eta \right] \tag{10.3.14}$$

and

$$\omega_{cv,\varepsilon\delta}(\eta, \varepsilon, \delta) = \mathrm{E} \left[\varepsilon \sup_\bullet |\eta| + \delta \left(\sup_\bullet \eta - \inf_\bullet \eta \right) \right] \tag{10.3.15}$$

$$= \mathrm{E} \left\{ \left[(\varepsilon + \delta) \sup_\bullet \eta - \delta \inf_\bullet \eta \right] \vee \left[\delta \sup_\bullet \eta - (\varepsilon + \delta) \inf_\bullet \eta \right] \right\} \tag{10.3.16}$$

where $\sup_P$ and $\sup_\bullet$ denote the essential suprema with respect to P and F, respectively.

10.3.2 Risk of an Asymptotic Estimator

This subsection is based on Section 2 of Rieder (1980). We consider only asymptotic estimators $S_n\colon \mathbb{R}^{2n} \to \mathbb{R}$ that are asymptotically linear at P with IC $\eta \in \Psi_2$ (cf. Definition 1.1.5), respectively $\eta \in \Psi_{2\bullet}$ with

$$\Psi_{2\bullet}(\theta) = \big\{\psi_\theta \in L_2(P_\theta) \mid \mathrm{E}_\bullet\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = 1\big\} \tag{10.3.17}$$

Moreover, given $\tau \in (0,\infty)$ we consider $\tau', \tau'' \in \mathbb{R}$ such that $\tau' + \tau'' = 2\tau$. We then lay the possibly asymmetric confidence intervals $\big[S_n - \tau'/\sqrt{n}\,,\ S_n + \tau''/\sqrt{n}\,\big]$ of given width $2\tau_n = 2\tau/\sqrt{n}$ around S_n. The following remark quotes pp 108 of Rieder (1980).

Remark 10.3.3 It is only reasonable to let the width shrink at an appropriate rate since with increasing n even the contaminated sample provides more and more information to estimate θ. [...] The partition of the width is to compensate possible asymmetries of S_n, which, in general, render the use of S_n as the strict midpoint unoptimal, in the sense defined subsequently. ////

Given ε, δ and simple perturbations $Q_{t,n}(q_1, q_2, \varepsilon, \delta)$ of form (10.3.1) and (10.3.5), respectively, we obtain analogously to Proposition 1.2.5

$$\sqrt{n}\, S_n\big(Q_{t,n}^n(q_1, q_2, \varepsilon, \delta)\big) \xrightarrow{\mathrm{w}} \mathcal{N}\big(\mathrm{E}(\varepsilon q_1 + \delta q_2)\eta,\ \mathrm{E}\,\eta^2\big) \tag{10.3.18}$$

According to the finite-sample risk given in Definition 10.2.1, we are interested in the maximum asymptotic under-/overshooting probability of $S = S_n$; i.e.,

$$\begin{aligned}
\mathrm{Risk}\,(S, \tau', \tau''; t) \quad &:= \quad \sup_{q_1 \in \mathcal{G}_{c,t},\, q_2 \in \mathcal{G}_{v,t}} \Big\{ \lim_{n\to\infty} Q_{t,n}^n(q_1, q_2, \varepsilon, \delta)(S_n < -\tau_n'/\sqrt{n}\,) \\
&\qquad\qquad \vee \lim_{n\to\infty} Q_{t,n}^n(q_1, q_2, \varepsilon, \delta)(S_n > \tau_n''/\sqrt{n}\,) \Big\} \qquad (10.3.19) \\
&\overset{(10.3.18)}{=} \sup_{q_1 \in \mathcal{G}_{c,t},\, q_2 \in \mathcal{G}_{v,t}} \Big\{ \Phi\Big(\frac{\mathrm{E}(\varepsilon q_1 + \delta q_2)\eta - \tau'}{\sqrt{\mathrm{E}\,\eta^2}}\Big) \\
&\qquad\qquad \vee \Phi\Big(\frac{\mathrm{E}(\varepsilon q_1 + \delta q_2)(-\eta) - \tau''}{\sqrt{\mathrm{E}\,\eta^2}}\Big) \qquad (10.3.20)
\end{aligned}$$

In view of (10.3.13)–(10.3.16) this leads us to the following limiting risk

$$\mathrm{Risk}\,(S, \tau', \tau''; t) \ = \ \Phi\Big(\frac{\big[s_t'(\eta) - \tau'\big] \vee \big[s_t''(\eta) - \tau''\big]}{\sqrt{\mathrm{E}\,\eta^2}}\Big) \qquad (t = 0,\ \varepsilon\delta) \tag{10.3.21}$$

with

$$s_0'(\eta) := (\varepsilon + \delta)\sup_P \eta - \delta \inf_P \eta \tag{10.3.22}$$

$$s_{\varepsilon\delta}'(\eta) := \mathrm{E}\,\big[(\varepsilon + \delta)\sup_\bullet \eta - \delta \inf_\bullet \eta\big] \tag{10.3.23}$$

and

$$s_0''(\eta) := \delta \sup_P \eta - (\varepsilon + \delta)\inf_P \eta \tag{10.3.24}$$

$$s_{\varepsilon\delta}''(\eta) := \mathrm{E}\,\big[\delta \sup_\bullet \eta - (\varepsilon + \delta)\inf_\bullet \eta\big] \tag{10.3.25}$$

Furthermore, we choose τ', τ'' in favor of the estimate and obtain the following risk of S.

Definition 10.3.4 *Given $\tau \in (0, \infty)$, the risk of an asymptotic estimator $S = S_n \colon \mathbb{R}^{2n} \to \mathbb{R}$ that is asymptotically linear at P with IC $\eta \in \Psi_2$, respectively $\eta \in \Psi_{2\bullet}$ is*

$$\operatorname{Risk}(S; t) = \inf \left\{ \operatorname{Risk}(S, \tau', \tau''; t) \,\middle|\, \tau', \tau'' \in \mathbb{R} \colon \tau' + \tau'' = \tau \right\} \qquad (t = 0,\ \varepsilon\delta) \tag{10.3.26}$$

As $\operatorname{Risk}(S, \tau', \tau''; t)$ is minimized by the following choice of τ', τ''

$$\tau'(\eta) = \tau + \tfrac{1}{2}\left(s_t'(\eta) - s_t''(\eta)\right) \tag{10.3.27}$$

$$\tau''(\eta) = \tau - \tfrac{1}{2}\left(s_t'(\eta) - s_t''(\eta)\right) \tag{10.3.28}$$

we get analogously to Theorem 2.2 of Rieder (1980)

$$\operatorname{Risk}(S; t) = \Phi\left(\frac{\frac{1}{2}\left[s_t'(\eta) + s_t''(\eta)\right] - \tau}{\sqrt{\mathrm{E}\,\eta^2}} \right) \tag{10.3.29}$$

10.3.3 Minimax Estimator

In view of (10.3.29) and by the strict monotony of Φ we have to solve

$$\frac{s_t'(\eta) + s_t''(\eta) - 2\tau}{\sqrt{\mathrm{E}\,\eta^2}} = \min! \qquad (t = 0,\ \varepsilon\delta) \tag{10.3.30}$$

for given $\tau \in (0, \infty)$ in order to obtain the IC of the asymptotic minimax estimator, where $\eta \in \Psi_2$ or $\eta \in \Psi_{2\bullet}$, respectively. To make sure that $\operatorname{Risk}(S; t) < 50\%$ which we obtain if we randomize between $\pm\infty$ with equal probability, without looking at the observations, we only consider ε, δ such that

$$s_t'(\eta) + s_t''(\eta) < 2\tau \tag{10.3.31}$$

In case of unconditional regression neighborhoods ($t = 0$) the solution to this problem subject to the boundedness condition (10.3.39) was derived by Rieder (1980) using results of robust testing theory provided by Rieder (1978). The following lemma, whose proof is based on a simple direct argument provided by H. Rieder, provides a more general way to solve this problem in case of unconditional as well as conditional neighborhoods.

Lemma 10.3.5 *Problem (10.3.30) is equivalent to the Hampel type problem*

$$\mathrm{E}\,\eta^2 = \min! \qquad s_t'(\eta) + s_t''(\eta) \le s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) < 2\tau \qquad (t = 0,\ \varepsilon\delta) \tag{10.3.32}$$

with $\eta \in \Psi_2$, respectively $\eta \in \Psi_{2\bullet}$.

PROOF Assume $\tilde{\eta}$ is a solution to problem (10.3.30); i.e.,

$$\frac{s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) - 2\tau}{\sqrt{\mathrm{E}\,\tilde{\eta}^2}} \leq \frac{s_t'(\eta) + s_t''(\eta) - 2\tau}{\sqrt{\mathrm{E}\,\eta^2}} \tag{10.3.33}$$

for all $\eta \in \Psi_2$, respectively $\eta \in \Psi_{2\bullet}$. This is equivalent to

$$\big[s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) - 2\tau\big]\sqrt{\mathrm{E}\,\eta^2} \leq \big[s_t'(\eta) + s_t''(\eta) - 2\tau\big]\sqrt{\mathrm{E}\,\tilde{\eta}^2} \tag{10.3.34}$$

Now, let $s_t'(\eta) + s_t''(\eta) \leq s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) < 2\tau$, then

$$\big[s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) - 2\tau\big]\sqrt{\mathrm{E}\,\eta^2} \leq \big[s_t'(\tilde{\eta}) + s_t''(\tilde{\eta}) - 2\tau\big]\sqrt{\mathrm{E}\,\tilde{\eta}^2} \tag{10.3.35}$$

and therefore

$$\sqrt{\mathrm{E}\,\eta^2} \geq \sqrt{\mathrm{E}\,\tilde{\eta}^2} \tag{10.3.36}$$

That is, any solution $\tilde{\eta}$ to problem (10.3.30) also solves the Hampel-type problem (10.3.32). ////

Remark 10.3.6 **(a)** Various aspects of this Hampel type problem (minimum norm problem, well-posedness, ...) can be found in Rieder (1994), pp 196 ($t = 0$) and pp 274 ($t = \varepsilon\delta$), respectively.

(b) Ruckdeschel and Rieder (2004) introduce a more general way to solve problems of the form

$$G(\text{bias}, \text{variance}) = \min! \tag{10.3.37}$$

where G is some suitable function; confer Remark 1.3.3 (b). In particular, they derive the solution to problem (10.3.30) in case of an one-dimensional parametric family and infinitesimal contamination, respectively total variation neighborhoods. ////

Now, we can apply the results contained in Chapters 5 and 7 of Rieder (1994), with some obvious identifications, to specify the unique solution to problem (10.3.32), respectively problem (10.3.30). We first give the solution for unconditional regression neighborhoods ($t = 0$).

Theorem 10.3.7 *Let $\tau \in (0, \infty)$ be fixed.*

(a) *There exist some $b \in (0, \infty)$, $c \in (-b, 0)$ and $A \in (0, \infty)$ such that the solution is given by*

$$\tilde{\eta} = A\big[c \vee \Lambda \wedge (c + b)\big] \tag{10.3.38}$$

if

$$\mathrm{E}\,(\Lambda)_+ > \frac{\varepsilon + 2\delta}{2\tau} \tag{10.3.39}$$

where

$$\mathrm{E}(c - \Lambda)_+ = \mathrm{E}(\Lambda - (c + b))_+ = \frac{\varepsilon + 2\delta}{2\tau} \tag{10.3.40}$$

and

$$A^{-1} = \mathrm{E}\,\big[c \vee \Lambda \wedge (c + b)\big]\Lambda \tag{10.3.41}$$

Conversely, for any $A \in (0,\infty)$, *some* $b \in (0,\infty)$ *and* $c \in (-b,0)$ *are defined by (10.3.40) if (10.3.39) holds. Moreover, if some* $\tilde{\eta}$ *is of form (10.3.38)–(10.3.41), then* $\tilde{\eta} \in \Psi_2$ *and* $\tilde{\eta}$ *is the solution to problem (10.3.30).*

(b) *If* $\mathrm{E}\,(\Lambda)_+ = \frac{\varepsilon+2\delta}{2\tau}$, *the solution is*

$$\bar{\eta} = \frac{2\tau}{\varepsilon+2\delta}\left(\frac{P(\Lambda<0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda>0) - \frac{P(\Lambda>0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda<0)\right) \tag{10.3.42}$$

Proof

(a) We have

$$s_0'(\eta) + s_0''(\eta) = (\varepsilon+2\delta)(\sup_P \eta - \inf_P \eta) \tag{10.3.43}$$

Thus, the form (10.3.38) of the solution to (10.3.32) may be read off from Theorem 5.5.5 (a) of Rieder (1994) by identifying $(\varepsilon+2\delta)$ with $r \in (0,\infty)$ and $(\sup_P \eta - \inf_P \eta)$ with $\omega_v(\eta)$. Writing out $\mathrm{E}(\tilde{\eta} - A\Lambda) = 0$, we get

$$\mathrm{E}(c-\Lambda)_+ = \mathrm{E}\,\big(\Lambda - (c+b)\big)_+ = \gamma \tag{10.3.44}$$

where $\gamma \in [0,\infty)$ is yet to determine. We have

$$1 = \mathrm{E}\,\eta\Lambda = \tfrac{1}{A}\big[\mathrm{E}\,\eta\tilde{\eta} + \mathrm{E}\,\eta(A\Lambda - \tilde{\eta})\big] \le \tfrac{1}{A}\,\mathrm{E}\,\eta\tilde{\eta} + \tfrac{\gamma}{\varepsilon+2\delta}\big[s_0'(\eta) + s_0''(\eta)\big] \tag{10.3.45}$$

with equality if $\eta = \tilde{\eta}$. This implies

$$\frac{s_0'(\tilde{\eta}) + s_0''(\tilde{\eta}) - 2\tau}{\sqrt{\mathrm{E}\,\tilde{\eta}^2}} = \frac{\big[1 - \frac{2\tau}{\varepsilon+2\delta}\gamma\big]\big[s_0'(\tilde{\eta}) + s_0''(\tilde{\eta})\big] - 2\frac{\tau}{A}\,\mathrm{E}\,\tilde{\eta}^2}{\sqrt{\mathrm{E}\,\tilde{\eta}^2}} \tag{10.3.46}$$

$$\le \frac{\big[1 - \frac{2\tau}{\varepsilon+2\delta}\gamma\big]\big[s_0'(\eta) + s_0''(\eta)\big] - 2\frac{\tau}{A}\,\mathrm{E}\,\eta\tilde{\eta}}{\sqrt{\mathrm{E}\,\eta^2}} \tag{10.3.47}$$

We now choose $\gamma = \frac{\varepsilon+2\delta}{2\tau}$ if $\mathrm{E}(\Lambda)_+ > \frac{\varepsilon+2\delta}{2\tau}$. Since $\tau > 0$ and $A > 0$ this leads to $\mathrm{E}\,\eta\tilde{\eta} \le \sqrt{\mathrm{E}\,\eta}\,\sqrt{\mathrm{E}\,\tilde{\eta}}$ which holds by the Cauchy-Schwarz inequality. Thus, $\tilde{\eta}$ is indeed the solution to problem (10.3.30).

Since (10.3.40) entails $\mathrm{E}\,\tilde{\eta} = 0$ for $\tilde{\eta}$ of form (10.3.38)–(10.3.41), the converse follows.

(b) Immediately follows from Theorem 5.5.5 (b) of Rieder (1994). ////

Remark 10.3.8 The solution given in the previous theorem coincides with the solution for any arbitrary one-dimensional L_2 differentiable parametric family which additionally does not have to have a certain invariance structure. We just have to identify Λ with the L_2 derivative — at some parameter value $\theta \in \Theta$ — of the corresponding parametric family. ////

We now specify the corresponding solutions in case of conditional regression neighborhoods ($t = \varepsilon\delta$).

Theorem 10.3.9 *Let* $\tau \in (0,\infty)$ *be fixed and assume* $\mathrm{E}(\varepsilon + 2\delta) < \infty$.

(a) *There exist some Borel measurable functions* $c' : \mathbb{R} \to (-\infty, 0)$ *and* $c'' : \mathbb{R} \to (0,\infty)$ *and some constant* $A \in (0,\infty)$ *such that the solution is given by*

$$\tilde{\eta} = A\big[c' \vee \Lambda \wedge c''\big] \tag{10.3.48}$$

if

$$K\big(|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ > \gamma\big) > 0 \qquad \gamma := \frac{\varepsilon + 2\delta}{2\tau} \tag{10.3.49}$$

where

$$c' = c'' = 0 \qquad \text{if} \quad |x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ \le \gamma \tag{10.3.50}$$

$$\mathrm{E}_\bullet(c' - \Lambda)_+ = \mathrm{E}_\bullet(\Lambda - c'')_+ = \gamma \qquad \text{if} \quad |x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ > \gamma \tag{10.3.51}$$

and

$$A^{-1} = \mathrm{E}\left[c' \vee \Lambda \wedge c''\right]\Lambda \tag{10.3.52}$$

Conversely, for any $A \in (0,\infty)$, *some Borel measurable functions* $c' : \mathbb{R} \to (-\infty, 0)$ *and* $c'' : \mathbb{R} \to (0,\infty)$ *are defined by (10.3.50) and (10.3.51) if (10.3.49) holds. Moreover, if some* $\tilde{\eta}$ *is of form (10.3.48)–(10.3.52), then* $\tilde{\eta} \in \Psi_{2\bullet}$ *and* $\tilde{\eta}$ *is the solution to problem (10.3.30).*

(b) *If* $K\big(|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ > \frac{\varepsilon+2\delta}{2\tau}\big) = 0$ *and* $K\big(|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ = \frac{\varepsilon+2\delta}{2\tau}\big) > 0$. *Then, under the assumption (10.1.5), the solution is*

$$\bar{\eta} = A|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ \left(\frac{\mathrm{I}(x\Lambda_f^{\mathrm{loc}} > 0)}{F(x\Lambda_f^{\mathrm{loc}} > 0)} - \frac{\mathrm{I}(x\Lambda_f^{\mathrm{loc}} < 0)}{F(x\Lambda_f^{\mathrm{loc}} < 0)}\right) \mathrm{I}\left(|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ = \tfrac{\varepsilon+2\delta}{2\tau}\right) \tag{10.3.53}$$

with the scalar $A \in (0,\infty)$ *determined by* $\mathrm{E}\,\bar{\eta}\Lambda = 1$.

Proof

(a) We have

$$s'_{\varepsilon\delta}(\eta) + s''_{\varepsilon\delta}(\eta) = \mathrm{E}(\varepsilon + 2\delta)(\sup_\bullet \eta - \inf_\bullet \eta) \tag{10.3.54}$$

To apply the results of Chapter 7 of Rieder (1994), we have to identify $(\varepsilon + 2\delta)$ with $r\varepsilon$ and $\mathrm{E}(\varepsilon + 2\delta)(\sup_\bullet \eta - \inf_\bullet \eta)$ with $r\omega_{v,\varepsilon}(\eta)$. The form (10.3.48) of the solution to (10.3.32) may then be read off from Theorem 7.4.4 (a), respectively Theorem 7.4.16 (a) of Rieder (1994) with $c' : \mathbb{R} \to (-\infty, 0)$, $c'' : \mathbb{R} \to (0,\infty)$ and $A \in (0,\infty)$. Writing out $\mathrm{E}_\bullet(\tilde{\eta} - A\Lambda) = 0$, we get

$$\mathrm{E}_\bullet(c' - \Lambda)_+ = \mathrm{E}_\bullet(\Lambda - c'')_+ = \gamma \tag{10.3.55}$$

Arguing as in part (a) of the proof of Theorem 10.3.7 for each x section yields

$$\gamma(x) = \frac{\varepsilon(x) + 2\delta(x)}{2\tau} \qquad \text{if} \quad |x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ > \frac{\varepsilon(x) + 2\delta(x)}{2\tau} \tag{10.3.56}$$

If $|x|\,\mathrm{E}_\bullet\left(\Lambda_f^{\mathrm{loc}}\right)_+ \le \gamma(x)$, we choose $c'(x) = c''(x) = 0$ and the corresponding $\tilde{\eta}$ solves problem (10.3.30). However, we have to assume (10.3.49) to make sure that

$\tilde{\eta} \in \Psi_2(\theta)$; i.e., $K(\tilde{\eta} \neq 0) > 0$. The measurability of the functions c', c'' follows analogously to the proof of Theorem 7.4.16 (a) of Rieder (1994).

Since (10.3.50) and (10.3.51) entail $\mathrm{E}_\bullet\, \tilde{\eta} = 0$ for $\tilde{\eta}$ of form (10.3.48)–(10.3.52), the converse follows.

(b) Follows from Theorem 7.4.16 (c) of Rieder (1994) by identifying ε with $\varepsilon + 2\delta$ and σ_1 with $2\tau\, \mathrm{E}_\bullet \left(\Lambda_f^{\mathrm{loc}}\right)_+$. ////

Remark 10.3.10 **(a)** The assumption $\mathrm{E}(\varepsilon+2\delta) < \infty$ guarantees that $s'(\eta)+s''(\eta)$ is finite on $L_\infty(P)$ since bounded ICs exist; confer also Rieder (1994), p 275.

(b) In case of Rieder (1980) the boundedness condition (10.3.39), which is equivalent to the disjointness of the neighborhoods in the corresponding testing problem, is needed beforehand to be able to apply the results of robust testing given by Rieder (1978). Whereas in our case the boundedness condition (10.3.39) ($t = 0$), respectively its analogon (10.3.49) ($t = \varepsilon\delta$) occurs canonically in the derivation of the solution.

(c) By (10.3.45) it holds

$$\frac{1}{2}\left[s_t'(\tilde{\eta}) + s_t''(\tilde{\eta})\right] = \tau - \frac{\tau}{A}\,\mathrm{E}\,\tilde{\eta}^2 \tag{10.3.57}$$

where $t = 0, \varepsilon\delta$. Thus, the asymptotic risk (10.3.29) of the asymptotic minimax estimator $\tilde{S}_t^{\mathrm{as}}$ reads

$$\mathrm{Risk}\,(\tilde{S}_t^{\mathrm{as}}; t) = \Phi\left(-\frac{\tau\sqrt{\mathrm{E}\,\tilde{\eta}^2}}{A}\right) \tag{10.3.58}$$

Since $\tau > 0$ and $A > 0$ this also implies (10.3.31).

(d) By evaluating the under-/overshoot confidence risk (10.3.29) uniformly, superefficient estimators are cut out not be assumption but by a high maximum risk. Doing so, Rieder (1981a) establishes a local asymptotic minimax bound for arbitrary estimators (cf. Theorem 4.1) which equals (10.3.58). Thus, in case of unconditional regression neighborhoods the asymptotic minimax estimator is, in this sense, minimax for all arbitrary estimators. Moreover, Rieder (1981a) proves the local asymptotic admissibility of minimax estimators (cf. Theorem 4.2, ibid.) and shows that these estimators have an asymptotic expansion in terms of a truncated logarithmic derivative (cf. Theorem 4.3, ibid.). However, he conjectures that more general loss functions cannot be treated this way as the corresponding proofs utilize the correspondence between estimators and tests which is only possible by the particular choice of the risk. ////

Based on the optimal IC $\tilde{\eta}$, we can specify an estimator $\tilde{S} = \tilde{S}_n$ that is asymptotically normal not only under simple perturbations but on full total variation balls by means of one-step constructions. For more details we refer to Section 6.4 of Rieder (1994) and Section 2.3.

Proposition 10.3.11 *For all $r \in (0, \infty)$ and all arrays $Q_{n,i} \in B_v(P, r/\sqrt{n}\,)$,*

$$\sqrt{n}\left(\tilde{S}_n - \theta - \frac{1}{n}\sum_{i=1}^{n}\int \tilde{\eta}_{\theta_n}\, dQ_{n,i}\right)\left(Q_n^{(n)}\right) \xrightarrow{\mathrm{w}} \mathcal{N}\left(0, \mathrm{E}\,\tilde{\eta}^2\right) \tag{10.3.59}$$

where $Q_n^{(n)} = Q_{n,1} \otimes \cdots \otimes Q_{n,n}$ *and* $\theta_n = h/\sqrt{n}$ *for some* $h \in \mathbb{R}$.

PROOF Theorem 2.3.1 in combination with Proposition 2.4.1 and Remark 2.3.4 (d). ////

Remark 10.3.12 **(a)** In case of unconditional regression neighborhoods we have for $\varepsilon_n := \varepsilon/\sqrt{n} \in [0,1)$ and $\delta_n := \delta/\sqrt{n} \in [0,1)$ such that $\varepsilon_n + \delta_n \in (0,1)$

$$\mathcal{U}_{cv,0}(0;\varepsilon_n,\delta_n) \subset B_v\big(P,\,(\varepsilon_n+\delta_n)\big) \tag{10.3.60}$$

Thus, $\tilde{S}$ is asymptotically normal for all arrays $Q_{n,i} \in \mathcal{U}_{cv,0}(0;\varepsilon_n,\delta_n)$ where

$$\lim_{n\to\infty} \sup_{Q_{n,i}\in B_v\left(P,\,(\varepsilon_n+\delta_n)\right)} \sqrt{n}\left|\int \tilde{\eta}_{\theta_n}\, dQ_{n,i}\right| \overset{(2.3.74)}{\leq} (\varepsilon+\delta)b \qquad \text{a.e. } P \tag{10.3.61}$$

Consequentially, the maximum asymptotic risk (10.3.29) is identical to the asymptotic maximum risk. Thus, we can regard $\tilde{S}$ as the asymptotic minimax estimator $\tilde{S}_0^{\text{as}}$.

(b) In case of conditional regression neighborhoods we get with $\varepsilon_n(x) := \varepsilon(x)/\sqrt{n}$ and $\delta_n(x) := \delta(x)/\sqrt{n}$ such that $\varepsilon_n(x) + \delta_n(x) \in (0,1)$ a.e. $K(dx)$ analogously to Proposition 7.2.2 of Rieder (1994),

$$\mathcal{U}_{cv,\varepsilon\delta}(0;\varepsilon_n,\delta_n) \subset B_v\big(P,\, r/\sqrt{n}\,\big) \tag{10.3.62}$$

where $\mathrm{E}(\varepsilon+\delta) \leq r \in (0,\infty)$. However, we do not have $\omega_{cv,\varepsilon\delta} = \omega_{v,0}$ in general. Thus, the construction problem in case of conditional regression neighborhoods goes beyond Chapter 6 of Rieder (1994) and we treat it separated from this thesis. ////

Chapter 11

One-Dimensional Normal Location

We specialize the results of Chapter 10 to one-dimensional normal location,

$$y = \theta + u \tag{11.0.1}$$

with y the observed variable and u the statistical error. Given a fixed number of i.i.d. observations $y_1, \ldots, y_n$ $(n \in \mathbb{N})$, the unknown location parameter $\theta \in \mathbb{R}$ has to be estimated. We assume

$$F(du) = \varphi(u)\lambda(du) \tag{11.0.2}$$

where $\varphi(u)$ is the density of $\mathcal{N}(0,1)$. Thus, the ideal model distributions read

$$P_\theta(dy) = \varphi(y-\theta)\lambda(dy) \tag{11.0.3}$$

Moreover, we consider the following contamination/total variation neighborhoods of P_θ,

$$\mathcal{U}_{cv}(\theta, \varepsilon_n, \delta_n) = \left\{ Q \in \mathcal{M}_1(\mathbb{B}) \,\middle|\, Q(dy) \geq (1-\varepsilon_n)P_\theta(dy) - \delta_n \right\} \tag{11.0.4}$$

for given radii $\varepsilon_n, \delta_n \in [0,1]$ where $\varepsilon_n = \varepsilon/\sqrt{n}\,, \delta_n = \delta/\sqrt{n}$ and $n \in \mathbb{N}$ is fixed. Then, the finite-sample and the asymptotic minimax location estimator may be read off from Sections 10.2 and 10.3 by choosing regressor $x \equiv 1$. In Sections 11.1 and 11.2 we specify the solutions for contamination (i.e., $\delta = 0$) and total variation (i.e., $\varepsilon = 0$) neighborhoods, respectively. Subsequently, we introduce numerical algorithms for the computation of the finite-sample risks and compare the results of these algorithms with analytical and empirical results as well as with results obtain by means of higher order approximations; confer Section 11.3. The results of various numerical comparisons between finite-sample and asymptotic results are given in Section 11.4. Finally, we give a short description of our `R` package `ROptEst` which can be used to compute the finite-sample and asymptotic minimax estimator; confer Section 11.5.

11.1 Contamination Neighborhoods

We assume $\delta = 0$. That is, we consider the following contamination neighborhoods,

$$\mathcal{U}_c(\theta, \varepsilon_n) = \big\{Q \in \mathcal{M}_1(\mathbb{B}) \,\big|\, Q = (1-\varepsilon_n)\mathcal{N}(\theta, 1) + \varepsilon_n H, \text{ any } H \in \mathcal{M}_1(\mathbb{B})\big\} \quad (11.1.1)$$

for given radius $\varepsilon_n = \varepsilon/\sqrt{n} \in (0, 1)$ and fixed sample size $n \in \mathbb{N}$; confer (1.2.4).

11.1.1 Finite-Sample Minimax Estimator

For a given number $\tau_n = \tau/\sqrt{n} \in (0, \infty)$ a least favorable pair $(R'_{-\tau_n,c}, R''_{\tau_n,c})$ for $\mathcal{U}_c(-\tau_n, \varepsilon_n)$ versus $\mathcal{U}_c(\tau_n, \varepsilon_n)$ by Subsubsection 10.2.2.1 is

$$R'_{-\tau_n,c}(dy) = (1-\varepsilon_n)\varphi(y+\tau_n)\max\big\{1, \exp\big(2\tau_n(y - b_c^{\text{fi}})\big)\big\}\lambda(dy) \quad (11.1.2)$$

$$R''_{\tau_n,c}(dy) = (1-\varepsilon_n)\varphi(y-\tau_n)\max\big\{1, \exp\big(-2\tau_n(y + b_c^{\text{fi}})\big)\big\}\lambda(dy) \quad (11.1.3)$$

where b_c^{fi} is the unique solution to

$$\frac{\varepsilon_n}{1-\varepsilon_n} = \exp(-2b_c^{\text{fi}}\tau_n)\Phi(\tau_n - b_c^{\text{fi}}) - \Phi(-\tau_n - b_c^{\text{fi}}) \quad (11.1.4)$$

and Φ is the cumulative distribution function of $\mathcal{N}(0, 1)$. The disjointness condition (10.2.15) reads

$$\varepsilon_n < 1 - \big[2\Phi(\tau_n)\big]^{-1} \quad (11.1.5)$$

which immediately implies $\varepsilon_n < 0.5$ for all $\tau_n \in (0, \infty)$. For a plot of these upper bounds on the radius for different values of $n \in \mathbb{N}$ see Figure 11.1. The minimax robust test $\tilde{\phi}_c$ for $\mathcal{U}_c(-\tau_n)$ versus $\mathcal{U}_c(\tau_n)$, by Subsection 10.2.3 and symmetry, is

$$\tilde{\phi}_c = \tfrac{1}{2}\,\mathrm{I}(h_c > 0) + \tfrac{1}{2}\,\mathrm{I}(h_c \geq 0) \quad (11.1.6)$$

with

$$h_c(u_1, \ldots, u_n) = \sum_{i=1}^{n} \tilde{\psi}_c(u_i) \qquad \tilde{\psi}_c(u) := u \min\left\{1, \frac{b_c^{\text{fi}}}{|u|}\right\} \quad (11.1.7)$$

Thus, the minimax estimator $\tilde{S}_c^{\text{fi}}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\psi}_c(y_i - \tilde{S}_c^{\text{fi}}) = 0 \quad (11.1.8)$$

with equal randomization between the smallest and the largest solutions. The corresponding IC reads

$$\mathrm{IC}_{\tilde{\psi}_c}(u) = A_c^{\text{fi}}\tilde{\psi}_c(u) \qquad (A_c^{\text{fi}})^{-1} = 2\Phi(b_c^{\text{fi}}) - 1 \quad (11.1.9)$$

Remark 11.1.1 The limiting estimate (i.e., $\varepsilon_n = 1 - \big[2\Phi(\tau_n)\big]^{-1}$) is the sample median with equal randomization between the minimal and maximal median; confer Subsubsection 10.2.4.1. ////

11.1.2 Asymptotic Minimax Estimator

By Theorem 10.3.7 and symmetry, the IC of the asymptotic minimax estimator is

$$\tilde{\eta}_c(u) = A_c^{\mathrm{as}} u \min\left\{1, \frac{b_c^{\mathrm{as}}}{|u|}\right\} \qquad (A_c^{\mathrm{as}})^{-1} = 2\Phi(b_c^{\mathrm{as}}) - 1 \tag{11.1.10}$$

where b_c^{as} is the unique solution to

$$\varepsilon = 2\tau\left[\varphi(b_c^{\mathrm{as}}) - b_c^{\mathrm{as}}\Phi(-b_c^{\mathrm{as}})\right] \tag{11.1.11}$$

and the boundedness restriction (10.3.39) reads

$$\varepsilon < \sqrt{\tfrac{2}{\pi}}\,\tau \tag{11.1.12}$$

see also Figure 11.1. By Theorem 6.2.4 of Rieder (1994), we can regard the asymp-

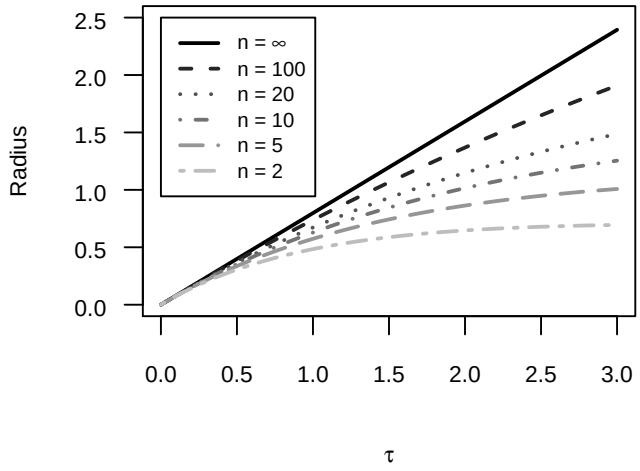

Figure 11.1: Upper bounds for radius ε given width τ derived by conditions (11.1.5) and (11.1.12) in case of contamination neighborhoods $(* = c)$.

totic minimax estimator $\tilde{S}_c^{\mathrm{as}}$ as an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\eta}_c(y_i - \tilde{S}_c^{\mathrm{as}}) = 0 \tag{11.1.13}$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2 which may be shown analogously to Theorem 10.2.9 (a).

Remark 11.1.2 The limiting estimate (i.e., $\varepsilon = \sqrt{\frac{2}{\pi}}\,\tau$) is the sample median with equal randomization between the minimal and maximal median. ////

In addition, we get the following relations between finite-sample and asymptotic results.

Lemma 11.1.3 *For $\varepsilon_n \in (0,1)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

(a)

$$\sqrt{n}\,\big(1 - \big[2\Phi(\tau_n)\big]^{-1}\big) < \sqrt{\tfrac{2}{\pi}}\,\tau \tag{11.1.14}$$

(b)

$$\begin{aligned}\frac{\varepsilon}{1-\varepsilon_n} - \sqrt{n}\,\big[\exp(-2b\tau_n)\Phi(\tau_n - b) - \Phi(-\tau_n - b)\big] \\ = \varepsilon - 2\tau\big[\varphi(b) - b\Phi(-b)\big] + \mathrm{O}(n^{-1/2})\end{aligned} \tag{11.1.15}$$

(c)

$$b_c^{\mathrm{fi}} = b_c^{\mathrm{as}} + \mathrm{O}(n^{-1/2}) \tag{11.1.16}$$

and the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is

$$b_c^{\mathrm{as.c}} := b_c^{\mathrm{as}} - \frac{1}{\sqrt{n}}\,\frac{\varepsilon(\varepsilon + b_c^{\mathrm{as}}\tau)}{2\tau\Phi(-b_c^{\mathrm{as}})} \tag{11.1.17}$$

where

$$b_c^{\mathrm{fi}} \;=\; b_c^{\mathrm{as.c}} + \mathrm{O}(n^{-1}) \tag{11.1.18}$$

Proof

(a) (11.1.14) is equivalent to

$$2\Phi(\tau_n)\left(1 - \sqrt{\tfrac{2}{\pi}}\,\tau_n\right) < 1 \tag{11.1.19}$$

where the left hand side is strictly decreasing in $\tau_n \in (0,\infty)$ because we get for its derivative

$$2\varphi(\tau_n)\left(1 - \sqrt{\tfrac{2}{\pi}}\,\tau_n\right) - 2\sqrt{\tfrac{2}{\pi}}\,\Phi(\tau_n) < 0 \tag{11.1.20}$$

since

$$2\varphi(\tau_n)\left(1 - \sqrt{\tfrac{2}{\pi}}\,\tau_n\right) < 2\varphi(\tau_n) < \sqrt{\tfrac{2}{\pi}} \tag{11.1.21}$$

and

$$2\sqrt{\tfrac{2}{\pi}}\,\Phi(\tau_n) > \sqrt{\tfrac{2}{\pi}} \tag{11.1.22}$$

Thus, the left hand side of (11.1.19) becomes maximal for $\tau_n \downarrow 0$ where $2\Phi(0) = 1$.

(b) We first rewrite the left hand side of (11.1.4),

$$\frac{\varepsilon_n}{1-\varepsilon_n} = \varepsilon_n \sum_{k=0}^{\infty} \varepsilon_n^k = \varepsilon_n + \mathrm{O}(n^{-1}) \tag{11.1.23}$$

Defining $t := 1/\sqrt{n}$, the right hand side of (11.1.4) reads

$$\begin{aligned}&\exp(-2\tau t)\Phi(\tau t - b) - \Phi(-\tau t - b)\\ &\quad = \big[\exp(-2b\tau t) - 1\big]\Phi(\tau t - b) + \Phi(\tau t - b) - \Phi(-\tau t - b) \qquad (11.1.24)\end{aligned}$$

and Taylor expansions in $t = 0$ yield,

$$\begin{aligned}&= \big[-2b\tau t + \mathrm{O}(t^2)\big]\big[\Phi(-b) + \varphi(b)\tau t + \mathrm{O}(t^2)\big]\\ &\qquad +\Phi(-b) + \varphi(b)\tau t + \mathrm{O}(t^2) - \Phi(-b) + \varphi(b)\tau t + \mathrm{O}(t^2) \qquad (11.1.25)\\ &= 2\tau\big[\varphi(b) - b\Phi(-b)\big]t + \mathrm{O}(t^2) \qquad (11.1.26)\end{aligned}$$

(c) Defining $t := 1/\sqrt{n}$, (11.1.4) is equivalent to

$$\frac{\varepsilon t}{1 - \varepsilon t} = \exp(-2b\tau t)\Phi(\tau t - b) - \Phi(-\tau t - b) \qquad (11.1.27)$$

and we obtain by Taylor expansions in $t = 0$

$$\begin{aligned}\varepsilon t + \varepsilon^2 t^2 + \mathrm{O}(t^3) = &\big[-2b\tau t + 2b^2\tau^2 t^2 + \mathrm{O}(t^3)\big]\\ &\big[\Phi(-b) + \varphi(b)\tau t + \tfrac{1}{2}\varphi(b)b\tau^2 t^2 + \mathrm{O}(t^3)\big]\\ &-\Phi(-b) - \varphi(b)\tau t - \tfrac{1}{2}\varphi(b)b\tau^2 t^2 + \mathrm{O}(t^3)\\ &+\Phi(-b) - \varphi(b)\tau t + \tfrac{1}{2}\varphi(b)b\tau^2 t^2 + \mathrm{O}(t^3) \qquad (11.1.28)\\ = &\ 2\tau\big[\varphi(b) - b\Phi(-b)\big]\big[t - b\tau t^2 + \mathrm{O}(t^3)\big] \qquad (11.1.29)\end{aligned}$$

We define

$$G(b,t) := \varepsilon - 2\tau\big[\varphi(b) - b\Phi(-b)\big] + \big[\varepsilon^2 + 2b\tau^2\big(\varphi(b) - b\Phi(-b)\big)\big]t + R(b,t) \qquad (11.1.30)$$

where $b = b(t)$, $b(0) = b_c^{\mathrm{as}}$ and $G(b_c^{\mathrm{as}}, 0) = 0$ by (11.1.11). Moreover, $R(b,t) = \mathrm{O}(t^2)$ as it is a bounded function in b and we omit it from the subsequent calculations. That is, we get

$$\frac{\partial}{\partial b}\, G(b,t)\Big|_{b=b_c^{\mathrm{as}},\, t=0} = \ 2\tau\Phi(-b_c^{\mathrm{as}}) \qquad (11.1.31)$$

$$\frac{\partial}{\partial t}\, G(b,t)\Big|_{b=b_c^{\mathrm{as}},\, t=0} \overset{(11.1.11)}{=} \varepsilon(\varepsilon + b_c^{\mathrm{as}}\tau) \qquad (11.1.32)$$

This leads us to

$$G(b,t) = 2\tau\Phi(-b_c^{\mathrm{as}})(b - b_c^{\mathrm{as}}) + \varepsilon(\varepsilon + b_c^{\mathrm{as}})t + \mathrm{O}\left(|b - b_c^{\mathrm{as}}|t\right) \qquad (11.1.33)$$

That is, $G(b_c^{\mathrm{as}}, t) = \mathrm{O}(t)$ and for any other $b_0 \neq b_c^{\mathrm{as}}$ we have $G(b_0, t) = \mathrm{O}(1)$; i.e., (11.1.16) holds. Furthermore, we obtain $G(b_c^{\mathrm{as.c}}, t) = \mathrm{O}(t^2)$ and for any other b_1 such that $\limsup_{t\to 0} \frac{|b_1 - b_c^{\mathrm{as.c}}|}{t} > 0$ we have $\limsup_{t\to 0} \frac{G(b_1,t)}{t} > 0$; i.e., (11.1.18) holds. ////

11.2 Total Variation Neighborhoods

We assume $\varepsilon = 0$. That is, we consider the following total variation neighborhoods,

$$\mathcal{U}_v(\theta, \delta_n) = \big\{ Q \in \mathcal{M}_1(\mathbb{B}) \,\big|\, d_v\big(Q, \mathcal{N}(\theta, 1)\big) \le \delta_n \big\} \tag{11.2.1}$$

for given radius $\delta_n = \delta/\sqrt{n} \in (0,1)$ and fixed sample size $n \in \mathbb{N}$; confer (1.2.5).

11.2.1 Finite-Sample Minimax Estimator

For a given number $\tau_n = \tau/\sqrt{n} \in (0, \infty)$ a least favorable pair $(R'_{-\tau_n,v}, R''_{\tau_n,v})$ for $\mathcal{U}_v(-\tau_n, \delta_n)$ versus $\mathcal{U}_v(\tau_n, \delta_n)$, by Subsubsection 10.2.2.1, is

$$R'_{-\tau_n,v}(dy) = \varphi(y+\tau_n) \begin{cases} \frac{1+\exp(2\tau_n y)}{1+\exp(-2\tau_n b_v^{\mathrm{fi}})} & y < -b_v^{\mathrm{fi}} \\ 1 & |y| \le b_v^{\mathrm{fi}} \\ \frac{1+\exp(2\tau_n y)}{1+\exp(2\tau_n b_v^{\mathrm{fi}})} & y > b_v^{\mathrm{fi}} \end{cases} \lambda(dy) \tag{11.2.2}$$

$$R''_{\tau_n,v}(dy) = \varphi(y-\tau_n) \begin{cases} \frac{1+\exp(-2\tau_n y)}{1+\exp(2\tau_n b_v^{\mathrm{fi}})} & y < -b_v^{\mathrm{fi}} \\ 1 & |y| \le b_v^{\mathrm{fi}} \\ \frac{1+\exp(-2\tau_n y)}{1+\exp(-2\tau_n b_v^{\mathrm{fi}})} & y > b_v^{\mathrm{fi}} \end{cases} \lambda(dy) \tag{11.2.3}$$

where b_v^{fi} is the unique solution to

$$\big[1 + \exp(-2b_v^{\mathrm{fi}}\tau_n)\big]\delta_n = \exp(-2b_v^{\mathrm{fi}}\tau_n)\Phi(\tau_n - b_v^{\mathrm{fi}}) - \Phi(-\tau_n - b_v^{\mathrm{fi}}) \tag{11.2.4}$$

The disjointness condition (10.2.15) reads

$$\delta_n < \Phi(\tau_n) - \tfrac{1}{2} \tag{11.2.5}$$

which immediately implies $\delta_n < 0.5$ for all $\tau_n \in (0, \infty)$. For a plot of these upper bounds on the radius for different values of $n \in \mathbb{N}$ see Figure 11.2. The minimax robust test $\tilde{\phi}_v$ for $\mathcal{U}_v(-\tau_n)$ versus $\mathcal{U}_v(\tau_n)$, by Subsection 10.2.3 and symmetry, is,

$$\tilde{\phi}_v = \tfrac{1}{2}\,\mathrm{I}(h_v > 0) + \tfrac{1}{2}\,\mathrm{I}(h_v \ge 0) \tag{11.2.6}$$

with

$$h_v(u_1, \ldots, u_n) = \sum_{i=1}^{n} \psi_v(u_i) \qquad \psi_v(u) := u \min\left\{1, \frac{b_v^{\mathrm{fi}}}{|u|}\right\} \tag{11.2.7}$$

Thus, the minimax estimator $\tilde{S}_v^{\mathrm{fi}}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \psi(y_i - \tilde{S}_v^{\mathrm{fi}}) = 0 \tag{11.2.8}$$

with equal randomization between the smallest and the largest solutions. The corresponding IC reads

$$\mathrm{IC}_{\tilde{\psi}_v}(u) = A_v^{\mathrm{fi}} \tilde{\psi}_c(u) \qquad (A_v^{\mathrm{fi}})^{-1} = 2\Phi(b_v^{\mathrm{fi}}) - 1 \tag{11.2.9}$$

Remark 11.2.1 (a) Assume $b_c^{\mathrm{fi}} = b_v^{\mathrm{fi}}$ for fixed $\tau \in (0, \infty)$, then $\varepsilon = \big(1 + \exp(-2b_v^{\mathrm{fi}}\tau_n)\big)\delta$ by (11.1.4) and (11.2.4) where $0 \leq \exp(-2b_v^{\mathrm{fi}}\tau_n) \leq 1$. That is, $\varepsilon \in [\delta, 2\delta]$.

(b) The limiting estimate (i.e., $\delta_n = \Phi(\tau_n) - \frac{1}{2}$) is the sample median with equal randomization between the minimal and maximal median; confer Subsubsection 10.2.4.1. ////

11.2.2 Asymptotic Minimax Estimator

By Theorem 10.3.7 and symmetry the IC of the asymptotic minimax estimator is

$$\tilde{\eta}_v(u) = A_v^{\mathrm{as}} u \min\left\{1, \frac{b_v^{\mathrm{as}}}{|u|}\right\} \qquad (A_v^{\mathrm{as}})^{-1} = 2\Phi(b_v^{\mathrm{as}}) - 1 \tag{11.2.10}$$

where b_v^{as} is the unique solution to

$$\delta = \tau\big[\varphi(b_v^{\mathrm{as}}) - b_v^{\mathrm{as}}\Phi(-b_v^{\mathrm{as}})\big] \tag{11.2.11}$$

and the boundedness restriction (10.3.39) reads

$$\delta < \tfrac{1}{\sqrt{2\pi}}\,\tau \tag{11.2.12}$$

see also Figure 11.2. By Theorem 6.2.4 of Rieder (1994), we can regard the asymp-

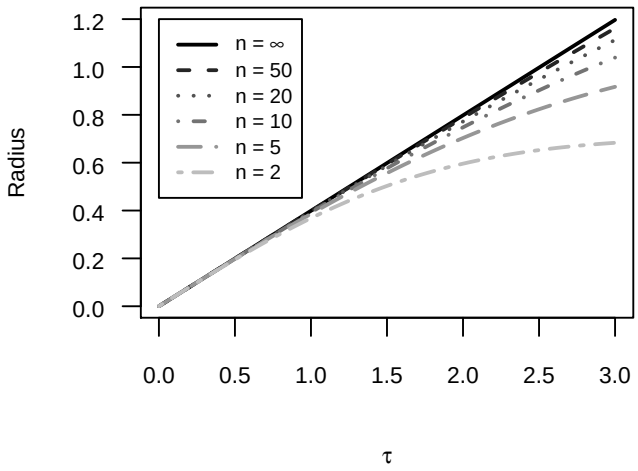

Figure 11.2: Upper bounds for the radius δ $(* = v)$ given width τ derived by conditions (11.2.5) and (11.2.12) in case of total variation neighborhoods $(* = v)$.

totic minimax estimator $\tilde{S}_v^{\text{as}}$ as an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\eta}_v(y_i - \tilde{S}_v^{\text{as}}) = 0 \tag{11.2.13}$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2, which may be shown analogously to Theorem 10.2.9 (a).

Remark 11.2.2 (a) Assume $b_c^{\text{as}} = b_v^{\text{as}}$ for fixed $\tau \in (0, \infty)$, then $\varepsilon = 2\delta$ by (11.1.11) and (11.2.11).

(b) The limiting estimate (i.e., $\delta = \frac{1}{\sqrt{2\pi}}\,\tau$) is the sample median with equal randomization between the minimal and maximal median. ////

In addition, we get the following relations between finite-sample and asymptotic results.

Lemma 11.2.3 *For $\delta_n \in (0,1)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

(a)

$$\sqrt{n}\,\big(\Phi(\tau_n) - \tfrac{1}{2}\big) < \tfrac{1}{\sqrt{2\pi}}\,\tau \tag{11.2.14}$$

(b)

$$\begin{aligned}\big(1 + \exp(-2b\tau_n)\big)\delta - \sqrt{n}\,\big[\exp(-2b\tau_n)\Phi(\tau_n - b) - \Phi(-\tau_n - b)\big] \\ = \delta - \tau\big[\varphi(b) - b\Phi(-b)\big] + \mathrm{O}(n^{-1/2})\end{aligned} \tag{11.2.15}$$

(c)

$$b_v^{\text{as}} = b_v^{\text{fi}} + \mathrm{O}(n^{-1}) \tag{11.2.16}$$

and the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is

$$b_v^{\text{as.c}} = b_v^{\text{as}} - \frac{1}{n}\,\frac{\tau\big[2\big(b_v^{\text{as}}\big)^2\delta + \tau\varphi(b_v^{\text{as}})\big]}{6\Phi(-b_v^{\text{as}})} \tag{11.2.17}$$

where

$$b_v^{\text{fi}} = b_v^{\text{as.c}} + \mathrm{O}(n^{-3/2}) \tag{11.2.18}$$

Proof

(a) (11.2.14) is equivalent to

$$\Phi(\tau_n) - \tfrac{1}{\sqrt{2\pi}}\,\tau_n < \tfrac{1}{2} \tag{11.2.19}$$

where the left hand side is strictly decreasing in $\tau_n \in (0,\infty)$ because we get for its derivative

$$\varphi(\tau_n) - \frac{1}{\sqrt{2\pi}} < 0 \qquad \text{since} \qquad \varphi(\tau_n) < \frac{1}{\sqrt{2\pi}} \tag{11.2.20}$$

Thus, the left hand side of (11.2.19) becomes maximal for $\tau_n \downarrow 0$ where $\Phi(0) = \frac{1}{2}$.

(b) Defining $t := 1/\sqrt{n}$, the left hand side of (11.2.4) reads,

$$\delta t\big(1 + \exp(-2b\tau t)\big) \tag{11.2.21}$$

and Taylor expansion in $t = 0$ yields,

$$\delta t(2 + \mathrm{O}(t)) = 2\delta_n + \mathrm{O}(n^{-1}) \tag{11.2.22}$$

This proves part (b) as the right hand side of (11.2.4) is identical to the right hand side of (11.1.4) which already occurred in part (b) of Lemma 11.1.3.

(c) Defining $t := 1/\sqrt{n}$, (11.2.4) is equivalent to

$$\delta t\big[1 + \exp(-2b\tau t)\big] = \exp(-2b\tau t)\Phi(\tau t - b) - \Phi(-\tau t - b)\big] \tag{11.2.23}$$

and we obtain by Taylor expansions in $t = 0$

$$\delta t\big[1 - b\tau t + b^2\tau^2 t^2\big] = \tau\big[\varphi(b) - b\Phi(-b)\big]\big[t - b\tau t^2 + \tfrac{1}{2}b^2\tau^2 t^3\big] + \mathrm{O}(t^4) \tag{11.2.24}$$

We define

$$\begin{aligned} G(b,t) := {} & \big\{\delta - \tau\big[\varphi(b) - b\Phi(-b)\big]\big\}(1 - b\tau t) \\ & + \big\{\delta - \tfrac{2}{3}\tau\big[\varphi(b) - b\Phi(-b)\big]\big\}b^2\tau^2 t^2 + \tfrac{1}{6}\tau^3\varphi(b)t^2 + R(b,t) \end{aligned} \tag{11.2.25}$$

where the remainder $R(b,t) = \mathrm{O}(t^3)$ as it is a smooth and bounded function in b. Moreover, $b = b(t)$, $b(0) = b_v^{\mathrm{as}}$ and $G(b_v^{\mathrm{as}}, 0) = 0$ by (11.2.11) and we get

$$\frac{\partial}{\partial b} G(b,t)\Big|_{b=b_v^{\mathrm{as}},\, t=0} = \tau\Phi(-b_v^{\mathrm{as}}) \tag{11.2.26}$$

$$\frac{\partial}{\partial t} G(b,t)\Big|_{b=b_v^{\mathrm{as}},\, t=0} \overset{(11.2.11)}{=} 0 \tag{11.2.27}$$

$$\frac{\partial^2}{\partial t^2} G(b,t)\Big|_{b=b_v^{\mathrm{as}},\, t=0} \overset{(11.2.11)}{=} \tfrac{1}{3}\tau^2\big[2\big(b_v^{\mathrm{as}}\big)^2\delta + \tau\varphi(b_v^{\mathrm{as}})\big] \tag{11.2.28}$$

This leads us to

$$G(b,t) = \tau\Phi(-b_v^{\mathrm{as}})(b - b_v^{\mathrm{as}}) + \tfrac{1}{6}\tau^2\big[2\big(b_v^{\mathrm{as}}\big)^2\delta - \tau\varphi(b_v^{\mathrm{as}})\big]t^2 + \mathrm{O}\left(|b - b_v^{\mathrm{as}}|t\right) \tag{11.2.29}$$

That is, $G(b_v^{\mathrm{as}}, t) = \mathrm{O}(t^2)$ and for any other $b_0 \neq b_v^{\mathrm{as}}$ we have $G(b_0, t) = \mathrm{O}(1)$; i.e., (11.2.16) holds. Furthermore, we obtain $G(b_v^{\mathrm{as.c}}, t) = \mathrm{O}(t^3)$ and for any other b_1 such that $\limsup_{t\to 0} \frac{|b_1 - b_v^{\mathrm{as.c}}|}{t^2} > 0$ we have $\limsup_{t\to 0} \frac{G(b_1,t)}{t^2} > 0$; i.e., (11.2.18) holds. ////

Remark 11.2.4 (a) Part (c) of the preceding lemma and of Lemma 11.1.3 as well as the corresponding results in case of simple linear regression (cf. Chapter 12) were initiated by P. Ruckdeschel's recent work on higher order asymptotics; confer Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c), Ruckdeschel (2005e) and Ruckdeschel and Kohl (2005).

(b) Lemma 11.1.3 and Lemma 11.2.3 show that there is a clear difference between contamination and total variation neighborhoods concerning the speed of convergence of the optimal clipping bounds ($\mathrm{O}(n^{-1/2})$ vs. $\mathrm{O}(n^{-1})$). Thus, one might expect that this also holds for the speed of convergence of the finite-sample risks. Indeed, this is indicated by our numerical results; confer Subsubsections 11.4.1.2 and 11.4.2.2. We conjecture that this is caused by the higher symmetry of total variation neighborhoods, which also shows up by calculating the Edgeworth expansions; confer Remark 11.3.7. ////

11.3 Computation of the Finite-Sample Risk

In this section we first take a closer look at the finite-sample risk; confer Subsection 11.3.1. Since the analytically exact determination of this risk is very difficult or even impossible for $n \geq 3$, we present two procedures for the numerical computation of the finite-sample risk; confer Subsection 11.3.2. Different checks which, to some extent, are included in Subsection 11.3.2 show that these two algorithms yield astonishingly accurate results. Since we consider contamination and total variation neighborhoods simultaneously in this section, the index $*$ substitutes the indices c (contamination neighborhoods) and v (total variation neighborhoods).

11.3.1 Finite-Sample Risk

We fix $n \in \mathbb{N}$ and radius $\varepsilon_n \in [0,1)$ $(* = c)$, respectively radius $\delta_n \in [0,1)$ $(* = v)$. Given some clipping bound $b \in (0,\infty)$, we then want to determine the finite-sample risk of an M estimator S satisfying

$$\sum_{i=1}^{n} \chi_0(y_i - S) = 0 \qquad \chi_0(u) = u \min\left\{1, \frac{b}{|u|}\right\} \tag{11.3.1}$$

with equal randomization between the smallest and the largest solutions. In particular, we are interested in the finite-sample minimax estimator $\tilde{S}_*^{\mathrm{fi}}$, the asymptotic minimax estimator $\tilde{S}_*^{\mathrm{as}}$ and the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected $(* = c)$, respectively $\mathrm{O}(n^{-1})$-corrected $(* = v)$ asymptotic optimal clipping bound. All these estimators are equivariant in sense of Definition 10.2.2; confer Theorem 10.2.9. As already stated in Remark 10.2.3, the finite-sample risk of any equivariant estimator S for given $\tau_n \in (0,\infty)$ reads,

$$\mathrm{Risk}\,(S;*) = \max\left\{ \sup_{Q_{-\tau_n} \in \mathcal{U}_*(-\tau_n)} Q^n_{-\tau_n}(S > 0), \sup_{Q_{\tau_n} \in \mathcal{U}_*(\tau_n)} Q^n_{\tau_n}(S < 0) \right\} \tag{11.3.2}$$

with $\mathcal{U}_*(-\tau_n) = g_{-\theta-\tau_n}\big(\mathcal{U}_*(\theta)\big)$ and $\mathcal{U}_*(\tau_n) = g_{-\theta+\tau_n}\big(\mathcal{U}_*(\theta)\big)$ by (10.1.12). Moreover, for any M estimator that is based on a score function $\chi_\theta(u)$, which is measurable in u and monotone increasing in $\theta \in \mathbb{R}$, from strictly positive to strictly

negative values the following inclusions

$$\{S' > \theta\} \subseteq \Big\{ \textstyle\sum_{i=1}^n \chi_\theta(y_i) > 0 \Big\} \subseteq \{S' \geq \theta\} \tag{11.3.3}$$

$$\{S'' > \theta\} \subseteq \Big\{ \textstyle\sum_{i=1}^n \chi_\theta(y_i) \geq 0 \Big\} \subseteq \{S'' \geq \theta\} \tag{11.3.4}$$

where

$$S'(y_1, \ldots, y_n) = \sup_{\theta \in \mathbb{R}} \Big\{ \textstyle\sum_{i=1}^n \chi_\theta(y_i) > 0 \Big\} \tag{11.3.5}$$

$$S''(y_1, \ldots, y_n) = \inf_{\theta \in \mathbb{R}} \Big\{ \textstyle\sum_{i=1}^n \chi_\theta(y_i) < 0 \Big\} \tag{11.3.6}$$

for any given $y_1, \ldots, y_n$. For more details we refer to pp 45 of Huber (1981). We have $\chi_\theta(y) = \chi_0(y - \theta)$. Therefore, we obtain for any such M estimator S and any $Q_{-\tau_n} \in \mathcal{U}_*(-\tau_n)$, respectively $Q_{\tau_n} \in \mathcal{U}_*(\tau_n)$,

$$Q^n_{-\tau_n}(S > 0) = \tfrac{1}{2}\big[Q^n_{-\tau_n}(S' > 0) + Q^n_{-\tau_n}(S'' > 0)\big] \tag{11.3.7}$$

$$\leq \tfrac{1}{2}\left[Q^n_{-\tau_n}\left(\sum_{i=1}^n \chi_0(y_i) > 0\right) + Q^n_{-\tau_n}\left(\sum_{i=1}^n \chi_0(y_i) \geq 0\right)\right] \tag{11.3.8}$$

$$\leq \tfrac{1}{2}\big[Q^n_{-\tau_n}(S' \geq 0) + Q^n_{-\tau_n}(S'' \geq 0)\big] \tag{11.3.9}$$

respectively

$$Q^n_{\tau_n}(S < 0) = \tfrac{1}{2}\big[Q^n_{\tau_n}(S' < 0) + Q^n_{\tau_n}(S'' < 0)\big] \tag{11.3.10}$$

$$\leq \tfrac{1}{2}\left[Q^n_{\tau_n}\left(\sum_{i=1}^n \chi_0(y_i) \leq 0\right) + Q^n_{\tau_n}\left(\sum_{i=1}^n \chi_0(y_i) < 0\right)\right] \tag{11.3.11}$$

$$\leq \tfrac{1}{2}\big[Q^n_{\tau_n}(S' \leq 0) + Q^n_{\tau_n}(S'' \leq 0)\big] \tag{11.3.12}$$

Remark 11.3.1 If $Q_\theta \in \mathcal{U}_*(\theta)$ $(\theta \in \mathbb{R})$ is absolutely continuous, we get by Lemma 10.2.4, respectively Lemma 10.6.1 of Huber (1981) that the distribution of any estimator $S = S_n$ $(n \in \mathbb{N}$ fixed), which is equivariant in sense of Definition 10.2.2, is absolutely continuous under Q^n_θ. As a direct consequence, we obtain equality in (11.3.8) and (11.3.9), respectively in (11.3.11) and (11.3.12) if $Q_{-\tau_n}$, respectively Q_{τ_n} is absolutely continuous. ////

By monotonicity of χ_0 the distribution of $\sum_{i=1}^n \chi_0(y_i)$ is monotone with respect to stochastic order; i.e., the probability of $\sum_{i=1}^n \chi_0(y_i) > 0$ and $\sum_{i=1}^n \chi_0(y_i) \geq 0$, respectively, under $Q_{-\tau_n} \in \mathcal{U}_*(-\tau_n)$ is maximal if

$$Q_{-\tau_n}(\chi_0(y) = b) = Q_{-\tau_n}(y \geq b) \overset{(10.1.12)}{=} Q_0(y \geq b + \tau_n) = \max! \tag{11.3.13}$$

where $b \in (0, \infty)$ is some given clipping bound. Analogously, we obtain that the probability of $\sum_{i=1}^n \chi_0(y_i) < 0$ and $\sum_{i=1}^n \chi_0(y_i) \leq 0$, respectively, under $Q_{\tau_n} \in \mathcal{U}_*(\tau_n)$ is maximal if

$$Q_{\tau_n}(\chi_0(y) = -b) = Q_{\tau_n}(y \leq -b) \overset{(10.1.12)}{=} Q_0(y \leq -b - \tau_n) = \max! \tag{11.3.14}$$

In case of contamination neighborhoods $(* = c)$ we get $Q'_{-\tau_n} \in \mathcal{U}_c(-\tau_n)$ and $Q''_{\tau_n} \in \mathcal{U}_c(\tau_n)$ with

$$Q'_{-\tau_n} = (1-\varepsilon_n)\mathcal{N}(-\tau_n, 1) + \varepsilon_n H'_{-\tau_n} \tag{11.3.15}$$

and

$$Q''_{\tau_n} = (1-\varepsilon_n)\mathcal{N}(\tau_n, 1) + \varepsilon_n H''_{\tau_n} \tag{11.3.16}$$

where $H'_{-\tau_n}$ and H''_{τ_n} are concentrated on $[\tau_n + b, \infty)$ and $(-\infty, -\tau_n - b]$, respectively. In case of total variation neighborhoods $(* = v)$ this leads us to $Q'_{-\tau_n} \in \mathcal{U}_v(-\tau_n)$ and $Q''_{\tau_n} \in \mathcal{U}_v(\tau_n)$ having cumulative distribution functions (cf. also Rieder (1994), pp 174)

$$Q'_{-\tau_n}(t) = \big(\Phi(t+\tau_n) - \delta_n\big)_+ + \delta_n H'_{-\tau_n}(t) \tag{11.3.17}$$

and

$$Q''_{\tau_n}(t) = \min\big\{\big[\Phi(t-\tau_n) + \delta_n H''_{\tau_n}(t)\big], 1\big\} \tag{11.3.18}$$

for all $t \in \mathbb{R}$, where $H'_{-\tau_n}$ and H''_{τ_n} are concentrated on $[\tau_n+b, \infty)$ and $(-\infty, -\tau_n-b]$, respectively. That is, in case of $Q'_{-\tau_n}$ mass δ_n is moved from the left tail to $[\tau_n+b, \infty)$, whereas in case of Q''_{τ_n} it is moved from the right tail to $(-\infty, -\tau_n-b]$.

Remark 11.3.2 **(a)** In view of Remark 11.3.1 we have to choose absolutely continuous distributions $H'_{-\tau_n}$ and H''_{τ_n} to obtain equality in (11.3.8) and (11.3.9), respectively in (11.3.11) and (11.3.12). Then, we may compute the finite-sample risk (11.3.2) by determining the distribution of $\sum_{i=1}^n \chi_0(y_i)$ under $Q'_{-\tau_n}$ and Q''_{τ_n} where it holds by symmetry,

$$\text{Risk}\,(S; *) = \big(Q'_{-\tau_n}\big)^n(S > 0) = \big(Q''_{\tau_n}\big)^n(S < 0) \tag{11.3.19}$$

(b) If $H'_{-\tau_n}$ or H''_{τ_n} is not absolutely continuous, we cannot apply Lemma 10.2.4, respectively Lemma 10.6.1 of Huber (1981) to obtain equality in (11.3.8) and (11.3.9), respectively in (11.3.11) and (11.3.12). However, if a random variable X has an arbitrary distribution and the random variable $Y > 0$ is stochastically independent of X with an absolutely continuous distribution, then the distribution of $X + Y$, respectively $X - Y$ is absolutely continuous and stochastically larger, respectively smaller than the distribution of X; i.e.,

$$P(X + Y > t) \geq P(X > t) \geq P(X - Y > t) \tag{11.3.20}$$

Thus, for any given $H'_{-\tau_n}$ or H''_{τ_n}, that is not absolutely continuous, we can specify a distribution that is absolutely continuous and stochastically larger, respectively smaller than $H'_{-\tau_n}$ or H''_{τ_n} and equality holds in (11.3.8) and (11.3.9), respectively in (11.3.11) and (11.3.12) for this new distribution.

(c) By part (b) of this remark and as the distributions of S' and S'' are monotone with respect to stochastic order, there is at least equality in (11.3.9) and (11.3.12) even if $H'_{-\tau_n}$ or H''_{τ_n} is not absolutely continuous. But, (11.3.7) or (11.3.10), respectively could be really smaller than (11.3.8) or (11.3.11), respectively. That is, to make sure that the finite-sample risk is really attained under

$Q'_{-\tau_n}$ and Q''_{τ_n}, (at least) 0 has to be a continuity point of the distributions of S' and S'' under $Q'_{-\tau_n}$ and Q''_{τ_n}. ////

We may now calculate the finite-sample risk by determining the distribution of $\sum_{i=1}^n \chi_0(y_i)$ under $\left(Q'_{-\tau_n}\right)^n$ or $\left(Q''_{\tau_n}\right)^n$, respectively. As a first step, we state the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n} where in case of contamination neighborhoods ($* = c$) we have

$$Q'_{-\tau_n}(\chi_0(y) = -b) = (1-\varepsilon_n)\Phi(-b+\tau_n) \tag{11.3.21}$$

$$Q'_{-\tau_n}(-b < \chi_0(y) < t) = (1-\varepsilon_n)\big[\Phi(t+\tau_n) - \Phi(-b+\tau_n)\big] \quad t \in (-b,b) \tag{11.3.22}$$

$$Q'_{-\tau_n}(\chi_0(y) = b) = (1-\varepsilon_n)\Phi(-b-\tau_n) + \varepsilon_n \tag{11.3.23}$$

respectively

$$Q''_{\tau_n}(\chi_0(y) = -b) = (1-\varepsilon_n)\Phi(-b-\tau_n) + \varepsilon_n \tag{11.3.24}$$

$$Q''_{\tau_n}(-b < \chi_0(y) < t) = (1-\varepsilon_n)\big[\Phi(t-\tau_n) - \Phi(-b-\tau_n)\big] \quad t \in (-b,b) \tag{11.3.25}$$

$$Q''_{\tau_n}(\chi_0(y) = b) = (1-\varepsilon_n)\Phi(-b+\tau_n) \tag{11.3.26}$$

In case of total variation neighborhoods ($* = v$) we get,

$$Q'_{-\tau_n}(\chi_0(y) = -b) = \big(\Phi(-b+\tau_n) - \delta_n\big)_+ \tag{11.3.27}$$

$$\begin{aligned} &Q'_{-\tau_n}(-b < \chi_0(y) < t) \\ &\quad = \big(\Phi(t+\tau_n) - \delta_n\big)_+ - \big(\Phi(-b+\tau_n) - \delta_n\big)_+ \qquad t \in (-b,b) \end{aligned} \tag{11.3.28}$$

$$Q'_{-\tau_n}(\chi_0(y) = b) = 1 - \big(\Phi(b+\tau_n) - \delta_n\big)_+ \tag{11.3.29}$$

respectively

$$Q''_{\tau_n}(\chi_0(y) = -b) = \min\big\{\Phi(-b-\tau_n) + \delta_n, 1\big\} \tag{11.3.30}$$

$$\begin{aligned} &Q''_{\tau_n}(-b < \chi_0(y) < t) \\ &\quad = \min\big\{\Phi(t-\tau_n) + \delta_n, 1\big\} - \min\big\{\Phi(-b-\tau_n) + \delta_n, 1\big\} \quad t \in (-b,b) \end{aligned} \tag{11.3.31}$$

$$Q''_{\tau_n}(\chi_0(y) = b) = 1 - \min\big\{\Phi(b-\tau_n) + \delta_n, 1\big\} \tag{11.3.32}$$

11.3.2 Numerical Algorithms

In this subsection we present two algorithms for the numerical computation of the finite-sample risk and check their precision where χ_0 is defined as in the preceding subsection; confer (11.3.1).

11.3.2.1 Algorithm A

This procedure directly uses the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n}, respectively, which can be read off from equations (11.3.21)–(11.3.32). The n-fold convolution of this distribution is calculated using Algorithm C.2.2 which is based on the fast Fourier transform; confer Appendix C.2. A comparison with (numerically) exact convolution results shows that this procedure yields very accurate results; confer Appendix C.3.

Remark 11.3.3 Since the law of $\tilde{\psi}_*$ $(* = c, v)$ under $Q'_{-\tau_n}$ and Q''_{τ_n} puts mass at the points $-b_*, b_*$, the n-fold convolution for even n, among other points (multiples of $-b_*, b_*$), puts mass at zero. Thus,

$$(Q'_{-\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) > 0\Big) < (Q'_{-\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) \geq 0\Big) \tag{11.3.33}$$

and

$$(Q''_{\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) < 0\Big) < (Q''_{\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) \leq 0\Big) \tag{11.3.34}$$

and the corresponding non-randomized estimate $\frac{1}{2}(S' + S'')$ is not minimax; confer also Section 4 of Huber (1968). In contrast, if n is odd, there is no mass at zero and therefore equality holds in (11.3.33) and (11.3.34). However, for increasing n the difference between $>$ and $\geq$ in (11.3.33) and $<$ and $\leq$ in (11.3.34) becomes small very fast as the mass at zero decays exponentially in n. This immediately follows from Theorem 2.3 of Hall (1992). ////

11.3.2.2 Algorithm B

This algorithm was developed in joint work with P. Ruckdeschel. Since all four cases $(\pm\tau_n,\ * = c, v)$ may be treated analogously, we only specify the case $(Q'_{-\tau_n})^n$ and $(* = c)$. To lighten the notation, we define $R := Q'_{-\tau_n}$. In view of (11.3.21)–(11.3.23) we can rewrite,

$$R^n\Big(\sum_{i=1}^n \chi_0(y_i) > 0\Big) = R^n\Big(\sum_{i=1}^n \big[(1 - V_i)Z_i + V_i W_i\big] > 0\Big) \tag{11.3.35}$$

with the following stochastically independent random variables

$$V_i \overset{\text{i.i.d.}}{\sim} \text{Binom}\,(1, p_1) \qquad p_1 := (1 - \varepsilon_n)\big[\Phi(-b + \tau_n) + \Phi(-b - \tau_n)\big] + \varepsilon_n \tag{11.3.36}$$

$$Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{L}\left(\tilde{Z}_i \,\middle|\, \tilde{Z}_i \in [-b, b]\right) \qquad \tilde{Z}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\,(-\tau_n, 1) \tag{11.3.37}$$

$$W_i := 2b\tilde{W}_i - b \qquad \tilde{W}_i \overset{\text{i.i.d.}}{\sim} \text{Binom}\,(1, p_2) \tag{11.3.38}$$

where

$$p_2 := \big[(1 - \varepsilon_n)\Phi(-b - \tau_n) + \varepsilon_n\big]/p_1 \tag{11.3.39}$$

and we abbreviate sums of these random variables of length $m \leq n$ by a superscript (m) at the random variable. By stochastic independence we get,

$$\begin{aligned} &R^n\Big(\sum_{i=1}^n \big[(1 - V_i)Z_i + V_i W_i\big] > 0\Big) \\ &\quad = R^n\big(Z^{(n)} > 0\big)R^n\big(V^{(n)} = 0\big) + \sum_{j=1}^{n-1} R^n\big(2b\tilde{W}^{(j)} + Z^{(n-j)} > jb\big)R^n\big(V^{(n)} = j\big) \\ &\quad\quad + R^n\big(\tilde{W}^{(n)} > n/2\big)R^n\big(V^{(n)} = n\big) \end{aligned} \tag{11.3.40}$$

with

$$R^n\big(2b\tilde{W}^{(j)} + Z^{(n-j)} > jb\big) = \sum_{k=0}^{j} R^{n-j}\big(Z^{(n-j)} > b(j-2k)\big) R^j\big(\tilde{W}^{(j)} = k\big) \tag{11.3.41}$$

Of course, we can do the same calculations for $>$ replaced by $\geq$. Thus, the finite-simple risk by absolute continuity of the law of $Z^{(m)}$ under R^m reads,

$$\begin{aligned}\text{Risk}(S;c) = {} & R^n\big(Z^{(n)} > 0\big) R^n\big(V^{(n)} = 0\big) \\ & + \sum_{j=1}^{n-1} \sum_{k=0}^{j} R^{n-j}\big(Z^{(n-j)} > b(j-2k)\big) R^j\big(\tilde{W}^{(j)} = k\big) R^n\big(V^{(n)} = j\big) \\ & + \tfrac{1}{2}\Big[R^n\big(\tilde{W}^{(n)} > n/2\big) + R^n\big(\tilde{W}^{(n)} \geq n/2\big)\Big] R^n\big(V^{(n)} = n\big) \end{aligned} \tag{11.3.42}$$

The m-fold convolution of the law of Z_i is calculated using Algorithm C.2.2, which is based on the fast Fourier transform; confer Appendix C.2.

Remark 11.3.4 **(a)** If n is even, the law of $\tilde{W}^{(n)}$ puts mass at $n/2$. Thus,

$$R^n\big(\tilde{W}^{(n)} > n/2\big) < R^n\big(\tilde{W}^{(n)} \geq n/2\big) \tag{11.3.43}$$

and again it follows that the non-randomized estimate $\frac{1}{2}(S' + S'')$ is not minimax; confer Remark 11.3.3.

(b) In case of the sample median (limiting estimate) we obtain

$$\text{Risk}\,(\text{Med};c) = \tfrac{1}{2}\Big[R^n\big(\tilde{W}^{(n)} > n/2\big) + R^n\big(\tilde{W}^{(n)} \geq n/2\big)\Big] \tag{11.3.44}$$

where $\varepsilon_n = 1 - \big[2\Phi(\tau_n)\big]^{-1}$; i.e., $p_2 = \frac{1}{2}$.

(c) We use similar algorithms to compute the finite-sample maximum MSE of M estimators; confer Ruckdeschel and Kohl (2005) and Ruckdeschel (2004b). ////

11.3.2.3 Checks

Based on Algorithm B, we may compute the finite-sample risk numerically exact at least for sample size $n = 2$ where analytic calculations yield

$$R^2\big(Z^{(2)} > 0\big) = 1 - R^2\big(Z^{(2)} \leq 0\big) \tag{11.3.45}$$

$$\begin{aligned} &= 1 - \Bigg[2 \int_{-\sqrt{2}\,b}^{0} \varphi(y + \sqrt{2}\,\tau_2)\Phi(y + \sqrt{2}\,b)dy \\ &\qquad -\Phi(\sqrt{2}\,\tau_2) + \Phi\big(-\sqrt{2}\,(b - \tau_2)\big)\Bigg] \Big/ \big[\Phi(b + \tau_2) - \Phi(-b + \tau_2)\big]^2 \end{aligned} \tag{11.3.46}$$

Moreover, if we choose ε, respectively δ such that the corresponding optimal clipping bound $b_* = \tau_2$ $(* = c, v)$ we obtain

$$R^2(Z^{(2)} \leq 0) = \frac{0.25 - \Phi(\sqrt{2}\,b_*)\Phi(-\sqrt{2}\,b_*)}{\big(\Phi(2b_*) - \frac{1}{2}\big)^2} \tag{11.3.47}$$

We restrict our considerations to the computation of the finite-sample risk of the asymptotic minimax estimator $\tilde{S}_*^{\text{as}}$ since $b_*^{\text{as}} = \tau_2$ leads to a larger radius than $b_*^{\text{fi}} = \tau_2$ and Algorithm A is less accurate for this larger radius. Furthermore, we denote the absolute deviation of the finite-sample risk obtained with Algorithm A, respectively Algorithm B from the numerically exact finite-sample risk $\text{Risk}^\natural(\tilde{S}_*^{\text{as}}; *)$ by error_A and error_B, respectively. The results for contamination neighborhoods are contained in Table 11.1 where 2^q denotes the number of lattice points used in step 2 of Algorithm C.2.2. The errors of these algorithms for total variation neighborhoods may be read off from Table 11.2.

ε	τ_2	b_c^{as}	$\text{Risk}^\natural(\tilde{S}_c^{\text{as}}; c)$	q	error_A	error_B
0.0480	2.000	2.000	0.052560	10	5.1e−05	2.4e−08
				12	1.3e−05	1.5e−09
				14	3.2e−06	9.5e−11
0.1243	1.500	1.500	0.142437	10	9.0e−05	5.1e−08
				12	2.2e−05	3.2e−09
				14	5.6e−06	2.0e−10
0.2357	1.000	1.000	0.294290	10	8.4e−05	3.9e−08
				12	2.1e−05	2.4e−09
				14	5.2e−06	1.5e−10

Table 11.1: The precision of the computation of the finite-sample risk for sample size $n = 2$ based on Algorithms A and B in case of contamination neighborhoods $(* = c)$. [$\text{Risk}^\natural(\tilde{S}_c^{\text{as}}; c)$ denotes the numerically exact finite-sample risk of $\tilde{S}_c^{\text{as}}$, error_A, respectively error_B the error of Algorithm A and Algorithm B, respectively and 2^q the number of lattice points used in Algorithm C.2.2.]

Remark 11.3.5 **(a)** Algorithms A and B yield small, respectively very small errors for sample size $n = 2$ which once more confirms the precision of Algorithm C.2.2. In addition, it is not surprising that the accurateness of these algorithms turns out to be independent of the neighborhood-type. Conversely, since both algorithms are based on Algorithm C.2.2 which maintains its high precision with increasing convolution power, we expect the same behavior for Algorithms A and B for increasing sample size $n \in \mathbb{N}$. However, to have another cross check, we compare the results of our algorithms with results obtained by numerical simulations; see Table 11.3.

(b) Huber (1968), p 278, determines the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_*^{\text{fi}}$ in case of $n = 2$, $\tau_2 = 1.0$ and $\varepsilon_2 = 0.0430$, respectively $\delta_2 = 0.0396$ which by (11.1.4), respectively (11.2.4) leads to $b_c^{\text{fi}} = 1.0$, respectively $b_v^{\text{fi}} = 1.0$. Our calculations yield $\text{Risk}^\natural(\tilde{S}_c^{\text{fi}}; c) = 0.140359$, respectively $\text{Risk}^\natural(\tilde{S}_v^{\text{fi}}; v) = 0.142338$ which confirms the results contained in Huber (1968). ////

We next study the differences between Algorithm A and Algorithm B for increasing sample size n where we again consider only the finite-sample risk of the asymptotic

δ	τ_2	b_v^{as}	$\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{as}};v)$	q	error_A	error_B
0.0240	2.000	2.000	0.027821	10	2.6e−05	2.6e−08
				12	6.6e−06	1.6e−09
				14	1.6e−06	1.0e−10
0.0622	1.500	1.500	0.083216	10	4.7e−05	6.2e−08
				12	1.2e−05	3.9e−09
				14	2.9e−06	2.4e−10
0.1178	1.000	1.000	0.206917	10	3.9e−05	5.6e−08
				12	9.7e−06	3.5e−09
				14	2.4e−06	2.2e−10

Table 11.2: The precision of the computation of the finite-sample risk for sample size $n = 2$ based on Algorithms A and B in case of total variation neighborhoods $(* = v)$. $[\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{as}};v)$ denotes the numerically exact finite-sample risk of $\tilde{S}_v^{\mathrm{as}}$, error_A, respectively error_B the error of Algorithm A and Algorithm B, respectively, and 2^q the number of lattice points used in Algorithm C.2.2.$]$

minimax estimator. Since the precision of the results turned out to be (almost) independent of the neighborhood type, we restrict the comparison to contamination neighborhoods. Moreover, we choose one "typical" situation as the results are almost independent of τ and ε. That is, we fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$ which by (11.1.11) leads to $b_c^{\mathrm{as}} = 1.374$ and determine the distance $\mathrm{dist}_{\mathrm{AB}}$ between the results of Algorithm A and Algorithm B. In addition, we give the computation time in seconds T_A, respectively T_B on an Athlon with 2.5 GHz and 512 MB RAM. Table 11.3 shows the corresponding results where $\mathrm{Risk}_{\mathrm{B}}$ denotes the finite-sample risk of S_c^{as} computed with Algorithm B and 2^q stands for the number of lattice points used in step 2 of Algorithm C.2.2. Moreover, to give another cross check for the results of our algorithms, we do some numerical simulations. That is, we simulate 1e 06 samples of size n and compute the empirical finite-sample risk and the corresponding 95% confidence interval. As it turns out, the results of Algorithms A and B always lie well within the 95% confidence interval.

Remark 11.3.6 Obviously, Algorithm B is more accurate than Algorithm A (cf. Tables 11.1, 11.2) but Algorithm A is clearly faster (cf. Tables 11.3). Moreover, the differences between Algorithm A and Algorithm B are small and get even smaller with increasing sample size n. That is, for sample sizes $n \le 10$ we choose Algorithm B whereas for $n > 10$ we work with Algorithm A where we use $q = 12$ in almost all computations. ////

n	emp. risk	95% conf. int.	q	Risk_B	dist_AB	T_A	T_B
3	0.0899	[0.0892, 0.0904]	10	0.090029	3.4e−05	0.01	0.05
			12	0.090029	8.6e−06	0.03	0.19
			14	0.090029	2.1e−06	0.18	1.25
4	0.0810	[0.0804, 0.0816]	10	0.081366	4.1e−05	0.01	0.09
			12	0.081366	1.0e−05	0.04	0.33
			14	0.081366	2.1e−06	0.26	2.44
5	0.0634	[0.0628, 0.0638]	10	0.063232	3.3e−05	0.01	0.18
			12	0.063232	8.3e−06	0.04	0.72
			14	0.063232	2.1e−06	0.32	5.21
10	0.0387	[0.0382, 0.0391]	10	0.038605	2.0e−05	0.02	0.94
			12	0.038605	4.9e−06	0.12	5.30
			14	0.038605	1.2e−06	0.76	31.54
20	0.0275	[0.0272, 0.0278]	10	0.027266	1.1e−05	0.04	7.04
			12	0.027266	2.8e−06	0.36	51.55
			14	0.027266	1.2e−06	1.45	231.94

Table 11.3: A comparison between Algorithm A, Algorithm B and empirical results in case of contamination neighborhoods. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$ which leads to $b_c^{\text{as}} = 1.374$. Risk_B denotes the finite-sample risk of $\tilde{S}_c^{\text{as}}$ computed with Algorithm B, 2^q the number of lattice points used in Algorithm C.2.2, dist_AB the distance between the results of Algorithm A and Algorithm B and T_A, respectively T_B the computation time in seconds for Algorithm A and Algorithm B, respectively.]

11.3.3 Higher Order Approximations

As already noted before, it is very difficult or even impossible to determine the exakt finite-sample risk for sample size $n \geq 3$ analytically. Thus, one might think of higher order approximations as provided by Edgeworth expansions (cf. Subsubsection 11.3.3.1) or saddlepoint approximations (cf. Subsubsection 11.3.3.2) to compute (at least) an approximation of the exakt finite-sample risk. We will compare the results of these techniques with the (numerically) exact results of our algorithms presented in Subsection 11.3.2; confer Subsubsection 11.3.3.3.

11.3.3.1 Edgeworth Expansions

We consider

$$\xi_i = \frac{\chi_0(y_i) - \text{E}_R\, \chi_0}{\sqrt{\text{Var}_R\, \chi_0}} \tag{11.3.48}$$

with some $\chi_0(u)$ as defined in (11.3.1) and assume $R = Q'_{-\tau_n}, Q''_{\tau_n}$ of form (11.3.15)–(11.3.18) to be absolutely continuous. Thus, $\text{E}_R\, |\xi_i|^5 < \infty$ and the corresponding Edgeworth expansion may be read off from Theorem 1 of Ibragimov

(1967) or p 16 of Field and Ronchetti (1990), respectively. That is, we obtain

$$\begin{aligned} & R^n\left(\sum_{i=1}^{n} \frac{\chi_0(y_i) - \mathrm{E}_R\,\chi_0}{\sqrt{\mathrm{Var}_R\,\chi_0}} < \sqrt{n}\,t\right) \\ & = \Phi(t) - \varphi(t)\left[\frac{\rho_R}{6\sqrt{n}}(t^2-1) + \frac{1}{n}\left(\frac{\kappa_R}{24}(t^3-3t) + \frac{\rho_R^2}{72}(t^5-10t^3+15t)\right)\right] \\ & \quad + \mathrm{O}(n^{-3/2}) \end{aligned} \tag{11.3.49}$$

where

$$\rho_R = \mathrm{E}_R\left(\frac{\chi_0 - \mathrm{E}_R\,\chi_0}{\sqrt{\mathrm{Var}_R\,\chi_0}}\right)^3 \qquad \text{and} \qquad \kappa_R = \mathrm{E}_R\left(\frac{\chi_0 - \mathrm{E}_R\,\chi_0}{\sqrt{\mathrm{Var}_R\,\chi_0}}\right)^4 - 3 \tag{11.3.50}$$

To determine an approximation of the finite-sample risk (11.3.2), we have to choose

$$t = -\sqrt{n}\,\frac{\mathrm{E}_R\,\chi_0}{\sqrt{\mathrm{Var}_R\,\chi_0}} \tag{11.3.51}$$

Remark 11.3.7 If k is even, calculating $\mathrm{E}_R\,\chi_0^k$ ($k \in \mathbb{N}$) for total variation neighborhoods we obtain $\mathrm{E}_{Q'_{-\tau_n}}\,\chi_0^k = \int \chi_0^k\,d\mathcal{N}(-\tau_n, 1)$ and $\mathrm{E}_{Q''_{\tau_n}}\,\chi_0^k = \int \chi_0^k\,d\mathcal{N}(\tau_n, 1)$, respectively; i.e., contamination disappears. We conjecture that this is the reason — at least in case of Edgeworth expansions — why the convergence of the finite-sample risks is faster for total variation neighborhoods; confer Remark 11.3.8. ////

11.3.3.2 Saddlepoint Approximations

The assumptions A 4.1–A 4.5 of Theorem 4.3 of Field and Ronchetti (1990) are fulfilled for $\chi_t(y) = \chi_0(y-t)$ and $R = Q'_{-\tau_n}, Q''_{\tau_n}$ as defined in (11.3.15)–(11.3.18), where we assume R to be absolutely continuous (i.e., $dR = r\,d\lambda$). Thus, we can read off an asymptotic expansion of the density of the corresponding M estimator which is

$$f_n(t) = \sqrt{\frac{n}{2\pi}}\,c^{-n}(t)\frac{A(t)}{\sigma(t)}\left[1 + \mathrm{O}(n^{-1})\right] \qquad t \in \mathbb{R} \tag{11.3.52}$$

where

$$c^{-1}(t) = \int \exp\{\alpha(t)\chi_t(y)\}\,r(y)\,dy \tag{11.3.53}$$

$$\sigma^2(t) = \int \chi_t(y)^2 c(t)\exp\{\alpha(t)\chi_t(y)\}\,r(y)\,dy \tag{11.3.54}$$

$$A(t) = \int \left[\tfrac{\partial}{\partial t}\chi_t(y)\right] c(t)\exp\{\alpha(t)\chi_t(y)\}\,r(y)\,dy \tag{11.3.55}$$

and $\alpha(t)$ is the solution $\alpha \in \mathbb{R}$ to

$$0 = \int \chi_t(y)\exp\{\alpha\chi_t(y)\}\,r(y)\,dy \tag{11.3.56}$$

11.3.3.3 Numerical Results

We consider the finite-sample risk of the asymptotic minimax estimator and check the quality of the introduced higher order approximations via the very accurate results obtained by our Algorithms A and B. The first table (Table 11.4) shows the results for contamination neighborhoods where we choose $\varepsilon = 0.1, 0.5$ and $\tau = \Phi^{-1}(0.95), \Phi^{-1}(0.995) \approx 1.645, 2.576$. For radius $\varepsilon = 0.1$ the saddlepoint approximation performs well down to sample size 4 or even 3, for radius $\varepsilon = 0.5$ the results get a little bit worse and we need about $5 - 10$ observations to obtain a good approximation. For the smaller τ the results of the Edgeworth expansion up to second order are comparable with the results of the saddlepoint approximation; i.e., for radius $\varepsilon = 0.1$ already 3 observations are enough to get a good approximation, whereas for radius $\varepsilon = 0.5$ we need about $5-10$ observations. For $\tau = 2.576$ and $\varepsilon = 0.1$ the Edgeworth expansion up to second order performs a little bit worse compared to the saddlepoint approximation, whereas for $\tau = 2.576$ and $\varepsilon = 0.5$ the results are again comparable. The Edgeworth expansion up to first order in all cases yields acceptable results down to sample size 10. But, we need 20 or even 50 observations to obtain results that are as good as the results obtained by the expansion up to second order and the saddlepoint approximation, respectively.
In case of total variation neighborhoods (cf. Table 11.5) the results are very similar. If we consider the saddlepoint approximation or the Edgeworth expansion up to second order, respectively, a sample size of about 5 is enough to obtain good approximation to the exact finite-sample risk. Again, the Edgeworth expansion up to first order performs a little bit worse and we need $20 - 50$ observations to get comparable good approximations as in the two other cases.

Remark 11.3.8 **(a)** Apparently, the speed of convergence towards the asymptotic risk is much faster in case of total variation neighborhoods. As Lemma 11.1.3 (c) and Lemma 11.2.3 (c) show $b_c^{\text{fi}} = b_c^{\text{as}} + \mathrm{O}(n^{-1/2})$ whereas $b_v^{\text{fi}} = b_v^{\text{as}} + \mathrm{O}(n^{-1})$. Moreover, our numerical calculations yield that the same holds for the corresponding finite-sample risks of the finite-sample minimax estimator; confer Subsubsections 11.4.1.2 and 11.4.2.2. However, this also seems to be true for the finite-sample risk of the asymptotic minimax estimator as the results included in Tables 11.4 and 11.5 as well as in Tables 11.6 and 11.9 indicate.

(b) On the one hand, we could use these higher order approximations to obtain a better approximation for the finite-sample risk of the asymptotic minimax estimator since the asymptotic risk of this estimator is much to optimistic for small sample sizes; confer Section 11.4. On the other hand, as the computation time of our Algorithm A and even more of our Algorithm B strongly increases with increasing sample size one could also think of saddlepoint approximations, respectively Edgeworth expansions to determine the finite-sample risk of the finite-sample minimax estimator especially for larger sample sizes ($n > 20$) since the computation time of these approximations is independent of n. In particular, the results for the Edgeworth expansions can be obtained with very little computational effort.

////

ε	τ	b_c^{as}	n	$\mathrm{Risk}^{\natural}(\tilde{S}_c^{\mathrm{as}};c)$	$\mathrm{Risk}_{\mathrm{EW1}}$	$\mathrm{Risk}_{\mathrm{EW2}}$	$\mathrm{Risk}_{\mathrm{SP}}$
0.1	1.645	1.484	2	15.56	13.37	13.80	14.61
			3	12.08	11.83	12.05	12.22
			4	11.24	10.97	11.14	11.17
			5	10.62	10.43	10.57	10.57
			10	9.36	9.26	9.35	9.34
			20	8.64	8.60	8.64	8.64
			50	8.12	8.10	8.12	8.12
			100	7.89	7.88	7.89	7.89
			1000	7.56	7.56	7.56	7.56
			asymptotic risk: 7.42				
	2.576	1.675	2	10.41	8.56	5.85	8.35
			3	4.75	5.48	4.44	5.16
			4	3.87	4.06	3.57	3.84
			5	3.14	3.30	3.02	3.14
			10	1.96	2.01	1.94	1.95
			20	1.46	1.48	1.45	1.46
			50	1.18	1.19	1.18	1.18
			100	1.08	1.08	1.08	1.08
			1000	0.95	0.95	0.95	0.95
			asymptotic risk: 0.91				
0.5	1.645	0.663	2	45.06	42.08	42.89	45.50
			3	37.10	37.19	38.07	39.36
			4	36.18	34.33	35.12	35.72
			5	32.74	32.41	33.10	33.37
			10	28.24	27.84	28.23	28.20
			20	25.03	24.84	25.04	25.03
			50	22.47	22.39	22.47	22.47
			100	21.28	21.25	21.28	21.28
			1000	19.49	19.49	19.49	19.49
			asymptotic risk: 18.74				
	2.576	0.919	2	39.83	34.92	36.50	40.17
			3	26.97	27.37	28.84	30.53
			4	25.90	22.88	24.01	24.76
			5	20.10	19.90	20.75	21.08
			10	13.42	13.14	13.41	13.40
			20	9.38	9.30	9.37	9.37
			50	6.75	6.73	6.74	6.74
			100	5.73	5.73	5.73	5.73
			1000	4.44	4.44	4.44	4.44
			asymptotic risk: 3.98				

Table 11.4: Approximation of the finite-sample risk via Edgeworth expansions and saddlepoint approximations in case of contamination neighborhoods. [The risks are given in percent. $\mathrm{Risk}^{\natural}(\tilde{S}_c^{\mathrm{as}};c)$ denotes the finite-sample risk of $\tilde{S}_c^{\mathrm{as}}$, $\mathrm{Risk}_{\mathrm{EW1}}$ and $\mathrm{Risk}_{\mathrm{EW2}}$ the approximations by means of Edgeworth expansions up to first, respectively second order and $\mathrm{Risk}_{\mathrm{SP}}$ the approximation via saddlepoint approximations.]

δ	τ	b_v^{as}	n	$\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{as}};v)$	$\mathrm{Risk}_{\mathrm{EW1}}$	$\mathrm{Risk}_{\mathrm{EW2}}$	$\mathrm{Risk}_{\mathrm{SP}}$
0.05	1.645	1.484	2	10.80	10.13	9.54	10.04
			3	9.11	9.06	8.92	8.97
			4	8.67	8.58	8.56	8.55
			5	8.40	8.31	8.34	8.32
			10	7.89	7.83	7.88	7.87
			20	7.65	7.62	7.65	7.65
			50	7.51	7.50	7.51	7.51
			100	7.47	7.46	7.47	7.47
			1000	7.42	7.42	7.42	7.42
			asymptotic risk: 7.42				
	2.576	1.675	2	5.61	4.67	3.72	4.32
			3	2.62	2.82	2.60	2.66
			4	2.08	2.15	2.03	2.04
			5	1.75	1.82	1.72	1.73
			10	1.25	1.29	1.24	1.24
			20	1.06	1.08	1.05	1.06
			50	0.96	0.97	0.96	0.96
			100	0.93	0.94	0.93	0.93
			1000	0.91	0.91	0.91	0.91
			asymptotic risk: 0.91				
0.25	1.645	0.663	2	28.87	24.52	26.82	27.30
			3	23.47	22.73	24.19	24.23
			4	23.19	21.78	22.83	22.74
			5	21.84	21.19	22.00	21.90
			10	20.35	19.98	20.36	20.32
			20	19.54	19.36	19.55	19.54
			50	19.06	18.99	19.06	19.06
			100	18.90	18.86	18.90	18.90
			1000	18.75	18.75	18.75	18.75
			asymptotic risk: 18.74				
	2.576	0.919	2	21.56	16.14	18.04	19.13
			3	11.88	12.35	12.79	13.29
			4	11.12	10.18	10.24	10.48
			5	8.80	8.83	8.79	8.92
			10	6.20	6.20	6.16	6.18
			20	5.01	5.01	5.00	5.01
			50	4.37	4.37	4.37	4.37
			100	4.17	4.17	4.17	4.17
			1000	4.00	4.00	4.00	4.00
			asymptotic risk: 3.98				

Table 11.5: Approximation of the finite-sample risk via Edgeworth expansions and saddlepoint approximations in case of total variation neighborhoods. [The risks are given in percent. $\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{as}};v)$ denotes the finite-sample risk of $\tilde{S}_v^{\mathrm{as}}$, $\mathrm{Risk}_{\mathrm{EW1}}$ and $\mathrm{Risk}_{\mathrm{EW2}}$ the approximation by means of Edgeworth expansions up to first, respectively second order and $\mathrm{Risk}_{\mathrm{SP}}$ the approximation via saddlepoint approximations.]

11.4 Numerical Comparisons

In this section we present various numerical comparisons between finite-sample and asymptotic results; i.e., we numerically check the asymptotics against finite-sample results obtained for fixed neighborhoods. To restrict the amount of results, we choose $\tau = \Phi^{-1}(0.95), \Phi^{-1}(0.975), \Phi^{-1}(0.995)$ such that 2τ corresponds to the width of 90%, 95%, 99%-confidence intervals in case of the standard normal distribution. Moreover, in most cases we use $q = 12$ in Algorithms A and B.

11.4.1 Contamination Neighborhoods

11.4.1.1 Optimal Clipping Bound

We fix radius $\varepsilon = 0.1, 0.5$ and compute b_c^{fi}, b_c^{as} and $b_c^{\text{as.c}}$ by (11.1.4), (11.1.11) and (11.1.17) for sample size $n \leq 25$; confer Figure 11.3. For large width τ the $\mathrm{O}(n^{-1/2})$-corrected is too small compared to the finite-sample optimal clipping bound (cf. also Table 11.6) whereas for not too large width τ it comes very close to the finite-sample one. In particular, if the radius is not too large.

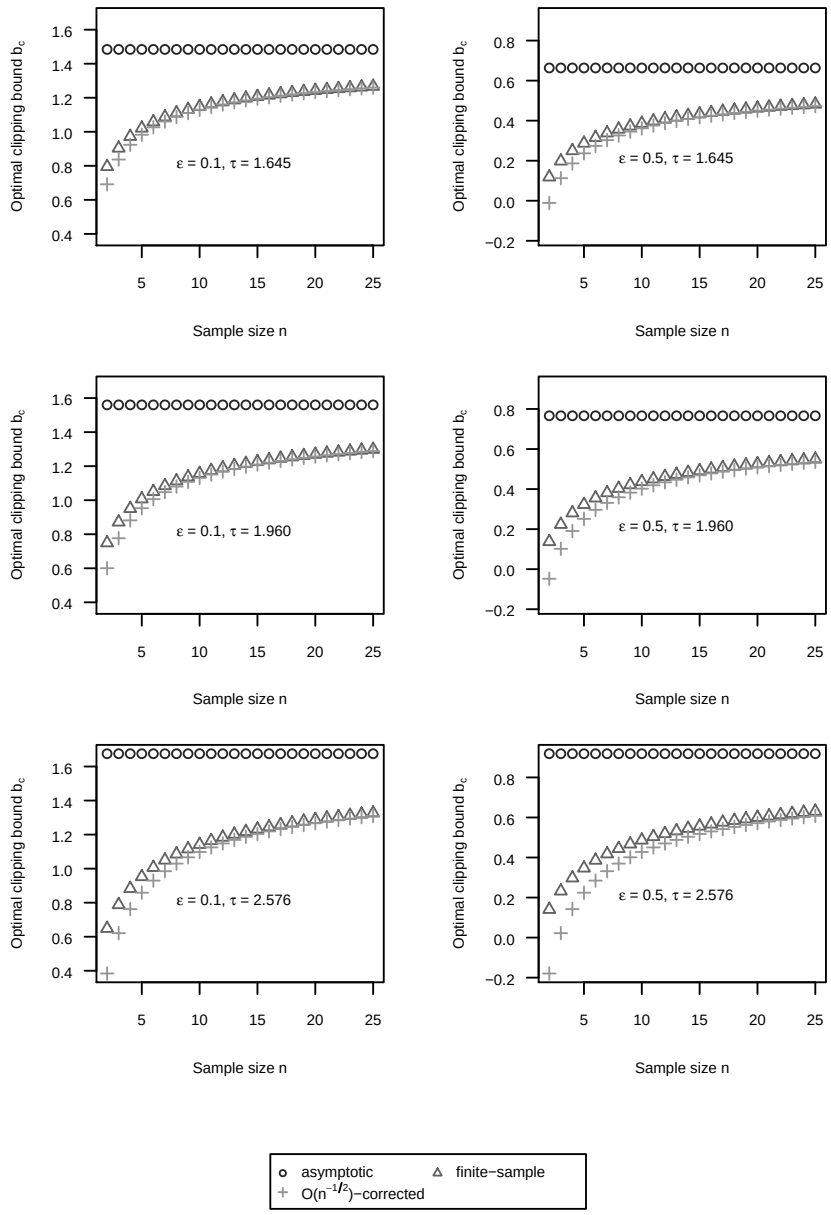

Figure 11.3: Optimal clipping bounds for sample size $n \leq 25$, radius $\varepsilon = 0.1, 0.5$ and width $\tau = 2.576, 1.960, 1.645$ in case of contamination neighborhoods $(* = c)$.

11.4.1.2 Finite-Sample Risk

We fix radius $\varepsilon = 0.1, 0.5$ and determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_c^{\text{fi}}$, the asymptotic minimax estimator $\tilde{S}_c^{\text{as}}$ and the estimator $S_c^{\text{as.c}}$ which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound. Although there are clear differences between the clipping bounds of the finite-sample and the asymptotic minimax estimator, the differences (in absolute values) concerning the corresponding finite-sample risks are only small; see Figure 11.4 and Table 11.6. Moreover, the finite-sample risk of the estimator which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is very close to the finite-sample risk of the finite-sample minimax estimator. Figure 11.5 shows the speed of convergence for the finite-sample risks. It seems to be of order $n^{-1/2}$ as in case of the optimal clipping bounds; confer Lemma 11.1.3. To get a better impression, we consider $y = \text{Risk}_{\text{as}}^{\natural} - \text{Risk}_{\text{fi}}^{\natural}$ and apply the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002); i.e., we estimate lambda by means of maximum likelihood such that $y^{\text{lambda}} \approx n$. That is, lambda ≈ -2 indicates $y = \mathrm{O}(n^{-1/2})$. Indeed, the estimated values of lambda are close to -2 which confirms our conjecture that we have a convergence of order $n^{-1/2}$; see Figure 11.6. Moreover, the speed seems to depend on τ and ε. A larger τ leads to a slightly faster convergence whereas a larger ε slightly decreases ($\tau = 1.645, 1.960$), respectively increases ($\tau = 2.576$) the speed of convergence.

So far, we compared absolute values for given radius ε and given width τ. But, now we study relative risks; i.e., how much efficiency do we lose if we take the asymptotic optimal clipping bound, respectively the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound instead of the finite-sample optimal. More precisely, given some estimator S, we consider

$$\text{relRisk}(S; c) = \frac{\text{Risk}(S; c)}{\text{Risk}(\tilde{S}_c^{\text{fi}}; c)} \tag{11.4.1}$$

and determine the radius ε_{as} and $\varepsilon_{\text{as.c}}$, respectively, such that the relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is maximal, respectively. To obtain valid results in case of the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound, we replace negative values of $b_c^{\text{as.c}}$ by 0. As the numerical results show (see Table 11.7), the maximum relative risks of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound are very small. Moreover, if we choose the width τ not too large and if we have a sample of moderate size (about $20-50$), there is no big difference between the finite-sample risks of all three estimators. These results are important if we know the radius.

But, in most applications the neighborhood radius ε is unknown ($\rho = 0$) or unknown except to belong to some radius interval ($\rho = 3, 2$), respectively. Thus, we also determine the least favorable radii for the finite-sample minimax estimator; confer Table 11.8. The corresponding radii are defined analogously to Section 2.2. Since the calculations of ε_3 and ε_2 are quite time consuming, we choose $q = 10$ in these two cases, for ε_0 we can use $q = 12$ as in the computations before. For $n < 10$ the least favorable radii and the corresponding relative risks behave different

for even and odd sample size, respectively. This is probably caused by the fact that there is mass at zero in case of an even sample size; confer Remark 11.3.3. However, this mass decays exponentially in n and therefore the effect disappears very fast for increasing sample size. Moreover, in all cases ε_0 is small compared to $\varepsilon_{\max}^{\mathrm{fi}}$ and also the relative risks for $\rho = 0$ are small if we have a moderate sample size $(20 - 50)$ and choose τ not too large. In most cases ε_3 and ε_2 are close to $\varepsilon_{\max}^{\mathrm{fi}}/3$ and $\varepsilon_{\max}^{\mathrm{fi}}/2$, respectively. But, the corresponding relative risks $(\rho = 3, 2)$ are small and in most cases stay well below 10% and 5%, respectively. Thus, in case that the true radius is completely unknown, respectively unknown except to belong to some radius interval, we strongly recommend to use the finite-sample minimax estimator which is given by the corresponding least favorable radius.

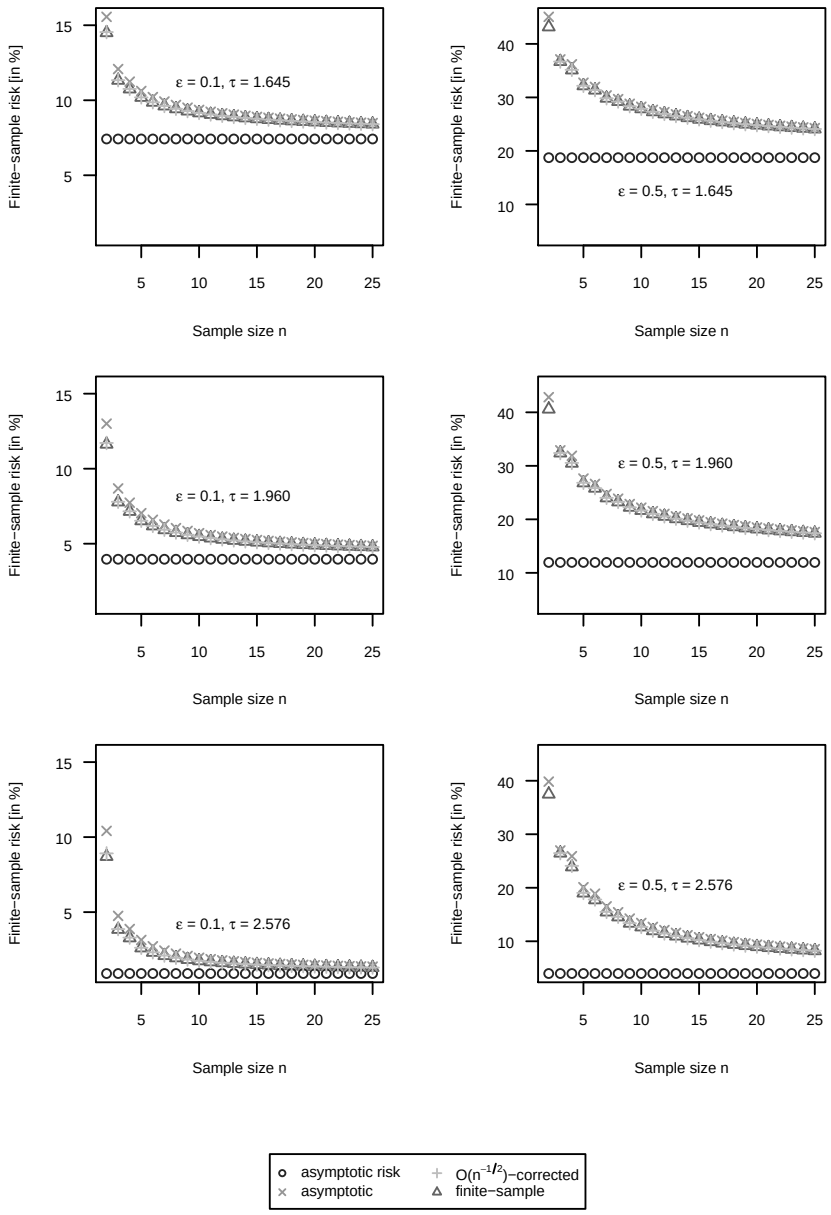

Figure 11.4: Finite-sample risk for sample size $n \leq 25$ given radius $\varepsilon = 0.1, 0.5$ and width $\tau = 2.576, 1.960, 1.645$ in case of contamination neighborhoods $(* = c)$.

ε	τ	n	b_c^{fi}	$b_c^{\mathrm{as.c}}$	b_c^{as}	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.1	1.645	2	0.795	0.691	1.484	14.502	14.546	15.562
		3	0.903	0.837		11.327	11.337	12.079
		4	0.972	0.924		10.749	10.754	11.240
		5	1.021	0.983		10.188	10.192	10.617
		10	1.148	1.130		9.156	9.156	9.357
		50	1.329	1.325		8.083	8.083	8.117
		100	1.374	1.372		7.872	7.872	7.888
	1.960	2	0.749	0.600	1.560	11.627	11.715	13.002
		3	0.871	0.777		7.777	7.794	8.682
		4	0.951	0.882		7.150	7.158	7.734
		5	1.007	0.953		6.523	6.529	7.035
		10	1.157	1.131		5.472	5.473	5.701
		50	1.373	1.368		4.498	4.498	4.533
		100	1.427	1.424		4.323	4.323	4.339
	2.576	2	0.649	0.384	1.675	8.706	8.916	10.415
		3	0.788	0.621		3.845	3.871	4.748
		4	0.883	0.762		3.280	3.299	3.873
		5	0.953	0.859		2.610	2.621	3.143
		10	1.142	1.098		1.751	1.753	1.958
		50	1.425	1.417		1.154	1.154	1.179
		100	1.497	1.493		1.067	1.067	1.078
0.5	1.645	2	0.118	−0.011	0.663	43.164	—	45.058
		3	0.198	0.113		36.695	36.699	37.100
		4	0.250	0.186		35.116	35.153	36.184
		5	0.287	0.237		32.161	32.167	32.741
		10	0.386	0.362		27.866	27.869	28.244
		50	0.533	0.528		22.378	22.378	22.468
		100	0.570	0.568		21.239	21.239	21.284
	1.960	2	0.137	−0.048	0.766	40.589	—	42.828
		3	0.223	0.102		32.385	32.392	32.892
		4	0.281	0.191		30.473	30.548	31.878
		5	0.323	0.251		26.812	26.825	27.616
		10	0.437	0.402		21.650	21.656	22.178
		50	0.610	0.603		15.564	15.564	15.682
		100	0.654	0.651		14.390	14.390	14.448
	2.576	2	0.140	−0.179	0.919	37.486	—	39.834
		3	0.232	0.022		26.461	26.468	26.966
		4	0.298	0.142		23.890	24.095	25.901
		5	0.347	0.224		19.006	19.033	20.099
		10	0.487	0.428		12.669	12.683	13.419
		50	0.710	0.699		6.606	6.606	6.745
		100	0.769	0.763		5.667	5.667	5.731

Table 11.6: Comparison of the optimal clipping bounds and the corresponding finite-sample risks in case of contamination neighborhoods ($* = c$). [The finite-sample risks are given in percent.]

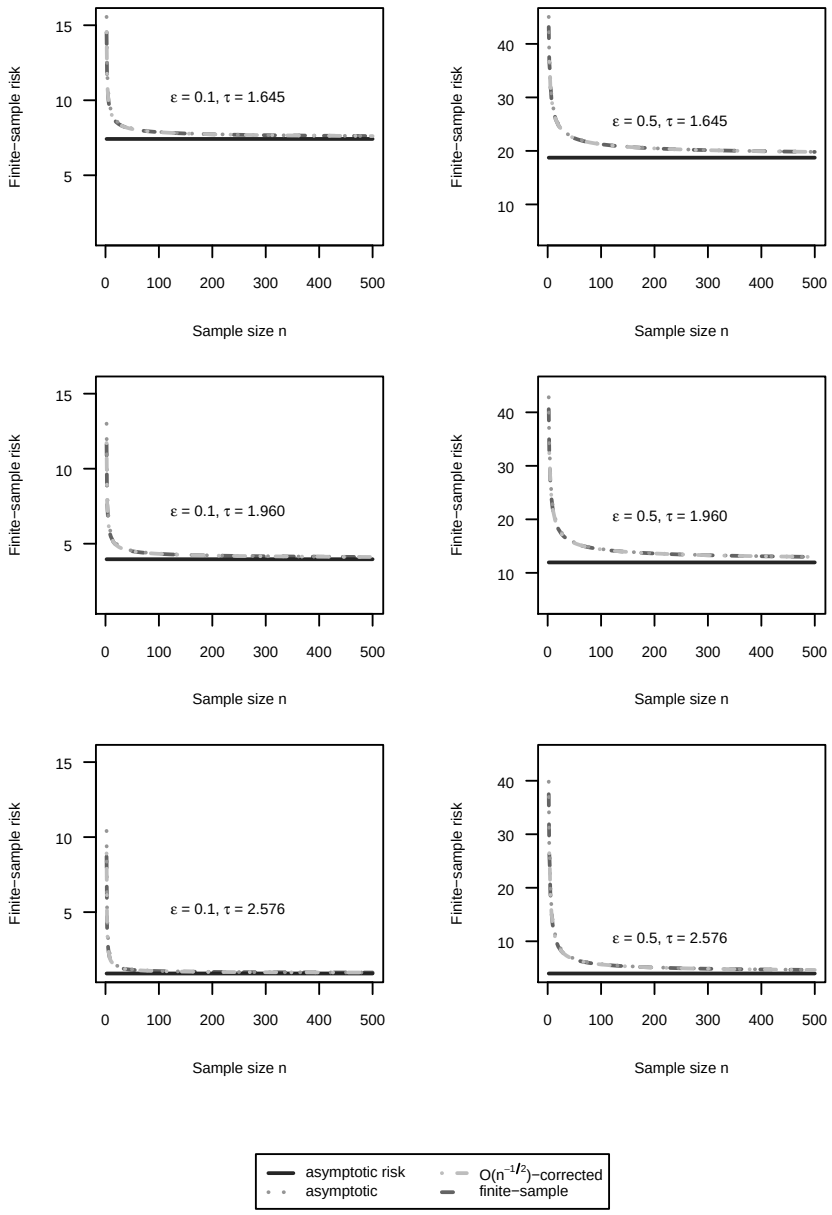

Figure 11.5: Finite-sample risk for increasing sample size n given radius $\varepsilon = 0.1, 0.5$ and width $\tau = 2.576, 1.960, 1.645$ in case of contamination neighborhoods ($* = c$).

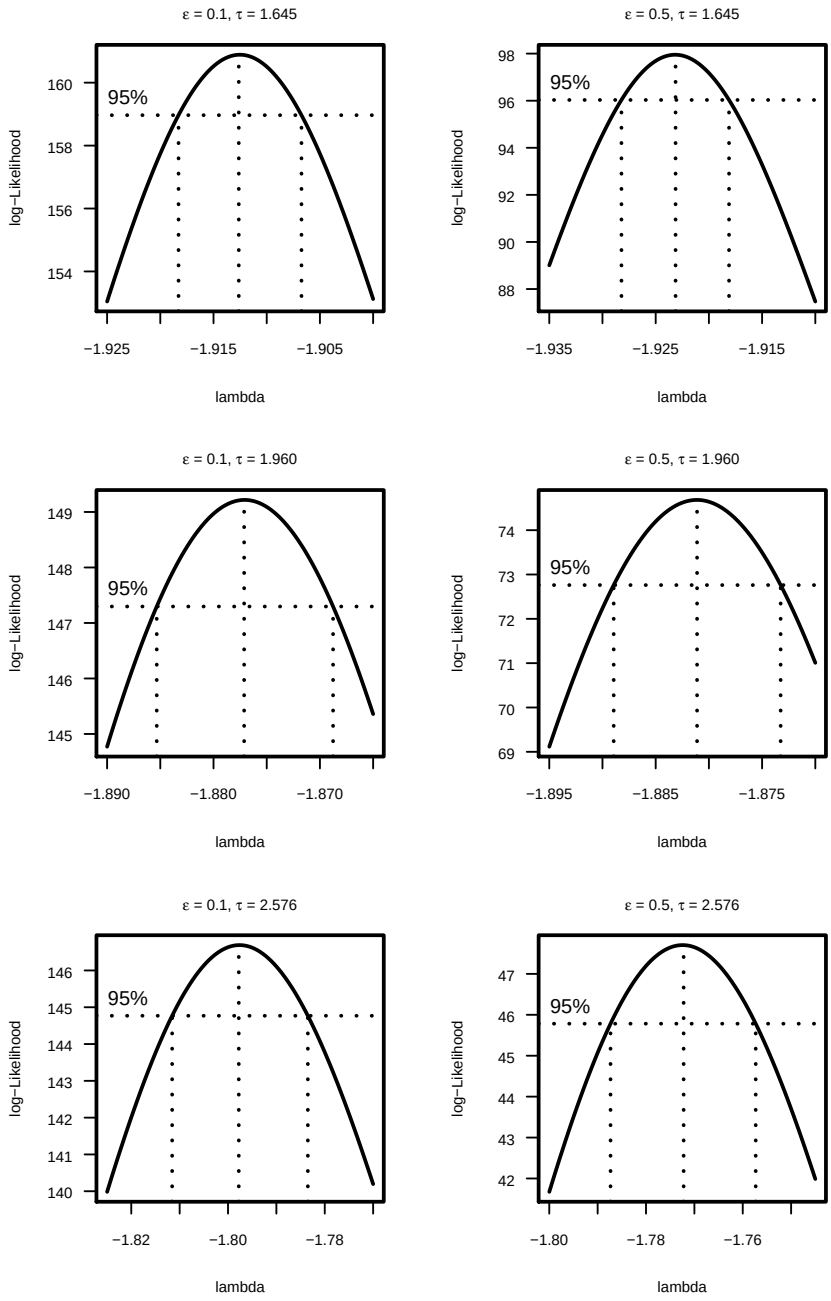

Figure 11.6: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator and contamination neighborhoods ($* = c$).

τ	n	ε_{as}	$\text{Risk}^{\natural}(\tilde{S}_c^{\text{fi}})$	$\text{relRisk}^{\natural}_{\text{as}}$	$\varepsilon_{\text{as.c}}$	$\text{Risk}^{\natural}(\tilde{S}_c^{\text{fi}})$	$\text{relRisk}^{\natural}_{\text{as.c}}$
1.645	2	0.131	17.04	1.074	0.248	26.06	1.004
	3	0.068	9.50	1.069	0.110	11.92	1.001
	4	0.062	8.67	1.048	0.502	35.25	1.001
	5	0.092	9.80	1.042	0.227	16.62	1.000
	10	0.125	10.14	1.022	0.325	18.99	1.000
	20	0.165	10.83	1.011	0.415	20.91	1.000
	50	0.215	11.69	1.005	0.459	20.69	1.000
	100	0.247	12.26	1.002	0.466	19.94	1.000
1.960	2	0.090	10.80	1.119	0.186	18.45	1.008
	3	0.052	5.41	1.131	0.083	6.93	1.002
	4	0.049	4.83	1.092	0.490	29.80	1.002
	5	0.076	5.56	1.080	0.203	10.95	1.001
	10	0.116	5.95	1.042	0.325	13.55	1.000
	20	0.164	6.53	1.022	0.446	15.97	1.000
	50	0.223	7.25	1.009	0.516	16.12	1.000
	100	0.263	7.74	1.004	0.537	15.61	1.000
2.576	2	0.033	3.43	1.233	0.075	6.80	1.024
	3	0.026	1.37	1.375	0.039	1.79	1.009
	4	0.026	1.16	1.263	0.390	17.18	1.009
	5	0.046	1.40	1.224	0.141	3.68	1.004
	10	0.093	1.66	1.118	1.029	41.40	1.001
	20	0.153	1.96	1.060	0.471	8.16	1.000
	50	0.232	2.36	1.024	0.599	8.90	1.000
	100	0.286	2.64	1.012	0.652	8.80	1.000

Table 11.7: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound in case of contamination neighborhoods ($* = c$). [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by ε_{as} and $\varepsilon_{\text{as.c}}$, respectively, and $\text{Risk}^{\natural}(\tilde{S}_c^{\text{fi}})$ is the finite-sample risk of $\tilde{S}_c^{\text{fi}}$ evaluated for the radius ε_{as} and $\varepsilon_{\text{as.c}}$, respectively. $\text{relRisk}^{\natural}_{\text{as}}$ and $\text{relRisk}^{\natural}_{\text{as.c}}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b_c^{\text{as.c}}$ by 0.]

τ	n	$\varepsilon_{\max}^{\text{fi}}$	ε_0	ε_3	ε_2	$\rho=0$	$\rho=3$	$\rho=2$
1.645	2	0.608	0.061	0.203	0.304	1.103	1.040	1.020
	3	0.687	0.033	0.078	0.093	1.069	1.027	1.012
	4	0.742	0.078	0.247	0.371	1.123	1.049	1.027
	5	0.782	0.068	0.157	0.207	1.107	1.038	1.018
	10	0.899	0.112	0.300	0.449	1.136	1.053	1.028
	11	0.913	0.118	0.270	0.383	1.139	1.054	1.026
	20	0.997	0.152	0.332	0.499	1.153	1.062	1.031
	50	1.099	0.195	0.366	0.549	1.168	1.072	1.036
	100	1.156	0.221	0.385	0.578	1.176	1.077	1.039
	∞	1.312	0.293	0.437	0.656	1.193	1.089	1.045
1.960	2	0.643	0.036	0.152	0.239	1.110	1.035	1.016
	3	0.738	0.019	0.055	0.065	1.070	1.029	1.013
	4	0.804	0.056	0.268	0.402	1.144	1.052	1.028
	5	0.855	0.049	0.135	0.177	1.121	1.043	1.020
	10	1.003	0.096	0.334	0.502	1.162	1.060	1.031
	11	1.022	0.102	0.266	0.381	1.166	1.062	1.030
	20	1.132	0.142	0.326	0.463	1.187	1.072	1.035
	50	1.267	0.196	0.397	0.575	1.210	1.087	1.043
	100	1.345	0.228	0.437	0.633	1.221	1.094	1.047
	∞	1.564	0.324	0.521	0.781	1.249	1.113	1.057
2.576	2	0.682	0.009	0.037	0.040	1.109	1.025	1.011
	3	0.802	0.004	0.023	0.026	1.060	1.033	1.015
	4	0.890	0.024	0.247	0.414	1.177	1.049	1.024
	5	0.959	0.020	0.086	0.189	1.134	1.049	1.022
	10	1.167	0.062	0.194	0.263	1.206	1.070	1.032
	11	1.194	0.068	0.234	0.337	1.212	1.076	1.035
	20	1.356	0.112	0.300	0.430	1.250	1.094	1.045
	50	1.566	0.181	0.404	0.589	1.292	1.119	1.058
	100	1.689	0.227	0.458	0.666	1.315	1.132	1.065
	∞	2.055	0.370	0.612	0.881	1.371	1.166	1.083

Table 11.8: Least favorable radii for the finite-sample minimax estimator in case of contamination neighborhoods ($* = c$). [The least favorable radii $\varepsilon_0, \varepsilon_3, \varepsilon_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\varepsilon_{\max}^{\text{fi}}$ denotes the upper bound on the radius given by the disjointness condition (11.1.5). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

11.4.1.3 Finite-Sample Distribution

We use Algorithm A (cf. Subsubsection 11.3.2.1) to compute the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance.
To restrict the amount of data, we treat only the two "extreme" situations $\varepsilon = 0.1$, $\tau = 1.645$ (see Figure 11.7) and $\varepsilon = 0.5$, $\tau = 2.576$ (see Figure 11.8).

Remark 11.4.1 If H''_{τ_n} is absolutely continuous, then also Q''_{τ_n} is absolutely continuous and by Remark 11.3.2 the distribution of $\tilde{S}_c^{\text{fi}}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^{n} \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. Moreover, by symmetry it suffices to consider the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$. ////

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^{n} \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ and of $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the `R` function `optim`; confer R Development Core Team (2005).
As we see, in both cases about 10 observations are enough to get already quite close to a normal distribution (i.e., $d_\kappa < 0.02$ for $n \geq 10$) where the jumps included in the cumulative distribution functions decay exponentially in n; confer also Remark 11.3.3.

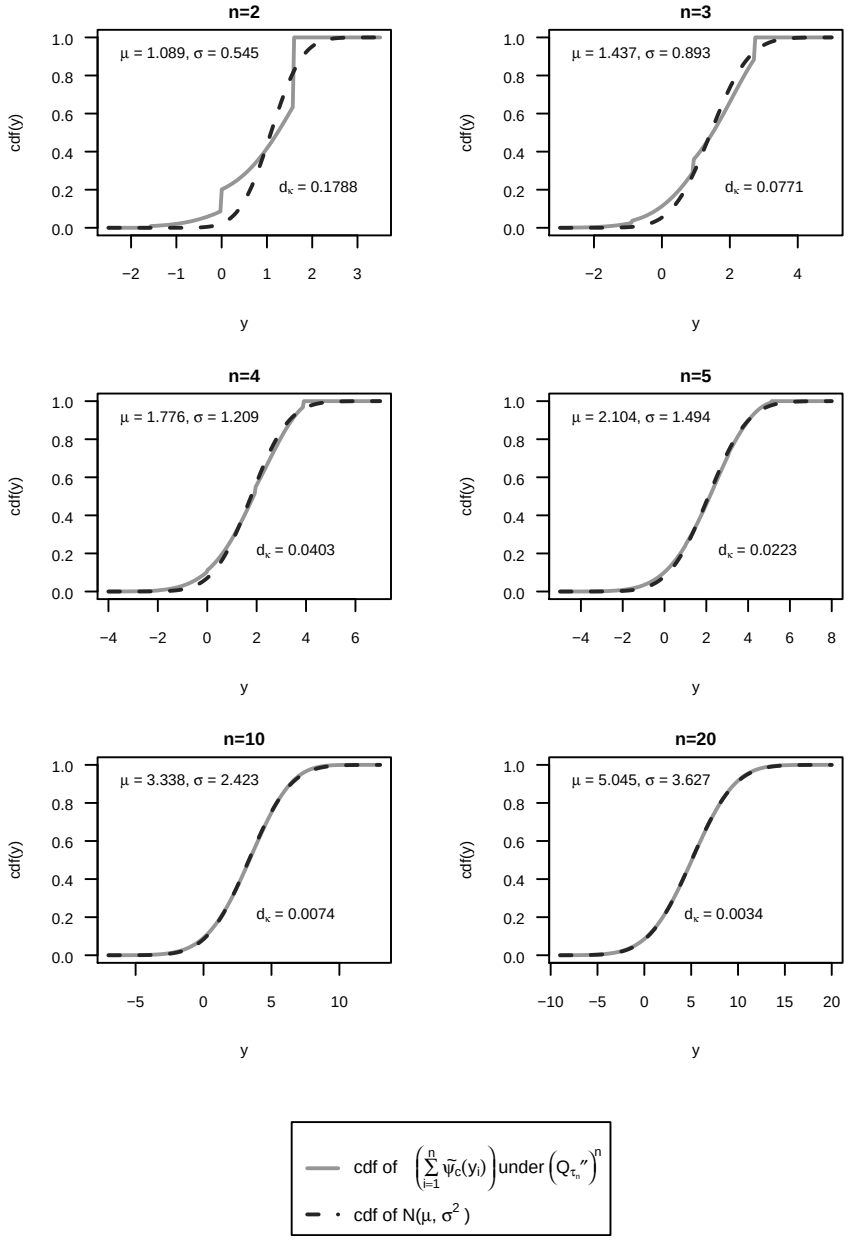

Figure 11.7: Finite-sample distributions for radius $\varepsilon = 0.1$ and $\tau = 1.645$ in case of contamination neighborhoods ($* = c$).

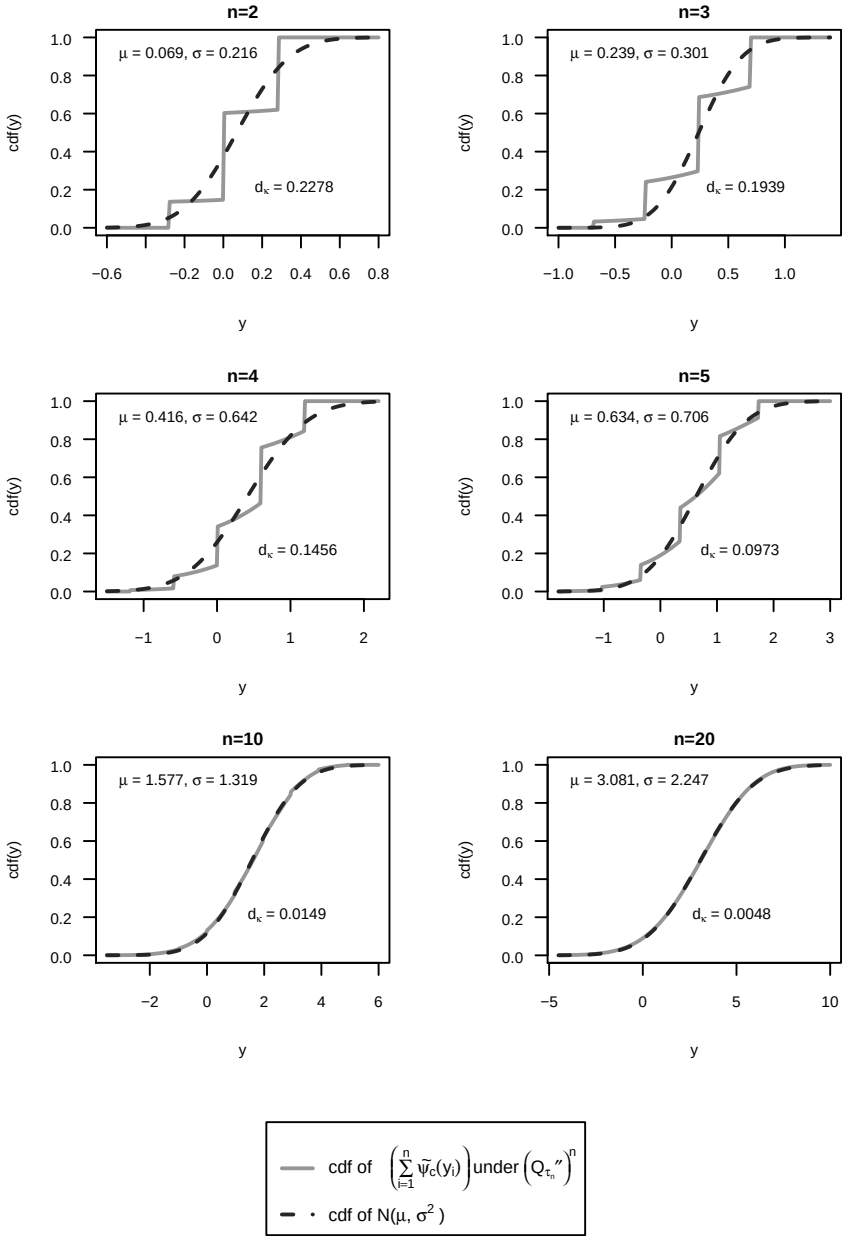

Figure 11.8: Finite-sample distributions for radius $\varepsilon = 0.5$ and $\tau = 2.576$ in case of contamination neighborhoods ($* = c$).

11.4.2 Total Variation Neighborhoods

11.4.2.1 Optimal Clipping Bound

We fix radius $\delta = 0.05, 0.25$. That is, $\delta = \varepsilon/2$, which by (11.1.11) and (11.2.11) leads to $b_c^{\mathrm{as}} = b_v^{\mathrm{as}}$. Thus, we have well comparable results for contamination and total variation neighborhoods. We compute b_v^{fi}, b_v^{as} and $b_v^{\mathrm{as.c}}$ by (11.2.4), (11.2.11) and (11.2.17). Figure 11.9 shows the results for sample size $n \leq 25$. The $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is too small compared with the finite-sample optimal clipping bound; confer also Table 11.9. But, already for moderate sample sizes n between 10 and 20 they are very close to each other where the approximation is worse for larger values of τ.

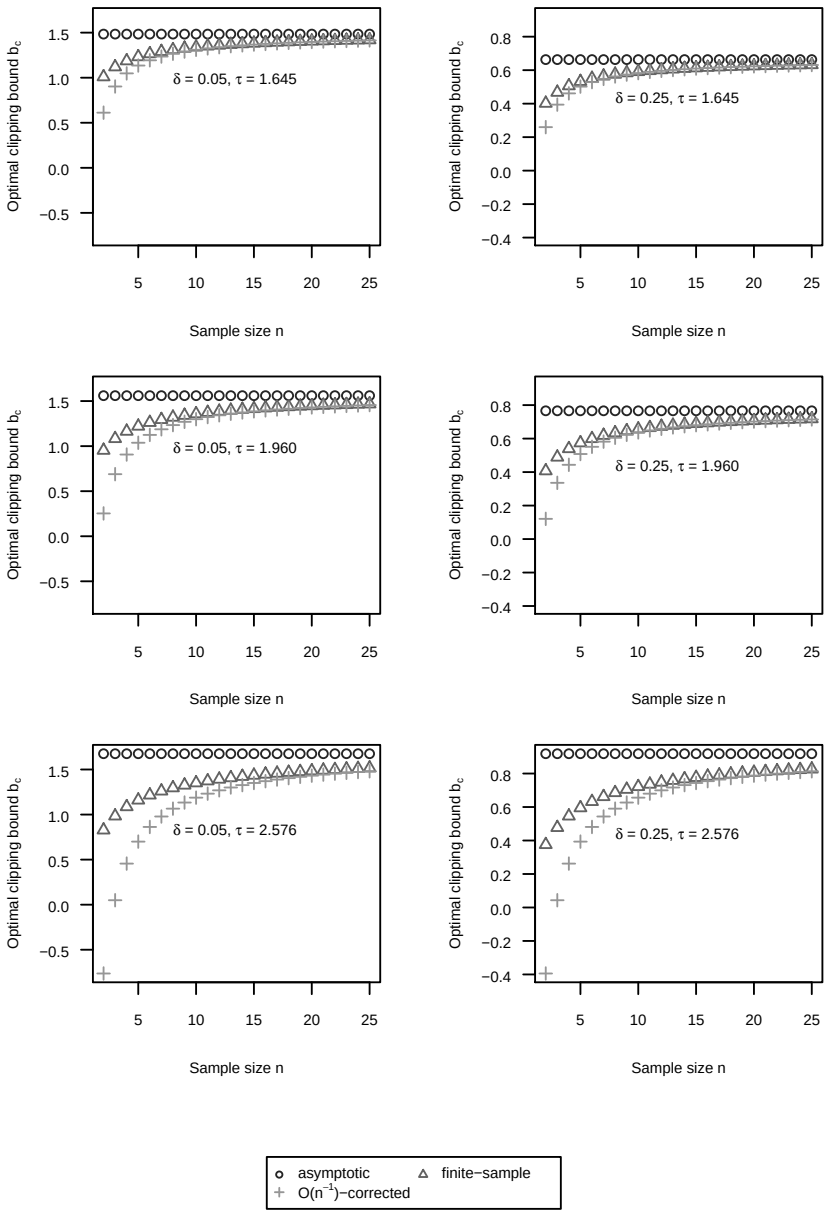

Figure 11.9: Optimal clipping bounds for sample sizes $n \leq 25$, radius $\delta = 0.05, 0.25$ and width $\tau = 2.576, 1.960, 1.645$ in case of total variation neighborhoods $(* = v)$.

11.4.2.2 Finite-Sample Risk

We fix radius $\delta = 0.05, 0.25$ and determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_v^{\text{fi}}$, the asymptotic minimax estimator $\tilde{S}_v^{\text{as}}$ and the estimator $S_v^{\text{as.c}}$ which is based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound. The differences between the finite-sample risks (in absolute values) are small already for very small sample sizes ($n \leq 5$); see Figure 11.10 and Table 11.9. In particular, the finite-sample risks of $S_v^{\text{as.c}}$ is very close to the finite-sample risk of $\tilde{S}_v^{\text{fi}}$, already for sample size $n = 5$. Moreover, the results suggest that not only in case of the optimal clipping bound (cf. Lemma 11.2.3) but also in case of the corresponding finite-sample risk of the finite-sample minimax estimator the speed of convergence is of order $\mathrm{O}(n^{-1})$. The second figure (Figure 11.11) and the results obtained by the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002) (see Figure 11.11) strongly confirm this conjecture. Moreover, the results are almost independent from τ and δ.

Second, we study relative risks; i.e., how much efficiency do we lose if we take the asymptotic optimal clipping bound, respectively the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound instead of the finite-sample optimal. We determine the radius δ_{as} and $\delta_{\text{as.c}}$, respectively, such that the relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is maximal, respectively. In case of the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound we replace negative values by 0 to obtain valid results. As the numerical results show (see Table 11.10), the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound has a quite large relative risk for very small sample sizes ($n \leq 5$) and large values of τ where the maxima are attained for very small radii δ. However, for moderate sample sizes n (between 10 and 20) there is already almost no difference between the finite-sample risk of the finite-sample minimax estimator and the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound. In case of the asymptotic minimax estimator, the maximum relative risks are small ($\tau = 2.576$), respectively very small ($\tau = 1.645, 1.960$) already at sample size 20. That is, in case of total variation neighborhoods and for moderate sample size n it almost does not matter which estimator we choose provided we know δ.

But, in most applications the neighborhood radius δ is unknown ($\rho = 0$) or unknown except to belong to some radius interval ($\rho = 3, 2$), respectively. Thus, we again determine the least favorable radii for the finite-sample minimax estimator; confer Table 11.11. The corresponding radii are defined analogously to Section 2.2. As in case of contamination neighborhoods (cf. Subsubsection 11.4.1.2), the least favorable radii and the corresponding relative risks behave different for small even and odd sample sizes ($n < 10$), respectively. This is probably caused by the fact that there is mass at zero in case of an even sample size; confer Remark 11.3.3. Moreover, the faster convergence towards the asymptotic values in case of total variation neighborhoods is also reflected by the least favorable radii. Since the asymptotic values are larger in case of total variation neighborhoods, we get larger relative risks for finite sample sizes. But nevertheless, in all cases δ_0 is small compared to $\delta_{\max}^{\text{fi}}$ and in most cases δ_3 and δ_2 are close to $\delta_{\max}^{\text{fi}}/3$ and $\delta_{\max}^{\text{fi}}/2$,

respectively. In addition, the corresponding relative risks ($\rho = 3, 2$) are small and in most cases stay well below 10% and 5%, respectively. Thus, as in case of contamination neighborhoods we strongly recommend to use the finite-sample minimax estimator which is given by the corresponding least favorable radius if the true radius is completely unknown, respectively unknown except to belong to some radius interval.

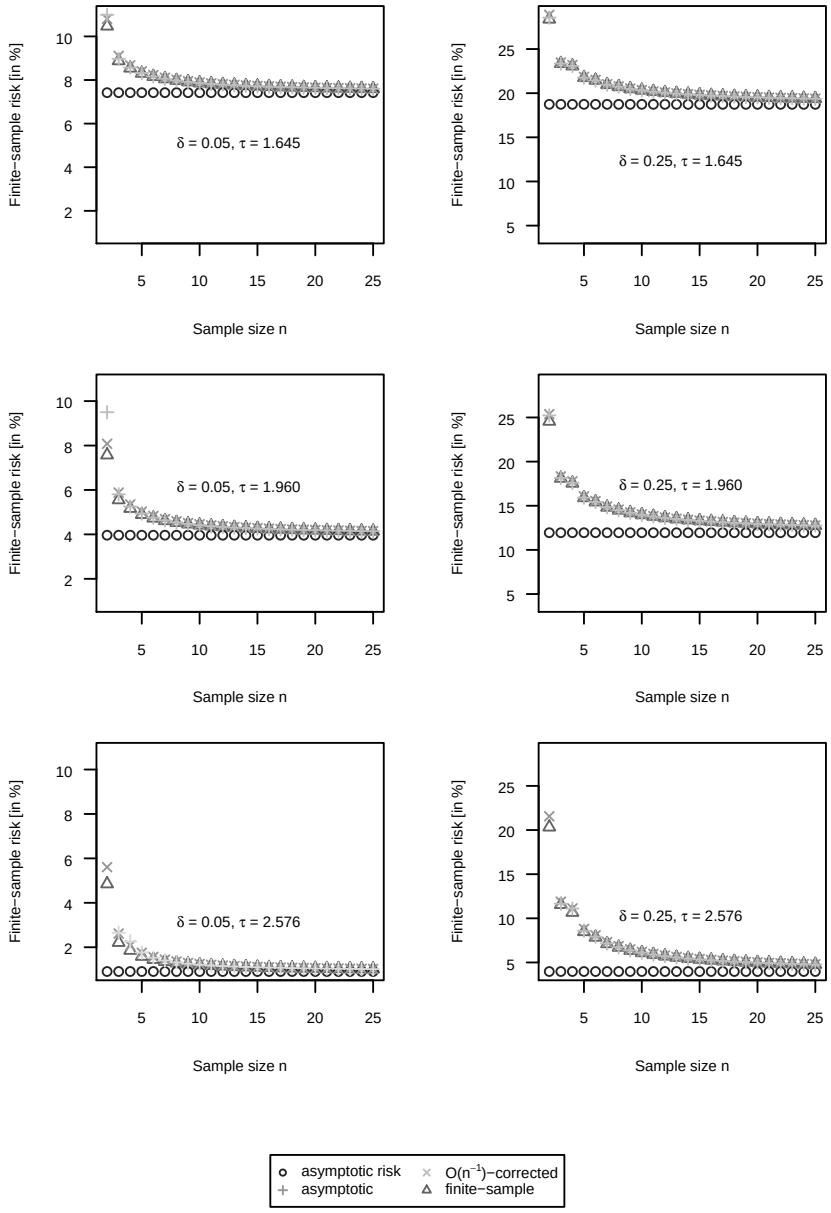

Figure 11.10: Finite-sample risk for sample size $n \leq 25$ given radius $\delta = 0.05, 0.25$ and width $\tau = 2.576, 1.960, 1.645$ in case of total variation neighborhoods $(* = v)$.

δ	τ	n	b_v^{fi}	$b_v^{\mathrm{as.c}}$	b_v^{as}	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.05	1.645	2	1.009	0.612	1.484	10.471	10.989	10.804
		3	1.120	0.903		8.879	8.972	9.106
		4	1.187	1.048		8.543	8.574	8.672
		5	1.232	1.135		8.306	8.321	8.399
		10	1.340	1.310		7.863	7.864	7.891
		50	1.451	1.449		7.510	7.510	7.511
		100	1.467	1.466		7.465	7.465	7.465
	1.960	2	0.953	0.253	1.560	7.574	9.501	8.079
		3	1.083	0.689		5.555	5.803	5.882
		4	1.165	0.907		5.167	5.254	5.351
		5	1.221	1.037		4.893	4.939	5.028
		10	1.359	1.299		4.414	4.417	4.453
		50	1.511	1.508		4.052	4.052	4.054
		100	1.535	1.534		4.008	4.008	4.009
	2.576	2	0.828	−0.764	1.675	4.851	—	5.606
		3	0.984	0.049		2.214	2.654	2.618
		4	1.086	0.456		1.859	2.264	2.080
		5	1.160	0.700		1.588	1.761	1.754
		10	1.351	1.188		1.201	1.215	1.247
		50	1.588	1.578		0.961	0.961	0.963
		100	1.630	1.627		0.934	0.934	0.935
0.25	1.645	2	0.402	0.260	0.663	28.379	28.554	28.874
		3	0.467	0.394		23.347	23.360	23.471
		4	0.506	0.461		23.069	23.081	23.190
		5	0.532	0.502		21.762	21.766	21.843
		10	0.591	0.583		20.326	20.326	20.350
		50	0.647	0.647		19.058	19.058	19.059
		100	0.655	0.655		18.898	18.898	18.898
	1.960	2	0.407	0.121	0.766	24.563	25.231	25.372
		3	0.489	0.336		18.106	18.151	18.324
		4	0.540	0.443		17.545	17.597	17.768
		5	0.574	0.508		15.903	15.921	16.061
		10	0.657	0.637		13.975	13.977	14.025
		50	0.741	0.740		12.352	12.352	12.354
		100	0.753	0.753		12.150	12.150	12.151
	2.576	2	0.375	−0.394	0.919	20.329	—	21.555
		3	0.477	0.043		11.559	11.658	11.882
		4	0.546	0.262		10.668	11.108	11.120
		5	0.595	0.393		8.484	8.589	8.796
		10	0.722	0.656		6.099	6.110	6.203
		50	0.870	0.866		4.364	4.364	4.370
		100	0.893	0.892		4.168	4.168	4.170

Table 11.9: Comparison of the optimal clipping bounds and the corresponding finite-sample risks in case of total variation neighborhoods ($* = v$). [The finite-sample risks are given in percent.]

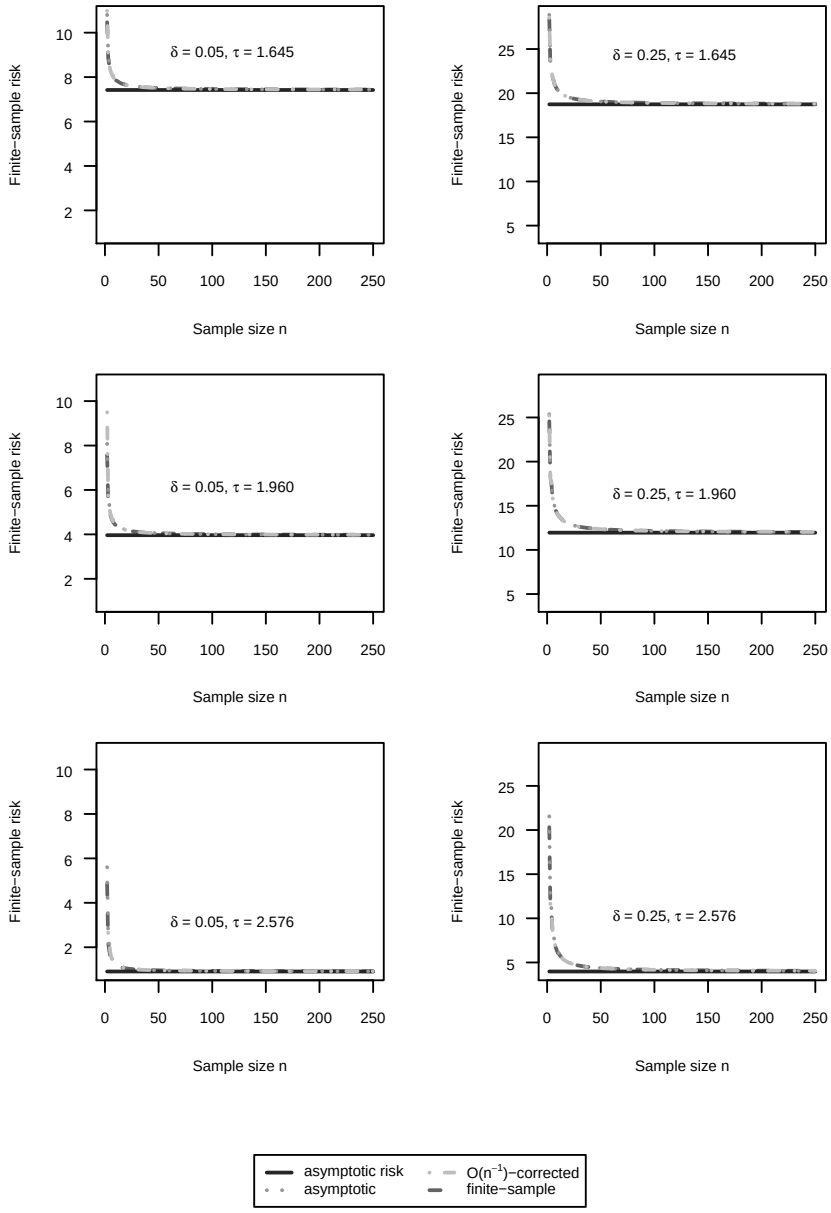

Figure 11.11: Finite-sample risk for increasing sample size n given radius $\delta = 0.05, 0.25$ and width $\tau = 2.576, 1.960, 1.645$ in case of total variation neighborhoods ($* = v$).

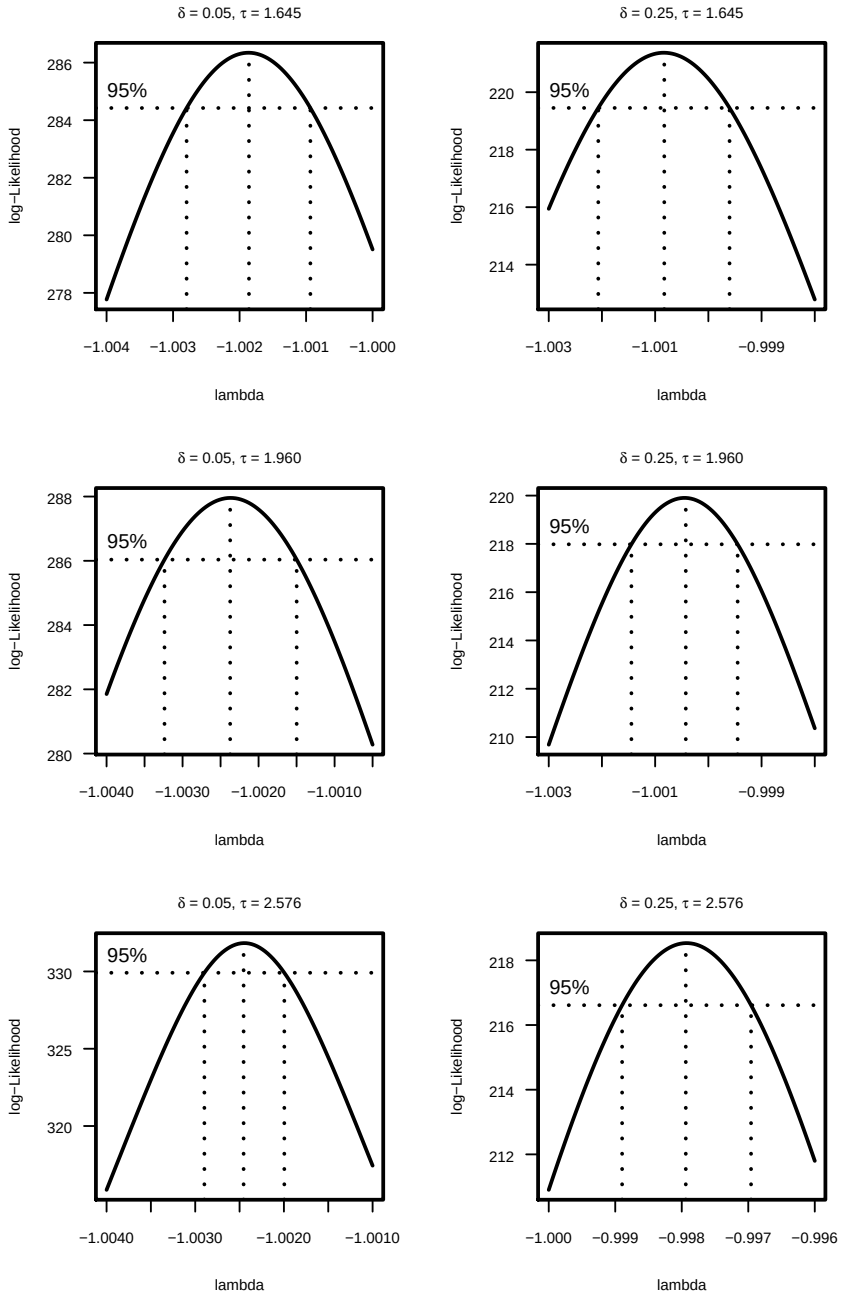

Figure 11.12: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator and total variation neighborhoods ($* = v$).

τ	n	δ_{as}	$\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{fi}})$	$\mathrm{relRisk}^{\natural}_{\mathrm{as}}$	$\delta_{\mathrm{as.c}}$	$\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{fi}})$	$\mathrm{relRisk}^{\natural}_{\mathrm{as.c}}$
1.645	2	0.068	12.28	1.033	0.008	5.96	1.086
	3	0.038	8.06	1.026	0.008	5.75	1.020
	4	0.033	7.38	1.016	0.008	5.68	1.007
	5	0.037	7.53	1.011	0.009	5.71	1.003
	10	0.040	7.33	1.004	0.010	5.66	1.000
	20	0.042	7.23	1.001	0.010	5.62	1.000
	50	0.043	7.15	1.000	0.012	5.66	1.000
	100	0.043	7.12	1.000	0.016	5.82	1.000
1.960	2	0.054	7.91	1.067	0.000	2.52	2.047
	3	0.033	4.59	1.062	0.004	2.80	1.118
	4	0.029	4.12	1.038	0.005	2.85	1.036
	5	0.036	4.26	1.028	0.007	2.93	1.016
	10	0.043	4.16	1.009	0.010	2.95	1.001
	20	0.047	4.09	1.003	0.012	2.93	1.000
	50	0.049	4.03	1.000	0.013	2.92	1.000
	100	0.050	4.01	1.000	0.016	3.00	1.000
2.576	2	0.023	2.59	1.171	0.000	0.50	6.855
	3	0.019	1.18	1.231	0.000	0.50	2.686
	4	0.019	1.01	1.148	0.001	0.53	1.614
	5	0.027	1.09	1.111	0.003	0.58	1.244
	10	0.044	1.11	1.039	0.009	0.64	1.017
	20	0.054	1.10	1.012	0.013	0.64	1.001
	50	0.061	1.08	1.002	0.015	0.64	1.000
	100	0.063	1.06	1.001	0.017	0.64	1.000

Table 11.10: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound in case of total variation neighborhoods ($* = v$). [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by $\varepsilon_{\mathrm{as}}$ and $\varepsilon_{\mathrm{as.c}}$, respectively, and $\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{fi}})$ is the finite-sample risk of $\tilde{S}_v^{\mathrm{fi}}$ evaluated for the radius δ_{as} and $\delta_{\mathrm{as.c}}$, respectively. $\mathrm{relRisk}^{\natural}_{\mathrm{as}}$ and $\mathrm{relRisk}^{\natural}_{\mathrm{as.c}}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b_v^{\mathrm{as.c}}$ by 0.]

τ	n	$\delta^{\mathrm{fi}}_{\max}$	δ_0	δ_3	δ_2	$\rho=0$	$\rho=3$	$\rho=2$
1.645	2	0.534	0.076	0.178	0.267	1.142	1.059	1.030
	3	0.570	0.054	0.107	0.118	1.114	1.033	1.015
	4	0.589	0.086	0.196	0.295	1.148	1.063	1.036
	5	0.602	0.086	0.184	0.228	1.148	1.053	1.024
	10	0.628	0.113	0.209	0.314	1.170	1.070	1.035
	11	0.630	0.115	0.210	0.307	1.172	1.072	1.035
	20	0.642	0.128	0.214	0.321	1.181	1.079	1.039
	50	0.650	0.139	0.217	0.325	1.188	1.085	1.043
	100	0.653	0.143	0.218	0.327	1.191	1.087	1.044
	∞	0.656	0.147	0.219	0.328	1.193	1.089	1.045
1.960	2	0.590	0.051	0.197	0.295	1.161	1.052	1.024
	3	0.643	0.036	0.072	0.080	1.127	1.033	1.015
	4	0.673	0.068	0.224	0.336	1.178	1.069	1.038
	5	0.692	0.072	0.178	0.220	1.178	1.056	1.025
	10	0.735	0.108	0.245	0.367	1.212	1.081	1.040
	11	0.739	0.112	0.242	0.341	1.215	1.083	1.040
	20	0.758	0.132	0.253	0.379	1.229	1.095	1.047
	50	0.772	0.149	0.257	0.386	1.241	1.105	1.052
	100	0.777	0.156	0.259	0.388	1.245	1.109	1.054
	∞	0.782	0.162	0.261	0.390	1.249	1.113	1.057
2.576	2	0.659	0.015	0.220	0.329	1.171	1.029	1.012
	3	0.747	0.011	0.026	0.028	1.128	1.035	1.015
	4	0.802	0.033	0.267	0.401	1.226	1.068	1.035
	5	0.839	0.039	0.118	0.142	1.224	1.057	1.025
	10	0.924	0.084	0.308	0.462	1.290	1.094	1.046
	11	0.933	0.090	0.258	0.372	1.297	1.100	1.047
	20	0.974	0.125	0.281	0.400	1.328	1.123	1.059
	50	1.005	0.158	0.297	0.429	1.353	1.147	1.072
	100	1.016	0.171	0.302	0.435	1.362	1.156	1.077
	∞	1.028	0.185	0.306	0.440	1.371	1.166	1.083

Table 11.11: Least favorable radii for the finite-sample minimax estimator in case of total variation neighborhoods ($* = v$). [The least favorable radii $\delta_0, \delta_3, \delta_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\delta^{\mathrm{fi}}_{\max}$ denotes the upper bound on the radius given by the disjointness condition (11.2.5). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

11.4.2.3 Finite-Sample Distribution

We use Algorithm A (cf. Subsubsection 11.3.2.1) to compute the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_v(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance d_κ. By symmetry, we only consider Q''_{τ_n} and again give only two "extreme" situations $\delta = 0.05$, $\tau = 1.645$ (see Figure 11.13) and $\delta = 0.25$, $\tau = 2.576$ (see Figure 11.14) analogously to the case of contamination neighborhoods (cf. Subsubsection 11.4.1.3). Moreover, if Q''_{τ_n} is absolutely continuous, the distribution of $\tilde{S}_v^{\text{fi}}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^{n} \tilde{\psi}_v(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide; confer Remark 11.4.1.

In the first case already 5 observations seem to be enough to get quite close to a normal distribution (i.e, $d_\kappa < 0.02$ for $n \geq 5$) whereas in the second case we need a sample size of about 10 (i.e., $d_\kappa < 0.02$ for $n \geq 10$). Moreover, the results indicate that the speed of convergence in case of total variation neighborhoods is not only faster at zero (corresponds to the finite-sample risk) but uniformly over the whole support of the law of $\sum_{i=1}^{n} \tilde{\psi}_v(y_i)$ under $\left(Q''_{\tau_n}\right)^n$. Hence, this could also be true if we consider other finite-sample risks like MSE.

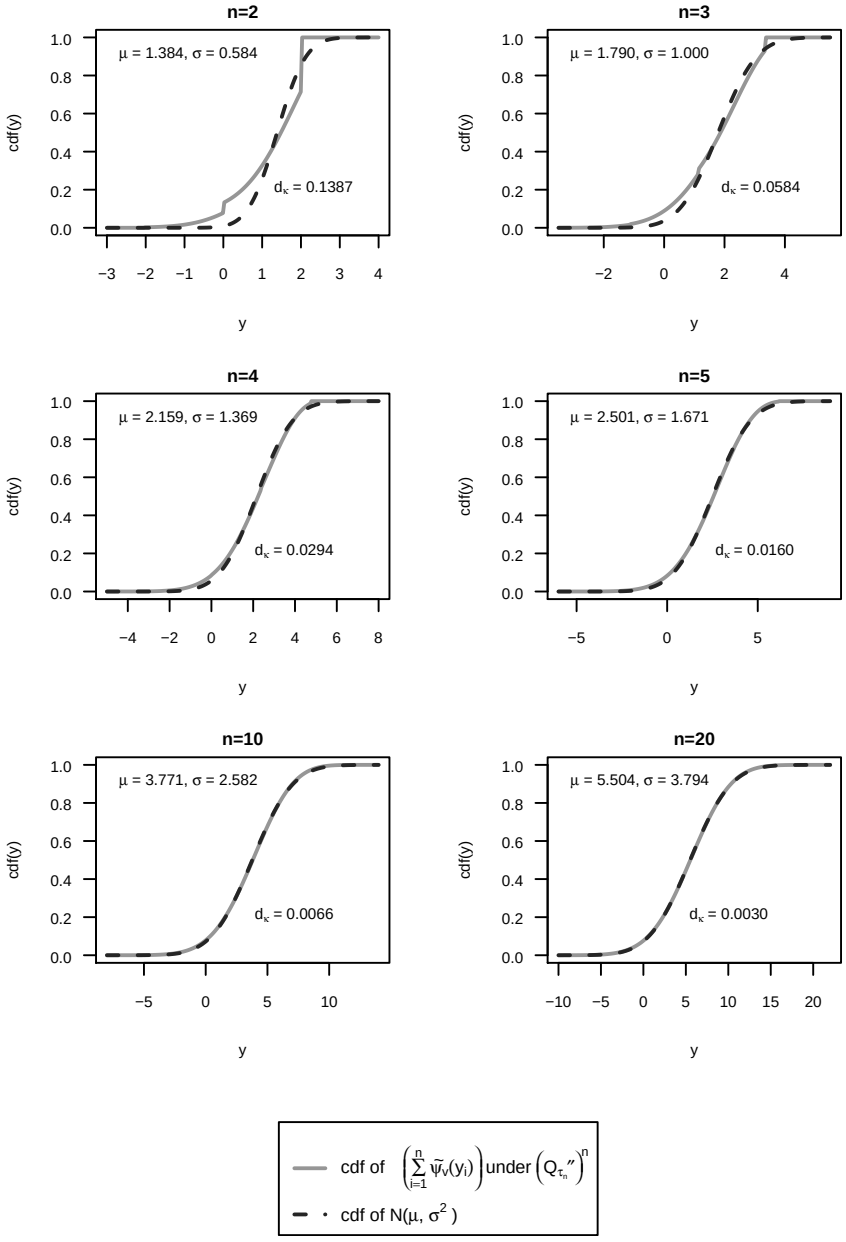

Figure 11.13: Finite-sample distributions for radius $\delta = 0.05$ and $\tau = 1.645$ in case of total variation neighborhoods ($* = v$).

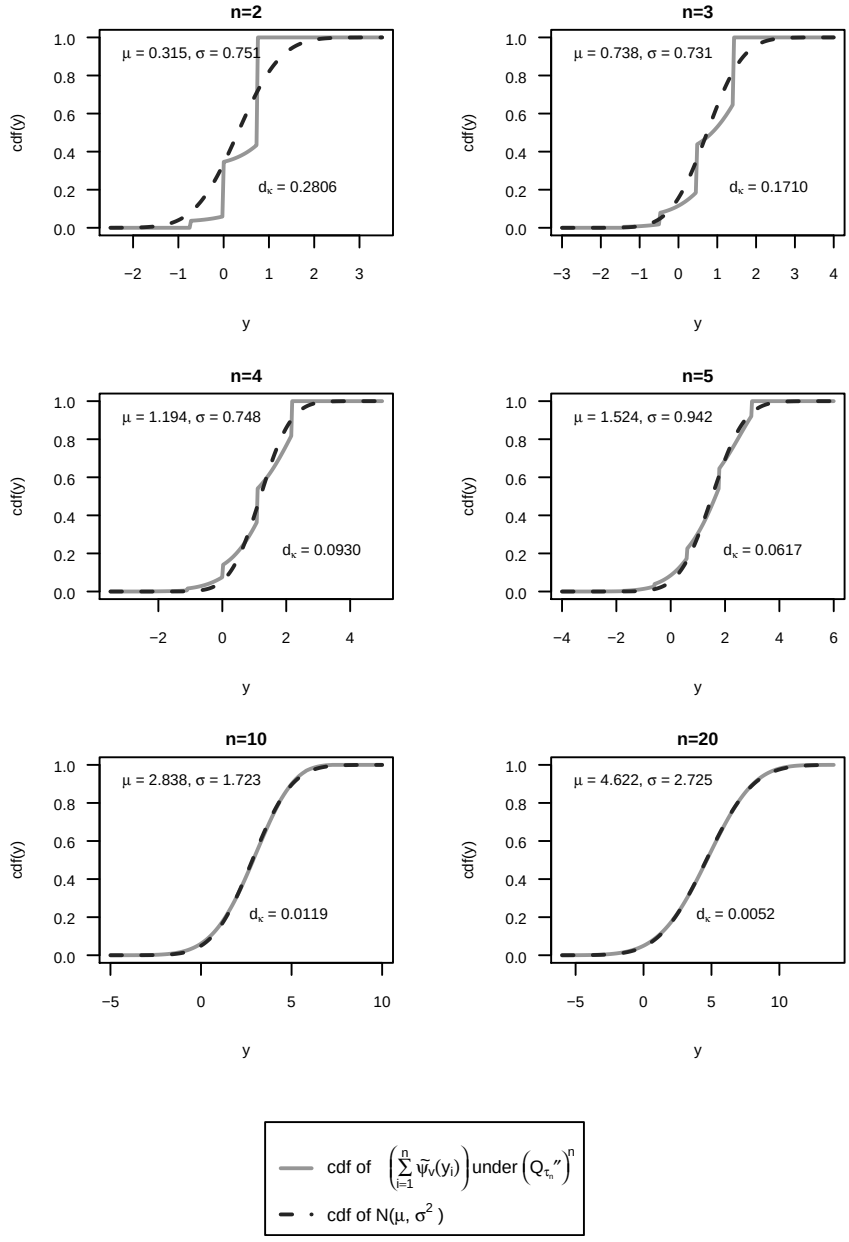

Figure 11.14: Finite-sample distributions for radius $\delta = 0.25$ and $\tau = 2.576$ in case of total variation neighborhoods ($* = v$).

11.5 Implementation Using R

The computation of the optimal ICs with respect to the finite-sample as well as asymptotic under-/overshoot confidence risk can be done via our `R` package `ROptEst` which is part of our `R` bundle `RobASt` and is in detail described in Appendix D.3. First, we instantiate a normal location family by

```
> N <- NormLocationFamily(mean = 0)
```

In addition to the `S4` class `InfRobModel` we provide the `S4` class `FixRobModel` which combines some parametric family with a fixed unconditional neighborhood. Now, given some sample of size n, some width $\tau \in (0, \infty)$ and some neighborhood radius $r \in (0, \infty)$ we can define the following robust models

```
> Rob1 <- InfRobModel(center = N,
+                     neighbor = ContNeighborhood(radius = r))
> Rob2 <- InfRobModel(center = N,
+                     neighbor=TotalVarNeighborhood(radius = r))
```

respectively

```
> Rob3 <-  FixRobModel(center = N,
+            neighbor = ContNeighborhood(radius = r/sqrt(n)))
> Rob4 <-  FixRobModel(center = N,
+            neighbor = TotalVarNeighborhood(radius = r/sqrt(n)))
```

where we use $r/\sqrt{n}$ in case of the robust models with fixed neighborhoods to get comparable results; confer Remark 10.1.4 (b). To compute the asymptotic optimal IC, we can proceed as follows

```
> IC1 <- optIC(model = Rob1, risk = asUnOvShoot(width = τ))
> IC2 <- optIC(model = Rob2, risk = asUnOvShoot(width = τ))
```

Remark 11.5.1 **(a)** In case of the asymptotic under-/overshoot confidence risk, the algorithm works for any one-dimensional L_2 differentiable parametric family.

(b) Since the solution in case of contamination neighborhoods is of total variation form (i.e., no centering constant but possibly asymmetric clipping), `IC1` is of class `TotalVarIC`. ////

The finite-sample optimal IC can be determined via

```
> IC3 <- optIC(model = Rob3,
+              risk = fiUnOvShoot(width = τ/sqrt(n)),
+              sampleSize = n)
> IC4 <- optIC(model = Rob4,
+              risk = fiUnOvShoot(width = τ/sqrt(n)),
+              sampleSize = n)
```

We use width $\tau/\sqrt{n}$ to get comparable results. In case of the finite-sample under-/overshoot confidence risk one can also specify whether Algorithm A (cf. Subsubsection 11.3.2.1) or Algorithm B (cf. Subsubsection 11.3.2.2) should be used in the call to `optIC` by setting the additional parameter `Algo` to `"A"` or `"B"`, respectively. By default the faster Algorithm A is called. In case of the finite-sample under-/overshoot confidence risk it is also possible to distinguish between contamination on the left or right hand side by setting the additional parameter `cont` to `"left"` or `"right"`, respectively. For details on this we refer to Subsection 11.3.1.

Remark 11.5.2 (a) In contrast to the asymptotic risk, the solution in case of the finite-sample risk is only implemented for normal location.

(b) As in case of the asymptotic risk, the solution in case of contamination neighborhoods is of total variation form; i.e., `IC3` is of class `TotalVarIC`. ////

To compute the $\mathrm{O}(n^{-1/2})$-corrected optimal IC ($* = c$), we start with `IC1` and modify the slots `clipUp`, `clipLo` and `stand` using equations (11.1.17) and (11.1.9). This can be done as follows

```
> IC5 <- IC1
> clipUp1 <- clipUp(IC1)/as.vector(stand(IC1))
> clipUp5 <- r*(r + clipUp1*τ)/(sqrt(n)*2*τ*pnorm(-clipUp1))
> clipUp5 <- max(0, clipUp1 - clipUp5)
> stand5 <- 1/(2*pnorm(clipUp5)-1)
> clipUp(IC5) <- stand5*clipUp5
> clipLo(IC5) <- -clipUp(IC5)
> stand(IC5) <- as.matrix(stand5)
```

Similarly, we can determine the $\mathrm{O}(n^{-1})$-corrected optimal IC ($* = v$) via

```
> IC6 <- IC2
> clipUp2 <- clipUp(IC2)/as.vector(stand(IC2))
> clipUp6 <- τ*(2*clipUp2^2*r + τ*dnorm(clipUp2))
> clipUp6 <- max(0, clipUp2 - clipUp6/(6*n*pnorm(-clipUp2)))
> stand6 <- 1/(2*pnorm(clipUp6)-1)
> clipUp(IC6) <- stand6*clip6
> clipLo(IC6) <- -clipUp(IC6)
> stand(IC6) <- as.matrix(stand6)
```

by using equations (11.2.17) and (11.2.9).

Remark 11.5.3 (a) Given some contaminated sample one can compute the finite-sample, respectively asymptotic minimax estimator using `locMEstimator`, respectively `ksEstimator` and `oneStepEstimator`. For more details we refer to Sections 3.5 and 5.5, respectively.

(b) After the installation of our `R` bundle `RobASt` one can find the `R` script `UnderOverShootRisk.R`, which contains some examples for the computation of the optimal ICs and the corresponding finite-sample, respectively asymptotic minimax estimator, in the directory "`.../RHome/library/ROptEst/scripts/`" where `RHome` stands for the local home directory of `R`. ////

Chapter 12

One-Dimensional Normal Linear Regression

We specialize the results of Chapter 10 to one-dimensional normal linear regression. That is, the ideal model distributions read

$$P_\theta(dx, dy) = \varphi(y - x\theta)\lambda(dy)K(dx) \qquad \theta \in \mathbb{R} \tag{12.0.1}$$

where we consider unconditional (cf. Section 12.1) and conditional (cf. Section 12.2) neighborhoods of contamination (i.e., $\delta = 0$ resp. $\delta(x) \equiv 0$) as well as total variation neighborhoods (i.e, $\varepsilon = 0$ resp. $\varepsilon(x) \equiv 0$)). In Section 12.3 we present various numerical results which enable us to check the asymptotics against finite-sample results. Finally, we explain how one can use our `R` package `ROptRegTS` to compute the finite-sample and asymptotic minimax estimator.

12.1 Unconditional neighborhoods

In this section we state the results for unconditional, or errors-in-variables, neighborhoods as given in Definition 10.1.3 (a).

12.1.1 Contamination Neighborhoods

We assume $\delta = 0$. That is, we consider the following unconditional contamination neighborhoods,

$$\mathcal{U}_{c,0}(\theta, \varepsilon_n) = \left\{ Q \in \mathcal{M}_1(\mathbb{B}^2) \,\middle|\, Q = (1 - \varepsilon_n)P_\theta + \varepsilon_n H, \text{ any } H \in \mathcal{M}_1(\mathbb{B}^2) \right\}$$

for given radius $\varepsilon_n = \varepsilon/\sqrt{n} \in (0, 1)$ and fixed sample size $n \in \mathbb{N}$.

12.1.1.1 Finite-Sample Minimax Estimator

For a given width $\tau_n = \tau/\sqrt{n} \in (0,\infty)$ a least favorable pair $(R'_{-\tau_n,c,0}, R''_{\tau_n,c,0})$ for $\mathcal{U}_{c,0}(-\tau_n,\varepsilon_n)$ versus $\mathcal{U}_{c,0}(\tau_n,\varepsilon_n)$ by Subsubsection 10.2.2.1 is

$$R'_{-\tau_n,c,0}(dx,dy) = (1-\varepsilon_n)\varphi(y+\tau_n x)\max\{1,\exp\big(2\tau_n(xy-b^{\text{fi}}_{c,0})\big)\}\lambda(dy)K(dx) \tag{12.1.1}$$

$$R''_{\tau_n,c,0}(dx,dy) = (1-\varepsilon_n)\varphi(y-\tau_n x)\max\{1,\exp\big(-2\tau_n(xy+b^{\text{fi}}_{c,0})\big)\}\lambda(dy)K(dx) \tag{12.1.2}$$

where $b^{\text{fi}}_{c,0}$ is the unique solution to

$$\frac{\varepsilon_n}{1-\varepsilon_n} = \int \big[\exp(-2b^{\text{fi}}_{c,0}\tau_n)\Phi\big(\tau_n|x| - b^{\text{fi}}_{c,0}/|x|\big) - \Phi\big(-\tau_n|x| - b^{\text{fi}}_{c,0}/|x|\big)\big]\,K(dx) \tag{12.1.3}$$

and the disjointness condition (10.2.18) reads

$$\varepsilon_n < 1 - \big[2\textstyle\int \Phi\big(\tau_n|x|\big)\,K(dx)\big]^{-1} \tag{12.1.4}$$

This immediately implies $\varepsilon_n < 0.5$ for all $\tau_n \in (0,\infty)$. The minimax robust test $\tilde{\phi}_{c,0}$ for $\mathcal{U}_{c,0}(-\tau_n)$ versus $\mathcal{U}_{c,0}(\tau_n)$, by Subsection 10.2.3 and symmetry, is

$$\tilde{\phi}_{c,0} = \tfrac{1}{2}\,\mathrm{I}(h_{c,0} > 0) + \tfrac{1}{2}\,\mathrm{I}(h_{c,0} \geq 0) \tag{12.1.5}$$

where $h_{c,0}$ is the loglikelihood of $(R'_{-\tau,c,0})^n$ with respect to $(R''_{\tau,c,0})^n$ as given in (10.2.23). Thus, the minimax estimator $\tilde{S}^{\text{fi}}_{c,0}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\psi}_{c,0}(x_i, y_i - x_i\tilde{S}^{\text{fi}}_{c,0}) = 0 \qquad \tilde{\psi}_{c,0}(x,u) = xu\min\left\{1, \frac{b^{\text{fi}}_{c,0}}{|xu|}\right\} \tag{12.1.6}$$

with equal randomization between the smallest and the largest solutions. The corresponding IC reads

$$\mathrm{IC}_{\tilde{\psi}_{c,0}}(x,u) = A^{\text{fi}}_{c,0}\tilde{\psi}_{c,0}(x,u) \qquad (A^{\text{fi}}_{c,0})^{-1} = 2\int x^2\Phi\big(b^{\text{fi}}_{c,0}/|x|\big)\,K(dx) - \mathcal{K} \tag{12.1.7}$$

where $\mathcal{K} = \int x^2\,K(dx)$.

12.1.1.2 Asymptotic Minimax Estimator

The IC of the asymptotic minimax estimator in case of unconditional contamination neighborhoods is given by Theorem 10.3.7 (a). By symmetry it has the following form,

$$\tilde{\eta}_{c,0}(x,u) = A^{\text{as}}_{c,0}xu\min\left\{1, \frac{b^{\text{as}}_{c,0}}{|xu|}\right\} \tag{12.1.8}$$

with

$$(A^{\text{as}}_{c,0})^{-1} = 2\int x^2\Phi\big(b^{\text{as}}_{c,0}/|x|\big)\,K(dx) - \mathcal{K} \tag{12.1.9}$$

where $\mathcal{K} = \int x^2\, K(dx)$ and $b^{\mathrm{as}}_{c,0}$ is the unique solution to

$$\varepsilon = 2\tau \int \left[|x| \varphi\big(b^{\mathrm{as}}_{c,0}/|x|\big) - b^{\mathrm{as}}_{c,0} \Phi\big(-b^{\mathrm{as}}_{c,0}/|x|\big) \right] K(dx) \tag{12.1.10}$$

and the boundedness restriction (10.3.39) reads

$$\varepsilon < \sqrt{\tfrac{2}{\pi}}\, \tau \int |x|\, K(dx) \tag{12.1.11}$$

However, it is not clear whether the corresponding asymptotic minimax estimator $\tilde{S}^{\mathrm{as}}_{c,0}$ can be constructed as an M estimator. Nevertheless, we can define the M estimator $S^{\mathrm{as}}_{c,0}$ satisfying

$$\sum_{i=1}^{n} \tilde{\eta}_{c,0}(x_i, y_i - x_i S^{\mathrm{as}}_{c,0}) = 0 \tag{12.1.12}$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2 (cf. Theorem 10.2.9 (a)) and serves as a comparative estimator for the finite-sample minimax estimator $\tilde{S}^{\mathrm{fi}}_{c,0}$. Moreover, we get the following relations between finite-sample and asymptotic results.

Lemma 12.1.1 *For $\varepsilon_n \in (0,1)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

(a)

$$\sqrt{n}\left(1 - \left[2 \textstyle\int \Phi\big(\tau_n |x|\big)\, K(dx)\right]^{-1}\right) < \sqrt{\tfrac{2}{\pi}}\, \tau \textstyle\int |x|\, K(dx) \tag{12.1.13}$$

(b)

$$\begin{aligned} &\frac{\varepsilon}{1-\varepsilon_n} - \sqrt{n} \int \left[\exp(-2b\tau_n) \Phi\big(\tau_n|x| - b/|x|\big) - \Phi\big(-\tau_n|x| - b/|x|\big) \right] K(dx) \\ &\qquad = \varepsilon - 2\tau \int \left[|x| \varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big) \right] K(dx) + \mathrm{o}(n^0) \end{aligned} \tag{12.1.14}$$

where we obtain a remainder $\mathrm{O}(n^{-1/2})$ instead of $\mathrm{o}(n^0)$ if $\mathrm{E}\,|x|^3 < \infty$.

(c)

$$b^{\mathrm{fi}}_{c,0} = b^{\mathrm{as}}_{c,0} + \mathrm{o}(n^0) \tag{12.1.15}$$

If $\mathrm{E}\,|x|^3 < \infty$, we obtain a remainder $\mathrm{O}(n^{-1/2})$ instead of $\mathrm{o}(n^0)$ and the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is

$$b^{\mathrm{as.c}}_{c,0} := b^{\mathrm{as}}_{c,0} - \frac{1}{\sqrt{n}} \frac{\varepsilon(\varepsilon + b^{\mathrm{as}}_{c,0}\tau)}{2\tau \int \Phi\big(-b^{\mathrm{as}}_{c,0}/|x|\big)\, K(dx)} \tag{12.1.16}$$

where

$$b^{\mathrm{fi}}_{c,0} = b^{\mathrm{as.c}}_{c,0} + \mathrm{o}(n^{-1/2}) \tag{12.1.17}$$

Moreover, the remainder $\mathrm{o}(n^{-1/2})$ can be replaced by $\mathrm{O}(n^{-1})$ if $\mathrm{E}\,|x|^5 < \infty$.

PROOF

(a)(12.1.13) is equivalent to

$$2\int \Phi\big(\tau_n|x|\big)\,K(dx)\Big(1-\sqrt{\tfrac{2}{\pi}}\,\tau_n\int |x|\,K(dx)\Big)<1 \tag{12.1.18}$$

As $x\mapsto\Phi(\tau_n|x|)$ is integrable for every $\tau_n\in(0,\infty)$ and $\tau_n\mapsto\Phi(\tau_n|x|)$ is differentiable for every $x\in\mathbb{R}$ with derivative $|x|\varphi(\tau_n|x|)\le|x|$ where $\int|x|\,K(dx)<\infty$ by (10.1.6), the assumptions of Lemma 16.2 of Bauer (1990) are fulfilled and we can differentiate under the integral sign; i.e.,

$$\frac{\partial}{\partial\tau_n}\int\Phi(\tau_n|x|)\,K(dx)=\int|x|\varphi(\tau_n|x|)K(dx) \tag{12.1.19}$$

Thus, the derivative of the left hand side of (12.1.18) reads,

$$\begin{aligned}2\int|x|\varphi(\tau_n|x|)K(dx)\Big(1-\sqrt{\tfrac{2}{\pi}}\,\tau_n\int|x|\,K(dx)\Big)-\\ -2\int\Phi(\tau_n|x|)\,K(dx)\sqrt{\tfrac{2}{\pi}}\int|x|\,K(dx)\end{aligned} \tag{12.1.20}$$

where we have

$$2\int|x|\varphi(\tau_n|x|)K(dx)\Big(1-\sqrt{\tfrac{2}{\pi}}\,\tau_n\int|x|\,K(dx)\Big)<2\int|x|\varphi(\tau_n|x|)K(dx) \tag{12.1.21}$$

$$<\sqrt{\tfrac{2}{\pi}}\int|x|\,K(dx) \tag{12.1.22}$$

and

$$2\int\Phi(\tau_n|x|)\,K(dx)\sqrt{\tfrac{2}{\pi}}\int|x|\,K(dx)>\sqrt{\tfrac{2}{\pi}}\int|x|\,K(dx) \tag{12.1.23}$$

as $K\big(\varphi(\tau_n|x|)>\varphi(0)\big)=1=K\big(\Phi(\tau_n|x|)>\frac12\big)$ by assumption (10.1.5), which is $K(x=0)=0$. That is, the left hand side of (12.1.18) is strictly decreasing in $\tau_n\in(0,\infty)$ and becomes maximal for $\tau_n\downarrow0$ where $2\int\Phi(0)K(dx)=1$.

(b) We first rewrite the left hand side of (12.1.3),

$$\frac{\varepsilon_n}{1-\varepsilon_n}=\varepsilon_n\sum_{k=0}^{\infty}\varepsilon_n^k=\varepsilon_n+\mathrm{O}(n^{-1}) \tag{12.1.24}$$

Defining $t:=1/\sqrt{n}$, the right hand side of (12.1.3) reads

$$\int\big[\exp(-2b\tau t)\Phi\big(\tau|x|t-b/|x|\big)-\Phi\big(-\tau|x|t-b/|x|\big)\big]\,K(dx)=:F(t) \tag{12.1.25}$$

As we integrate a smooth (in t) and bounded ($\le$ const) function whose derivative (in t) is dominated ($\le|x|+$const), the assumptions of Lemma 16.2 of Bauer (1990)

are easy to verify analogously to part (a) of this proof; i.e., we can differentiate under the integral sign which gives

$$F'(t) = \tau \int \Big\{ \exp(-2b\tau t)\big[|x|\varphi\big(\tau|x|t - b/|x|\big) - 2b\Phi\big(\tau|x|t - b/|x|\big)\big] + |x|\varphi\big(\tau|x|t + b/|x|\big)\Big\} K(dx) \tag{12.1.26}$$

Thus, a Taylor expansion of $F(t)$ in $t = 0$ yields

$$F(t) = 2\tau\Big[\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big]t + \mathrm{o}(t) \tag{12.1.27}$$

To obtain a remainder $\mathrm{O}(t^2)$ instead of $\mathrm{o}(t)$ we have to guarantee that $F(t)$ is twice differentiable. This can be achieved by another application of Lemma 16.2 of Bauer (1990). However, we additionally have to assume $\mathrm{E}\,|x|^3 < \infty$ to get an integrable dominant; confer (12.1.28).

(c) We continue part (b); i.e., if $\mathrm{E}\,|x|^3 < \infty$, the second derivative of F is,

$$F''(t) = \tau^2 \int \Big\{ \exp(-2b\tau t)\big[4b^2\Phi\big(\tau|x|t - b/|x|\big) - \big(\tau|x|^3 t + 3b|x|\big)\varphi\big(\tau|x|t - b/|x|\big)\big] - \big(\tau|x|^3 t + b|x|\big)\varphi\big(\tau|x|t + b/|x|\big)\Big\} K(dx) \tag{12.1.28}$$

Thus, a Taylor expansion of $F(t)$ in $t = 0$ yields

$$F(t) = 2\tau\Big[\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big]\big(t - b\tau t^2\big) + \mathrm{o}(t^2) \tag{12.1.29}$$

Moreover, if $\mathrm{E}\,|x|^5 < \infty$, $F(t)$ is even three times differentiable and we obtain a remainder $\mathrm{O}(t^3)$ instead of $\mathrm{o}(t^2)$. We now define

$$G(b,t) := \varepsilon - 2\tau \int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx) + \Big\{\varepsilon^2 - 2b\tau^2 \int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big\}t + R(b,t) \tag{12.1.30}$$

where $R(b,t) = \mathrm{o}(t)$ if $\mathrm{E}\,|x|^3 < \infty$, respectively $R(b,t) = \mathrm{O}(t^2)$ if $\mathrm{E}\,|x|^5 < \infty$ as $R(b,t)$ is a bounded function in b. Furthermore, $b = b(t)$, $b(0) = b^{\mathrm{as}}_{c,0}$ and $G(b^{\mathrm{as}}_{c,0}, 0) = 0$ by (12.1.10). We now expand $G(b,t)$ in $(b^{\mathrm{as}}_{c,0}, 0)$ where

$$\frac{\partial}{\partial b}\, G(b,t)\Big|_{b=b^{\mathrm{as}}_{c,0},\, t=0} = 2\tau \int \Phi\big(-b^{\mathrm{as}}_{c,0}/|x|\big)\, K(dx) \tag{12.1.31}$$

by an application of Lemma 16.2 of Bauer (1990) and

$$\frac{\partial}{\partial t}\, G(b,t)\Big|_{b=b^{\mathrm{as}}_{c,0},\, t=0} \overset{(12.1.10)}{=} \varepsilon(\varepsilon + b^{\mathrm{as}}_{c,0}\tau) \tag{12.1.32}$$

This leads us to

$$G(b,t) = 2\tau \int \Phi\big(-b^{\mathrm{as}}_{c,0}/|x|\big)\, K(dx)(b - b^{\mathrm{as}}_{c,0}) + \varepsilon(\varepsilon + b^{\mathrm{as}}_{c,0}\tau)t + R(b,t) \quad (12.1.33)$$

with $R(b,t) = \mathrm{o}(|b - b^{\mathrm{as}}_{c,0}|t^0)$ if $\mathrm{E}\,|x|^3 < \infty$, respectively $R(b,t) = \mathrm{O}\big(|b - b^{\mathrm{as}}_{c,0}|t\big)$ if $\mathrm{E}\,|x|^5 < \infty$. Thus, without any further assumptions on the absolute moments of x we have $G(b^{\mathrm{as}}_{c,0}, t) = \mathrm{o}(t^0)$. However, if we assume $\mathrm{E}\,|x|^3 < \infty$, then $G(b^{\mathrm{as}}_{c,0}, t) = \mathrm{O}(t)$. In both cases we get for any other $b_0 \neq b^{\mathrm{as}}_{c,0}$ that $G(b_0, t) = \mathrm{O}(1)$; i.e., (12.1.15) holds. Moreover, we obtain $G(b^{\mathrm{as.c}}_{c,0}, t) = \mathrm{o}(t)$ if $\mathrm{E}\,|x|^3 < \infty$, respectively $G(b^{\mathrm{as.c}}_{c,0}, t) = \mathrm{O}(t^2)$ if $\mathrm{E}\,|x|^5 < \infty$ and for any other b_1 such that $\limsup_{t\to 0} \frac{|b_1 - b^{\mathrm{as.c}}_{c,0}|}{t} > 0$ we have $\limsup_{t\to 0} \frac{G(b_1,t)}{t} > 0$; i.e., (12.1.17) holds. ////

12.1.2 Total Variation Neighborhoods

We assume $\varepsilon = 0$; i.e., we consider the following unconditional total variation neighborhoods,

$$\mathcal{U}_{v,0}(\theta, \delta_n) = \big\{Q \in \mathcal{M}_1(\mathbb{B}^2) \,\big|\, d_v\big(Q, P_\theta\big) \leq \delta_n\big\} \quad (12.1.34)$$

for given radius $\delta_n = \delta/\sqrt{n} \in (0,1)$ and fixed sample size $n \in \mathbb{N}$.

12.1.2.1 Finite-Sample Minimax Estimator

For a given width $\tau_n = \tau/\sqrt{n} \in (0,\infty)$ a least favorable pair $(R'_{-\tau_n,v,0}, R''_{\tau_n,v,0})$ for $\mathcal{U}_{v,0}(-\tau_n, \delta_n)$ versus $\mathcal{U}_{v,0}(\tau_n, \delta_n)$ by Subsubsection 10.2.2.1 is

$$R'_{-\tau_n,v,0}(dx,dy) = \varphi(y + \tau_n x) \begin{cases} \frac{1+\exp(2\tau_n xy)}{1+\exp(-2\tau_n b^{\mathrm{fi}}_{v,0})} & xy < -b^{\mathrm{fi}}_{v,0} \\ 1 & |xy| \leq b^{\mathrm{fi}}_{v,0} \\ \frac{1+\exp(2\tau_n xy)}{1+\exp(2\tau_n b^{\mathrm{fi}}_{v,0})} & xy > b^{\mathrm{fi}}_{v,0} \end{cases} \lambda(dy)K(dx) \quad (12.1.35)$$

$$R''_{\tau_n,v,0}(dx,dy) = \varphi(y - \tau_n x) \begin{cases} \frac{1+\exp(-2\tau_n xy)}{1+\exp(2\tau_n b^{\mathrm{fi}}_{v,0})} & xy < -b^{\mathrm{fi}}_{v,0} \\ 1 & |xy| \leq b^{\mathrm{fi}}_{v,0} \\ \frac{1+\exp(-2\tau_n xy)}{1+\exp(-2\tau_n b^{\mathrm{fi}}_{v,0})} & xy > b^{\mathrm{fi}}_{v,0} \end{cases} \lambda(dy)K(dx) \quad (12.1.36)$$

where $b^{\mathrm{fi}}_{v,0}$ is the unique solution to

$$\begin{aligned} &\big[1 + \exp(-2b^{\mathrm{fi}}_{v,0}\tau_n)\big]\delta_n \\ &= \int \big[\exp(-2b^{\mathrm{fi}}_{v,0}\tau_n)\Phi\big(\tau_n|x| - b^{\mathrm{fi}}_{v,0}/|x|\big) - \Phi\big(-\tau_n|x| - b^{\mathrm{fi}}_{v,0}/|x|\big)\big]\, K(dx) \end{aligned} \quad (12.1.37)$$

and the disjointness condition (10.2.15) reads

$$\delta_n < \textstyle\int \Phi\big(\tau_n|x|\big)\, K(dx) - \frac{1}{2} \quad (12.1.38)$$

This immediately implies $\delta_n < 0.5$ for all $\tau_n \in (0,\infty)$. The minimax robust test $\tilde{\phi}_{v,0}$ for $\mathcal{U}_{v,0}(-\tau_n)$ versus $\mathcal{U}_{v,0}(\tau_n)$, by Subsection 10.2.3 and symmetry, is

$$\tilde{\phi}_{v,0} = \tfrac{1}{2}\,\mathrm{I}(h_{v,0} > 0) + \tfrac{1}{2}\,\mathrm{I}(h_{v,0} \geq 0) \quad (12.1.39)$$

where $h_{0,v}$ is the loglikelihood of $(R'_{-\tau,v,0})^n$ with respect to $(R''_{\tau,v,0})^n$ as given in (10.2.23). Thus, the finite-sample minimax estimator $\tilde{S}^{\mathrm{fi}}_{v,0}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\psi}_{v,0}(x_i, y_i - x_i \tilde{S}^{\mathrm{fi}}_{v,0}) = 0 \qquad \tilde{\psi}_{v,0}(x,u) = xu \min\left\{1, \frac{b^{\mathrm{fi}}_{v,0}}{|xu|}\right\} \tag{12.1.40}$$

with equal randomization between the smallest and the largest solutions. The corresponding IC reads

$$\mathrm{IC}_{\tilde{\psi}_{v,0}}(x,u) = A^{\mathrm{fi}}_{v,0} \tilde{\psi}_{v,0}(x,u) \qquad (A^{\mathrm{fi}}_{v,0})^{-1} = 2 \int x^2 \Phi\big(b^{\mathrm{fi}}_{v,0}/|x|\big)\, K(dx) - \mathcal{K} \tag{12.1.41}$$

where $\mathcal{K} = \int x^2\, K(dx)$.

12.1.2.2 Asymptotic Minimax Estimator

The IC of the asymptotic minimax estimator in case of unconditional total variation neighborhoods by Theorem 10.3.7 (a) and symmetry is,

$$\tilde{\eta}_{v,0}(x,u) = A^{\mathrm{as}}_{v,0} xu \min\left\{1, \frac{b^{\mathrm{as}}_{v,0}}{|xu|}\right\} \tag{12.1.42}$$

with

$$(A^{\mathrm{as}}_{v,0})^{-1} = 2 \int x^2 \Phi\big(b^{\mathrm{as}}_{v,0}/|x|\big)\, K(dx) - \mathcal{K} \tag{12.1.43}$$

where $\mathcal{K} = \int x^2\, K(dx)$ and $b^{\mathrm{as}}_{v,0}$ is the unique solution to

$$\delta = \tau \int \big[|x| \varphi\big(b^{\mathrm{as}}_{v,0}/|x|\big) - b^{\mathrm{as}}_{v,0} \Phi\big(-b^{\mathrm{as}}_{v,0}/|x|\big)\big]\, K(dx) \tag{12.1.44}$$

and the boundedness restriction (10.3.39) reads

$$\delta < \tfrac{1}{\sqrt{2\pi}} \tau \int |x|\, K(dx) \tag{12.1.45}$$

As in case of contamination neighborhoods, it is not clear whether the asymptotic minimax estimator $\tilde{S}^{\mathrm{as}}_{v,0}$ can be constructed as an M estimator. Nevertheless, we define the M estimator $S^{\mathrm{as}}_{v,0}$ satisfying

$$\sum_{i=1}^{n} \tilde{\eta}_{v,0}(x_i, y_i - x_i S^{\mathrm{as}}_{v,0}) = 0 \tag{12.1.46}$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2 which may be shown analogously to Theorem 10.2.9 (a) and serves as a comparative estimator for the finite-sample minimax estimator $\tilde{S}^{\mathrm{fi}}_{v,0}$. Moreover, we get the following relations between finite-sample and asymptotic results.

Lemma 12.1.2 *For* $\delta_n \in (0,1)$, $\tau_n \in (0,\infty)$ *(*$n \in \mathbb{N}$ *fixed) it holds,*

(a)

$$\sqrt{n}\Big[\int \Phi(\tau_n|x|)\,K(dx) - \tfrac{1}{2}\Big] < \tfrac{1}{\sqrt{2\pi}}\,\tau \int |x|\,K(dx) \tag{12.1.47}$$

(b)

$$\begin{aligned}&\big[1+\exp(-2b\tau_n)\big]\delta\\ &\quad - \sqrt{n}\int \big[\exp(-2b\tau_n)\Phi\big(\tau_n|x| - b/|x|\big) - \Phi\big(-\tau_n|x| - b/|x|\big)\big]\,K(dx)\\ &= \delta - \tau\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\,K(dx) + \mathrm{o}(n^0)\end{aligned} \tag{12.1.48}$$

where we can replace the remainder $\mathrm{o}(n^0)$ *by* $\mathrm{O}(n^{-1/2})$ *if* $\mathrm{E}\,|x|^3 < \infty$.

(c)

$$b_{v,0}^{\mathrm{as}} = b_{v,0}^{\mathrm{fi}} + \mathrm{o}(n^0) \tag{12.1.49}$$

Moreover, if $\mathrm{E}\,|x|^3 < \infty$, *respectively* $\mathrm{E}\,|x|^5 < \infty$, *we obtain a remainder* $\mathrm{o}(n^{-1/2})$, *respectively* $\mathrm{O}(n^{-1})$ *instead of* $\mathrm{o}(n^0)$. *If* $\mathrm{E}\,|x|^5 < \infty$, *the* $\mathrm{O}(n^{-1})$*-corrected asymptotic optimal clipping bound is*

$$b_{v,0}^{\mathrm{as.c}} = b_{v,0}^{\mathrm{as}} - \frac{1}{n}\,\frac{\tau\Big[2\big(b_{v,0}^{\mathrm{as}}\big)^2\delta + \tau\int |x|^3\varphi\big(b_{v,0}^{\mathrm{as}}/|x|\big)\,K(dx)\Big]}{6\int \Phi\big(-b_{v,0}^{\mathrm{as}}/|x|\big)\,K(dx)} \tag{12.1.50}$$

where

$$b_{v,0}^{\mathrm{fi}} = b_{v,0}^{\mathrm{as.c}} + \mathrm{o}(n^{-1}) \tag{12.1.51}$$

Moreover, the remainder $\mathrm{o}(n^{-1})$ *can be replaced by* $\mathrm{O}(n^{-3/2})$ *if* $\mathrm{E}\,|x|^7 < \infty$.

Proof

(a) (12.1.47) is equivalent to

$$\int \Phi(\tau_n|x|)\,K(dx) - \tfrac{1}{\sqrt{2\pi}}\,\tau_n \int |x|\,K(dx) < \tfrac{1}{2} \tag{12.1.52}$$

where the left hand side is strictly decreasing in $\tau_n \in (0,\infty)$ because we get for its derivative (using (12.1.19))

$$\textstyle\int |x|\varphi(\tau_n|x|)\,K(dx) - \frac{1}{\sqrt{2\pi}}\int |x|\,K(dx) < 0 \tag{12.1.53}$$

since

$$|x|\varphi(\tau_n|x|) < \tfrac{1}{\sqrt{2\pi}}|x| \qquad \text{a.e. } K(dx) \tag{12.1.54}$$

by the assumption (10.1.5), which is $K(x=0) = 0$. Thus, the left hand side of (12.1.52) becomes maximal for $\tau_n \downarrow 0$ where $\Phi(0) = \frac{1}{2}$.

(b) Defining $t := 1/\sqrt{n}$, the left hand side of (12.1.37) reads,

$$\delta t\big(1+\exp(-2b\tau t)\big) \tag{12.1.55}$$

and Taylor expansion in $t = 0$ yields,

$$\delta t(2 + \mathrm{O}(t)) = 2\delta_n + \mathrm{O}(n^{-1}) \tag{12.1.56}$$

This proves part (b) as the right hand side of (12.1.37) is identical to the right hand side of (12.1.3) which already occurred in part (b) of Lemma 12.1.1.

(c) Defining $t := 1/\sqrt{n}$ (12.1.37) is equivalent to

$$\big[1 + \exp(-2b\tau t)\big]\delta t = \int \big[\exp(-2b\tau t)\Phi\big(\tau|x|t - b/|x|\big) - \Phi\big(-\tau|x|t - b/|x|\big)\big]\, K(dx) \tag{12.1.57}$$

and we get by Taylor expansions in $t = 0$ analogously to part (c) of Lemma 12.1.1

$$\delta t\big[1 - b\tau t + b^2\tau^2 t^2\big] = 2\tau\Big\{\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big\}\big(t - b\tau t^2\big) + R(t) \tag{12.1.58}$$

where $R(t) = \mathrm{o}(t^2)$ if $\mathrm{E}\,|x|^3 < \infty$ and $R(t) = \mathrm{O}(t^3)$ if $\mathrm{E}\,|x|^5 < \infty$. Furthermore, if $\mathrm{E}\,|x|^5 < \infty$, then $F(t)$ as given in (12.1.25) is three times differentiable (by Lemma 16.2 of Bauer (1990)) with third derivative

$$\begin{aligned} F'''(t) = \tau^3 \int \Big\{ & \exp(-2b\tau t)\big[-8b^3\Phi\big(\tau|x|t - b/|x|\big) \\ & - \big(7b^2|x| - |x|^3 + 4b\tau|x|^3 t + \tau^2|x|^5 t^2\big)\varphi\big(\tau|x|t - b/|x|\big)\big] \\ & + \big(b^2|x| - |x|^3 + 2b\tau|x|^3 t + \tau^2|x|^5 t^2\big)\varphi\big(\tau|x|t + b/|x|\big)\Big\}\, K(dx) \end{aligned} \tag{12.1.59}$$

We now define

$$\begin{aligned} G(b,t) := & \Big\{\delta - \tau\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big\}(1 - b\tau t) \\ & + \Big\{\delta - \tfrac{2}{3}\tau\int \big[|x|\varphi\big(b/|x|\big) - b\Phi\big(-b/|x|\big)\big]\, K(dx)\Big\}b^2\tau^2 t^2 \\ & + \tfrac{1}{6}\tau^3\int |x|^3\varphi\big(b/|x|\big)\, K(dx)t^2 + R(b,t) \end{aligned} \tag{12.1.60}$$

where $R(b,t) = \mathrm{o}(t^2)$ if $\mathrm{E}\,|x|^5 < \infty$ since $R(b,t)$ is a bounded function in b. Moreover, if $\mathrm{E}\,|x|^7 < \infty$, $F(t)$ is even four times differentiable and $R(b,t) = \mathrm{O}(t^3)$ ($R(b,t)$ bounded in b). In addition, $b = b(t)$, $b(0) = b^{\mathrm{as}}_{v,0}$ and $G(b^{\mathrm{as}}_{v,0}, 0) = 0$ by (12.1.44). Furthermore, we get

$$\frac{\partial}{\partial b}\, G(b,t)\bigg|_{b=b^{\mathrm{as}}_{v,0},\, t=0} = \tau\int \Phi\big(-b^{\mathrm{as}}_{v,0}/|x|\big)\, K(dx) \tag{12.1.61}$$

by an application of Lemma 16.2 of Bauer (1990) and

$$\frac{\partial}{\partial t}\, G(b,t)\bigg|_{b=b^{\mathrm{as}}_{v,0},\, t=0} \overset{(12.1.44)}{=} 0 \tag{12.1.62}$$

$$\frac{\partial^2}{\partial t^2}\, G(b,t)\bigg|_{b=b^{\mathrm{as}}_{v,0},\, t=0} \overset{(12.1.44)}{=} \tfrac{1}{3}\tau^2\Big[2\big(b^{\mathrm{as}}_{v,0}\big)^2\delta + \tau\int |x|^3\varphi\big(b^{\mathrm{as}}_{v,0}/|x|\big)\, K(dx)\Big] \tag{12.1.63}$$

This leads us to

$$G(b,t) = \tau \int \Phi\big(-b_{v,0}^{\mathrm{as}}/|x|\big)\, K(dx)(b - b_{v,0}^{\mathrm{as}}) + \tfrac{1}{6}\tau^2\Big[2\big(b_{v,0}^{\mathrm{as}}\big)^2\delta + \tau\int |x|^3\varphi\big(b_{v,0}^{\mathrm{as}}/|x|\big)\, K(dx)\Big]t^2 + R(b,t) \quad (12.1.64)$$

with $R(b,t) = \mathrm{o}(t^2) + \mathrm{O}(|b - b_{v,0}^{\mathrm{as}}|t)$ if $\mathrm{E}\,|x|^5 < \infty$ and $R(b,t) = \mathrm{O}\big(|b - b_{v,0}^{\mathrm{as}}|t\big)$ if $\mathrm{E}\,|x|^7 < \infty$. Thus, without any further assumptions on the absolute moments of x we have $G(b_{v,0}^{\mathrm{as}}, t) = \mathrm{o}(t^0)$. Moreover, if $\mathrm{E}\,|x|^3 < \infty$, then $G(b_{v,0}^{\mathrm{as}}, t) = \mathrm{o}(t)$, respectively if $\mathrm{E}\,|x|^5 < \infty$, then $G(b_{v,0}^{\mathrm{as}}, t) = \mathrm{O}(t^2)$. In all three cases we get for any other $b_0 \neq b_{v,0}^{\mathrm{as}}$ that $G(b_0, t) = \mathrm{O}(1)$; i.e., (12.1.49) holds. Moreover, we obtain $G(b_{v,0}^{\mathrm{as,c}}, t) = \mathrm{o}(t^2)$ if $\mathrm{E}\,|x|^5 < \infty$, respectively $G(b_{v,0}^{\mathrm{as,c}}, t) = \mathrm{O}(t^3)$ if $\mathrm{E}\,|x|^7 < \infty$ and for any other b_1 such that $\limsup_{t\to 0} \frac{|b_1 - b_{v,0}^{\mathrm{as,c}}|}{t^2} > 0$ we have $\limsup_{t\to 0} \frac{G(b_1,t)}{t^2} > 0$; i.e., (12.1.51) holds. ////

Remark 12.1.3 We conjecture the moment conditions included in Lemma 12.1.1 and Lemma 12.1.2 to be not only sufficient but also necessary. Some numerical calculations which we did for regressor distribution K equal to different F-distributions seem to confirm this conjecture. However, we will not investigate this further in this thesis. ////

12.1.3 Computation of the Finite-Sample Risk

In this section we take a closer look at the finite-sample risk in case of unconditional regression neighborhoods (cf. Subsubsection 12.1.3.1) and propose two procedures for the numerical computation of this risk (cf. Subsubsection 12.1.3.2). We check the results of these two algorithms by means of simulations.

12.1.3.1 Finite-Sample Risk

We fix $n \in \mathbb{N}$ and radius $\varepsilon_n \in [0,1)$ $(* = c)$, respectively radius $\delta_n \in [0,1)$ $(* = v)$. Given some clipping bound $b \in (0,\infty)$ we then want to determine the finite-sample risk of an M estimator S satisfying

$$\sum_{i=1}^{n} \chi_0(x_i, y_i - x_i S) = 0 \qquad \chi_0(x,u) = xu \min\left\{1, \frac{b}{|xu|}\right\} \quad (12.1.65)$$

with equal randomization between the smallest and the largest solutions. In particular, we are interested in the finite-sample minimax estimator $\tilde{S}_*^{\mathrm{fi}}$ and the estimator S_*^{as}, respectively. Moreover, if $\mathrm{E}\,|x|^3 < \infty$, respectively $\mathrm{E}\,|x|^5 < \infty$ (or even $\mathrm{E}\,|x|^7 < \infty$) we are also interested in the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected $(* = c)$, respectively $\mathrm{O}(n^{-1})$-corrected $(* = v)$ asymptotic optimal clipping bound. All these estimators are equivariant in sense of Definition 10.2.2; confer Theorem 10.2.9. As already stated in Remark 10.2.3, the finite-sample risk

of any equivariant estimator S for given $\tau_n \in (0,\infty)$ reads,

$$\text{Risk}\,(S;*,0) = \max\left\{ \sup_{Q_{-\tau_n}\in\mathcal{U}_{*,0}(-\tau_n)} Q^n_{-\tau_n}(S>0), \sup_{Q_{\tau_n}\in\mathcal{U}_{*,0}(\tau_n)} Q^n_{\tau_n}(S<0)\right\} \tag{12.1.66}$$

with $\mathcal{U}_{*,0}(-\tau_n) = g_{-\theta-\tau_n}\big(\mathcal{U}_{*,0}(\theta)\big)$ and $\mathcal{U}_{*,0}(\tau_n) = g_{-\theta+\tau_n}\big(\mathcal{U}_{*,0}(\theta)\big)$ by (10.1.12). Moreover, we have the following inclusions

$$\{S'>\theta\} \subseteq \Big\{ \textstyle\sum_{i=1}^n \chi_0(x_i, y_i - x_i\theta) > 0\Big\} \subseteq \{S'\geq\theta\} \tag{12.1.67}$$

$$\{S''>\theta\} \subseteq \Big\{ \textstyle\sum_{i=1}^n \chi_0(x_i, y_i - x_i\theta) \geq 0\Big\} \subseteq \{S''\geq\theta\} \tag{12.1.68}$$

where

$$S'(\mathbf{x},\mathbf{y}) = \sup_{\theta\in\mathbb{R}}\Big\{ \textstyle\sum_{i=1}^n \chi_0(x_i, y_i - x_i\theta) > 0\Big\} \tag{12.1.69}$$

$$S''(\mathbf{x},\mathbf{y}) = \inf_{\theta\in\mathbb{R}}\Big\{ \textstyle\sum_{i=1}^n \chi_0(x_i, y_i - x_i\theta) < 0\Big\} \tag{12.1.70}$$

for any given $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$; confer also Rieder (1989), p 218. Thus, we can proceed analogously to the location case (cf. Subsection 11.3.1) and obtain for any such M estimator S and any $Q_{-\tau_n}\in\mathcal{U}_{*,0}(-\tau_n)$, respectively $Q_{\tau_n}\in\mathcal{U}_{*,0}(\tau_n)$,

$$Q^n_{-\tau_n}(S>0) = \tfrac{1}{2}\big[Q^n_{-\tau_n}(S'>0) + Q^n_{-\tau_n}(S''>0)\big] \tag{12.1.71}$$

$$\leq \tfrac{1}{2}\left[Q^n_{-\tau_n}\left(\sum_{i=1}^n \chi_0(x_i,y_i)>0\right) + Q^n_{-\tau_n}\left(\sum_{i=1}^n \chi_0(x_i,y_i)\geq 0\right)\right] \tag{12.1.72}$$

$$\leq \tfrac{1}{2}\big[Q^n_{-\tau_n}(S'\geq 0) + Q^n_{-\tau_n}(S''\geq 0)\big] \tag{12.1.73}$$

respectively

$$Q^n_{\tau_n}(S<0) = \tfrac{1}{2}\big[Q^n_{\tau_n}(S'<0) + Q^n_{\tau_n}(S''<0)\big] \tag{12.1.74}$$

$$\leq \tfrac{1}{2}\left[Q^n_{\tau_n}\left(\sum_{i=1}^n \chi_0(x_i,y_i)\leq 0\right) + Q^n_{\tau_n}\left(\sum_{i=1}^n \chi_0(x_i,y_i)< 0\right)\right] \tag{12.1.75}$$

$$\leq \tfrac{1}{2}\big[Q^n_{\tau_n}(S'\leq 0) + Q^n_{\tau_n}(S''\leq 0)\big] \tag{12.1.76}$$

Remark 12.1.4 If $Q_\theta\in\mathcal{U}_{*,0}(\theta)$ $(\theta\in\mathbb{R})$ fulfills conditions (10.2.6) and (10.2.7) of Lemma 10.2.4, then the unconditional probability $Q_\theta(S=0)$ vanishes for every equivariant estimator. Thus, we obtain equality in (12.1.72) and (12.1.73), respectively in (12.1.75) and (12.1.76) if $Q_{-\tau_n}$, respectively Q_{τ_n} fulfills (10.2.6) and (10.2.7). ////

By monotonicity of χ_0, the distribution of $\sum_{i=1}^n \chi_0(x_i,y_i)$ is monotone with respect to stochastic order; i.e., the probability of $\sum_{i=1}^n \chi_0(x_i,y_i)>0$, respectively of $\sum_{i=1}^n \chi_0(x_i,y_i)\geq 0$ under $Q_{-\tau_n}\in\mathcal{U}_{*,0}(-\tau_n)$ is maximal if

$$Q_{-\tau_n}(\chi_0(x,y)=b) = Q_{-\tau_n}(xy\geq b) = \max! \tag{12.1.77}$$

where $b \in (0,\infty)$ is some given clipping bound and we only consider distributions $Q_{-\tau_n}$ such that $Q_{-\tau_n}(x=0)=0$; i.e., (10.2.6) holds. Moreover, we have

$$Q_{-\tau_n}(xy \geq b) = Q_{-\tau_n}(y \geq b/x,\, x>0) + Q_{-\tau_n}(y \leq b/x,\, x<0) \tag{12.1.78}$$

$$\overset{(10.1.12)}{=} \int_0^\infty \int_{b/x}^\infty Q_0(dy+\tau_n x \,|\, x)\, Q_0(dx) + \int_{-\infty}^0 \int_{-\infty}^{b/x} Q_0(dy+\tau_n x \,|\, x)\, Q_0(dx) \tag{12.1.79}$$

$$= \int_0^\infty \int_{b/x+\tau_n x}^\infty Q_0(dy \,|\, x)\, Q_0(dx) + \int_{-\infty}^0 \int_{-\infty}^{b/x+\tau_n x} Q_0(dy \,|\, x)\, Q_0(dx) \tag{12.1.80}$$

In case of unconditional contamination neighborhoods ($* = c$, $t=0$) this yields

$$\begin{aligned} &Q_{-\tau_n}(xy \geq b) \\ &\quad = (1-\varepsilon_n) \int \Phi(-b/|x| - \tau_n|x|)\, K(dx) \\ &\qquad + \varepsilon_n \Big\{ \int_0^\infty \int_{b/x+\tau_n x}^\infty H(dy \,|\, x)\, H(dx) + \int_{-\infty}^0 \int_{-\infty}^{b/x+\tau_n x} H(dy \,|\, x)\, H(dx) \Big\} \end{aligned} \tag{12.1.81}$$

for any $H \in \mathcal{M}_1(\mathbb{B}^2)$. To achieve the maximum, we may choose $H'_{-\tau_n} \in \mathcal{M}_1(\mathbb{B}^2)$

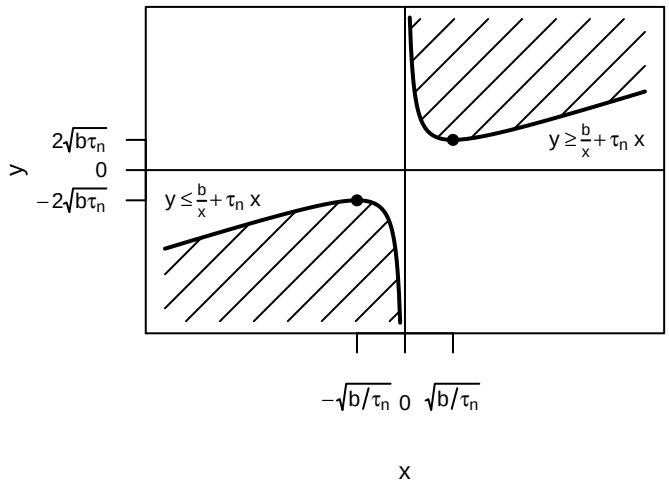

Figure 12.1: Support of $H'_{-\tau_n}$.

such that $H'_{-\tau_n}(dx)$ is some arbitrary distribution with $H'_{-\tau_n}(x=0)=0$. But, $H'_{-\tau_n}(dy\,|\,x)$ has to be concentrated on $[b/x+\tau_n x,\infty)$ if $x>0$, respectively on $(-\infty, b/x+\tau_n x]$ if $x<0$; confer Figure 12.1. This leads us to

$$Q'_{-\tau_n} = (1-\varepsilon_n)P_{-\tau_n} + \varepsilon_n H'_{-\tau_n} \qquad (*=c,\, t=0) \tag{12.1.82}$$

Of course, we should choose $H'_{-\tau_n}$ such that the support of $H'_{-\tau_n}(dx)$ is included in the support of $K(dx)$ or even such that $H'_{-\tau_n}(dx)=K(dx)$. Otherwise, one could identify outliers with probability 1. In case of unconditional total variation neighborhoods we can use similar arguments and obtain some $Q'_{-\tau_n}$ such that

$$Q'_{-\tau_n}(xy\le t) = \Big(\int \Phi\big(t/|x|+\tau_n|x|\big)\,K(dx) - \delta_n\Big)_+ + \delta_n H'_{-\tau_n}(xy\le t) \tag{12.1.83}$$

for all $t\in\mathbb{R}$. That is, mass δ_n is moved from the region where xy is smaller than the corresponding δ_n quantile to the region where $xy\ge b$.

Analogously, we obtain the probability of $\sum_{i=1}^n \chi_0(x_i,y_i)<0$, respectively of $\sum_{i=1}^n \chi_0(x_i,y_i)\le 0$ under $Q_{\tau_n}\in\mathcal{U}_{*,0}(\tau_n)$ is maximal if

$$Q_{\tau_n}(\chi_0(x,y)=-b) = Q_{\tau_n}(xy\le -b) = \max! \tag{12.1.84}$$

where $b\in(0,\infty)$ is some given clipping bound and we again only consider distributions Q_{τ_n} such that $Q_{\tau_n}(x=0)=0$; i.e., (10.2.6) holds. By similar arguments as in case of $Q_{-\tau_n}$ we get for unconditional contamination neighborhoods $(*=c,\, t=0)$

$$Q''_{\tau_n} = (1-\varepsilon_n)P_{\tau_n} + \varepsilon_n H''_{\tau_n} \tag{12.1.85}$$

where $H''_{\tau_n}(x=0)=0$ and the support of $H''_{\tau_n}(dx)$ is included in the support of $K(dx)$ or even $H'_{-\tau_n}(dx)=K(dx)$. Moreover, $H''_{\tau_n}(dy\,|\,x)$ is concentrated on $(-\infty, -b/x-\tau_n x]$ if $x>0$, respectively on $[-b/x-\tau_n x,\infty)$ if $x<0$; confer Figure 12.2. In case of unconditional total variation neighborhoods $(*=v,\, t=0)$ this yields some Q''_{τ_n} such that

$$Q''_{\tau_n}(xy\le t) = \min\Big\{\int \Phi\big(t/|x|-\tau_n|x|\big)\,K(dx) + \delta_n H''_{\tau_n}(xy\le t),\, 1\Big\} \tag{12.1.86}$$

for all $t\in\mathbb{R}$. That is, mass δ_n is moved from the region where xy is larger than the corresponding $1-\delta_n$ quantile to the region where $xy\le -b$.

Remark 12.1.5 In view of Remark 12.1.4 we additionally have to choose $H'_{-\tau_n}$ and H''_{τ_n} such that $H'_{-\tau_n}(dy\,|\,x)$, respectively $H''_{\tau_n}(dy\,|\,x)$ is absolutely continuous a.e. $H'_{-\tau_n}(dx)$, respectively a.e. $H''_{\tau_n}(dx)$ to obtain equality in (12.1.72) and (12.1.73), respectively in (12.1.75) and (12.1.76). Then, we may compute the finite-sample risk (12.1.66) by determining the distribution of $\sum_{i=1}^n \chi_0(x_i,y_i)$ under $Q'_{-\tau_n}$ and Q''_{τ_n} where it holds by symmetry,

$$\text{Risk}\,(S;*,0) = \big(Q'_{-\tau_n}\big)^n(S>0) = \big(Q''_{\tau_n}\big)^n(S<0) \tag{12.1.87}$$

Concerning the absolute continuity of $H'_{-\tau_n}(dy\,|\,x)$ and $H''_{\tau_n}(dy\,|\,x)$, we could also argue analogously to Remark 11.3.2 (b) and (c). ////

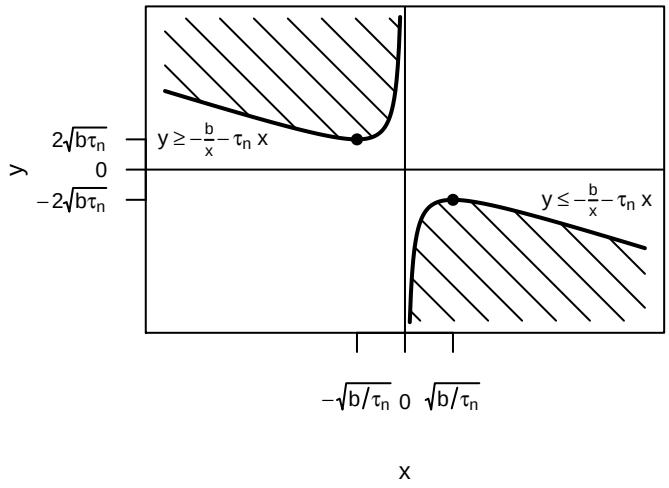

Figure 12.2: Support of H''_{τ_n}.

We may now calculate the finite-sample risk by determining the distribution of $\sum_{i=1}^n \chi_0(x_i, y_i)$ under $\big(Q'_{-\tau_n}\big)^n$ or $\big(Q''_{\tau_n}\big)^n$, respectively. As a first step, we state the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n} where in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$ we have

$$Q'_{-\tau_n}(\chi_0(x,y) = -b) = (1-\varepsilon_n) \int \Phi\big(-b/|x| + \tau_n|x|\big)\, K(dx) \tag{12.1.88}$$

$$\begin{aligned} &Q'_{-\tau_n}(-b < \chi_0(x,y) < t) \\ &= (1-\varepsilon_n) \int \Big[\Phi\big(t/|x| + \tau_n|x|\big) - \Phi(-b/|x| + \tau_n|x|)\Big]\, K(dx) \quad t \in (-b,b) \end{aligned} \tag{12.1.89}$$

$$Q'_{-\tau_n}(\chi_0(x,y) = b) = (1-\varepsilon_n) \int \Phi\big(-b/|x| - \tau_n|x|\big)\, K(dx) + \varepsilon_n \tag{12.1.90}$$

respectively

$$Q''_{\tau_n}(\chi_0(x,y) = -b) = (1-\varepsilon_n) \int \Phi\big(-b/|x| - \tau_n|x|\big)\, K(dx) + \varepsilon_n \tag{12.1.91}$$

$$\begin{aligned} &Q''_{\tau_n}(-b < \chi_0(x,y) < t) \\ &= (1-\varepsilon_n) \int \Big[\Phi\big(t/|x| - \tau_n|x|\big) - \Phi(-b/|x| - \tau_n|x|)\Big]\, K(dx) \quad t \in (-b,b) \end{aligned} \tag{12.1.92}$$

$$Q''_{\tau_n}(\chi_0(x,y) = b) = (1-\varepsilon_n) \int \Phi\big(-b/|x| + \tau_n|x|\big)\, K(dx) \tag{12.1.93}$$

In case of unconditional total variation neighborhoods ($* = v$, $t = 0$) we get,

$$Q'_{-\tau_n}(\chi_0(x,y) = -b) = \Big(\int \Phi\big(-b/|x| + \tau_n|x| \big)\, K(dx) - \delta_n \Big)_+ \tag{12.1.94}$$

$$\begin{aligned} &Q'_{-\tau_n}(-b < \chi_0(x,y) < t) \\ &\quad = \Big(\int \Phi\big(t/|x| + \tau_n|x|\big)\, K(dx) - \delta_n \Big)_+ \\ &\qquad - \Big(\int \Phi\big(-b/|x| + \tau_n|x| \big)\, K(dx) - \delta_n \Big)_+ \qquad t \in (-b, b) \end{aligned} \tag{12.1.95}$$

$$Q'_{-\tau_n}(\chi_0(x,y) = b) = 1 - \Big(\int \Phi\big(b/|x| + \tau_n|x|\big)\, K(dx) - \delta_n \Big)_+ \tag{12.1.96}$$

respectively

$$Q''_{\tau_n}(\chi_0(x,y) = -b) = \min\Big\{ \int \Phi\big(-b/|x| - \tau_n|x| \big)\, K(dx) + \delta_n,\, 1 \Big\} \tag{12.1.97}$$

$$\begin{aligned} &Q''_{\tau_n}(-b < \chi_0(x,y) < t) \\ &\quad = \min\Big\{ \int \Phi\big(t/|x| - \tau_n|x|\big)\, K(dx) + \delta_n,\, 1 \Big\} \\ &\qquad - \min\Big\{ \int \Phi\big(-b/|x| - \tau_n|x| \big)\, K(dx) + \delta_n,\, 1 \Big\} \qquad t \in (-b, b) \end{aligned} \tag{12.1.98}$$

$$Q''_{\tau_n}(\chi_0(x,y) = b) = 1 - \min\Big\{ \int \Phi\big(b/|x| - \tau_n|x|\big)\, K(dx) + \delta_n,\, 1 \Big\} \tag{12.1.100}$$

12.1.3.2 Numerical Algorithms

We determine the finite-sample risk by computing a numerical approximation of the distribution of $\sum_{i=1}^n \chi_0(x_i, y_i)$ under $\big(Q'_{-\tau_n}\big)^n$ and $\big(Q''_{\tau_n}\big)^n$, respectively. The first procedure corresponds to Algorithm A given in Subsubsection 11.3.2.1 and directly uses the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n} which can be read off from (12.1.88)–(12.1.100). The n-fold convolution of this distribution is calculated using Algorithm C.2.2 which is based on the fast Fourier transform; confer Appendix C.2. The second procedure corresponds to Algorithm B given in Subsubsection 11.3.2.2 where we now have

$$p_1 := (1 - \varepsilon_n) \int \Big[\Phi\big(-b/|x| + \tau_n|x| \big) + \Phi\big(-b/|x| - \tau_n|x| \big) \Big]\, K(dx) + \varepsilon_n \tag{12.1.101}$$

$$p_2 := \Big[(1 - \varepsilon_n) \int \Phi\big(-b/|x| - \tau_n|x| \big)\, K(dx) + \varepsilon_n \Big] \Big/ p_1 \tag{12.1.102}$$

and $\tilde{Z}_i \overset{\text{i.i.d.}}{\sim} Q$ with cumulative distribution function

$$Q(t) = \int \Phi\big(t/|x| - \tau_n|x|\big)\, K(dx) \qquad (t \in \mathbb{R}) \tag{12.1.103}$$

We expect the precision of Algorithms A and B in case of simple linear regression and unconditional neighborhoods to be similar to the location case (cf. Subsubsection 11.3.2.3) but we do not have analytical results to check this. However, to

test our programs we made several cross checks. We for instance choose $K = \mathrm{I}_{\{1\}}$, respectively $K = \mathrm{Unif}([0.95, 1.05])$ such that we could expect results that are identical, respectively close to the location case. Moreover, we do some numerical simulations to check the algorithms. That is, we simulate 1e06 samples of size n and compute the empirical finite-sample risk and the corresponding 95% confidence interval. As it turns out, the results of Algorithms A and B always lie well within the 95% confidence interval; confer Tables 12.1–12.4. Of course, another cross check is that both algorithms yield approximately the same results. Thus, we determine the distance $\mathrm{dist}_{\mathrm{AB}}$ between the results of Algorithm A and Algorithm B where we additionally give the computation time in seconds T_A, respectively T_B on an Athlon with 2.5 GHz and 512 MB RAM. To restrict the amount of results, we consider only the finite-sample risk of the estimator $S_{*,0}^{\mathrm{as}}$ and choose one "typical" situation as the results are almost independent of τ and ε, respectively δ. That is, we fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$, respectively $\delta = 0.1$. $\mathrm{Risk}_{\mathrm{B}}$ denotes the finite-sample risk of $S_{*,0}^{\mathrm{as}}$ computed with Algorithm B and 2^q stands for the number of lattice points used in step 2 of Algorithm C.2.2. The first two tables (see Tables 12.1 and 12.2) show the results for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional contamination, respectively total variation neighborhoods. The remaining two tables (see Tables 12.3 and 12.4) contain the results for $K = \mathrm{Unif}([-1, 2])$ and unconditional contamination, respectively total variation neighborhoods. As in the location case (cf. Subsubsection 11.3.2.3) the precision of the results turn out to be (almost) independent of the neighborhood type.

Remark 12.1.6 Since the differences between the two algorithms are only small and since Algorithm B is much slower, we in most cases use Algorithm A and choose $q = 12$. ////

n	emp. risk	95% conf. int.	q	Risk_{B}	dist_{AB}	T_A	T_B
2	0.2140	[0.2132, 0.2148]	10	0.214298	6.7e−05	0.31	0.33
			12	0.214298	1.7e−05	0.85	0.90
			14	0.214298	4.2e−06	3.10	3.30
3	0.1338	[0.1331, 0.1345]	10	0.133351	5.5e−05	0.33	0.75
			12	0.133351	1.4e−05	0.88	2.50
			14	0.133351	3.4e−06	3.15	10.20
4	0.1080	[0.1074, 0.1086]	10	0.107819	5.1e−05	0.34	0.95
			12	0.107819	1.3e−05	0.88	3.30
			14	0.107819	3.2e−06	3.30	14.20
5	0.0874	[0.0869, 0.0880]	10	0.087459	4.6e−05	0.33	1.75
			12	0.087459	1.1e−05	0.90	6.90
			14	0.087459	2.9e−06	3.30	28.30
10	0.0468	[0.0464, 0.0472]	10	0.047159	3.0e−05	0.35	5.50
			12	0.047159	7.5e−06	0.98	22.90
			14	0.047159	1.9e−06	3.90	97.00

Table 12.1: A comparison between Algorithm A, Algorithm B and empirical results for $K = \frac{1}{3}\big(\text{I}_{\{0.5\}} + \text{I}_{\{1.0\}} + \text{I}_{\{1.5\}}\big)$ in case of unconditional contamination neighborhoods $(* = c, t = 0)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$ which leads to $b^{\text{as}}_{c,0} = 1.690$. Risk_B denotes the finite-sample risk of $S^{\text{as}}_{c,0}$ computed with Algorithm B, 2^q the number of lattice points used in Algorithm C.2.2, dist_{AB} the distance between the results of Algorithm A and Algorithm B and T_A and T_B the computation time in seconds for Algorithm A and Algorithm B, respectively.]

n	emp. risk	95% conf. int.	q	$\mathrm{Risk_B}$	$\mathrm{dist_{AB}}$	T_A	T_B
2	0.1228	[0.1221, 0.1234]	10	0.122721	3.9e−05	0.32	0.34
			12	0.122721	9.7e−06	0.86	0.93
			14	0.122721	2.4e−06	3.17	3.55
3	0.0751	[0.0746, 0.0756]	10	0.075221	3.5e−05	0.34	0.75
			12	0.075221	8.6e−06	0.88	2.60
			14	0.075221	2.2e−06	3.20	10.50
4	0.0569	[0.0564, 0.0573]	10	0.056804	3.0e−05	0.34	0.96
			12	0.056804	7.4e−06	0.90	3.50
			14	0.056804	1.8e−06	3.40	14.80
5	0.0454	[0.0450, 0.0458]	10	0.045296	2.4e−05	0.35	1.80
			12	0.045296	6.0e−06	0.91	7.00
			14	0.045296	1.5e−06	3.45	29.15
10	0.0247	[0.0244, 0.0250]	10	0.024619	1.2e−05	0.35	5.60
			12	0.024619	3.0e−06	0.98	23.00
			14	0.024619	7.4e−07	4.05	98.25

Table 12.2: A comparison between Algorithm A, Algorithm B and empirical results for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ in case of unconditional total variation neighborhoods $(* = v, t = 0)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\delta = 0.1$ which leads to $b^{\mathrm{as}}_{v,0} = 1.690$. Risk_B denotes the finite-sample risk of $S^{\mathrm{as}}_{v,0}$ computed with Algorithm B, 2^q the number of lattice points used in Algorithm C.2.2, $\mathrm{dist_{AB}}$ the distance between the results of Algorithm A and Algorithm B and T_A and T_B the computation time in seconds for Algorithm A and Algorithm B, respectively.]

n	emp. risk	95% conf. int.	q	Risk_B	dist_AB	T_A	T_B
2	0.2547	[0.2539, 0.2556]	10	0.254439	5.7e−05	1.10	1.60
			12	0.254438	1.4e−05	4.10	5.80
			14	0.254438	3.5e−06	15.80	24.00
3	0.1960	[0.1952, 0.1968]	10	0.195739	5.3e−05	1.12	4.50
			12	0.195739	3.2e−05	4.10	18.00
			14	0.195739	3.3e−06	15.90	71.00
4	0.1643	[0.1636, 0.1651]	10	0.163796	5.9e−05	1.12	5.80
			12	0.163796	1.5e−05	4.10	22.70
			14	0.163796	3.7e−06	15.95	92.00
5	0.1400	[0.1393, 0.1407]	10	0.140584	5.6e−05	1.20	11.25
			12	0.140584	1.4e−05	4.10	45.00
			14	0.140584	2.1e−06	15.95	182.00

Table 12.3: A comparison between Algorithm A, Algorithm B and empirical results for $K = \text{Unif}([-1, 2])$ in case of unconditional contamination neighborhoods $(* = c, t = 0)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$ which leads to $b^{\text{as}}_{c,0} = 1.649$. Risk_B denotes the finite-sample risk of $S^{\text{as}}_{c,0}$ computed with Algorithm B, 2^q the number of lattice points used in Algorithm C.2.2, dist_AB the distance between the results of Algorithm A and Algorithm B and T_A and T_B the computation time in seconds for Algorithm A and Algorithm B, respectively.]

n	emp. risk	95% conf. int.	q	Risk_B	dist_AB	T_A	T_B
2	0.1652	[0.1644, 0.1659]	10	0.165362	3.3e−05	1.10	1.55
			12	0.165361	8.2e−06	4.00	6.00
			14	0.165361	2.1e−06	15.90	23.90
3	0.1258	[0.1251, 0.1264]	10	0.125791	5.3e−05	1.10	4.45
			12	0.125791	8.3e−06	4.05	17.75
			14	0.125791	2.1e−06	15.80	71.00
4	0.1021	[0.1015, 0.1027]	10	0.101998	3.7e−05	1.07	5.75
			12	0.101997	9.1e−06	3.88	22.50
			14	0.101997	2.3e−06	15.55	90.50
5	0.0859	[0.0853, 0.0864]	10	0.085828	3.4e−05	1.11	11.05
			12	0.085828	8.5e−06	3.95	44.45
			14	0.085828	2.1e−06	15.65	180.00

Table 12.4: A comparison between Algorithm A, Algorithm B and empirical results for $K = \text{Unif}([-1, 2])$ in case of unconditional total variation neighborhoods $(* = v, t = 0)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\delta = 0.1$ which leads to $b^{\text{as}}_{v,0} = 1.649$. Risk_B denotes the finite-sample risk of $S^{\text{as}}_{v,0}$ computed with Algorithm B, 2^q the number of lattice points used in Algorithm C.2.2, dist_AB the distance between the results of Algorithm A and Algorithm B and T_A and T_B the computation time in seconds for Algorithm A and Algorithm B, respectively.]

12.2 Conditional Neighborhoods

We now specialize the results to conditional, or error-free-variables, neighborhoods as given in Definition 10.1.3 (b).

12.2.1 Contamination Neighborhoods

Let $\delta = 0$; i.e., consider the following conditional contamination neighborhoods,

$$\begin{aligned}\mathcal{U}_{c,\varepsilon}(\theta, \varepsilon_n) = \big\{ & Q \in \mathcal{M}_1(\mathbb{B}^2) \,\big|\, Q(dx) = K(dx), \\ & Q(dy \,|\, x) = (1 - \varepsilon_n(x))\mathcal{N}(x\theta, 1)(dy) + \varepsilon_n(x) H(dy \,|\, x) \quad \text{a.e. } K(dx), \\ & \text{for any Markov kernel } H : \mathbb{B} \times \mathbb{R} \to [0,1] \big\} \end{aligned} \tag{12.2.1}$$

with given contamination curve $\varepsilon \colon \mathbb{R} \to [0, \infty)$ such that $\varepsilon_n(x) = \varepsilon(x)/\sqrt{n} \in (0,1)$ a.e. $K(dx)$ and fixed sample size $n \in \mathbb{N}$.

12.2.1.1 Finite-Sample Minimax Estimator

For a given width $\tau_n = \tau/\sqrt{n} \in (0, \infty)$ a least favorable pair $(R'_{-\tau_n,c,\varepsilon}, R''_{\tau_n,c,\varepsilon})$ for $\mathcal{U}_{c,\varepsilon}(-\tau_n, \varepsilon_n)$ versus $\mathcal{U}_{c,\varepsilon}(\tau_n, \varepsilon_n)$ by Subsubsection 10.2.2.2 is

$$\begin{aligned} & R'_{-\tau_n,c,\varepsilon}(dx, dy) \\ & = (1 - \varepsilon_n(x))\varphi(y + \tau_n x) \max\big\{1, \exp\big(2\tau_n[xy - b^{\text{fi}}_{c,\varepsilon}(x)|x|]\big)\big\}\lambda(dy)K(dx) \end{aligned} \tag{12.2.2}$$

$$\begin{aligned} & R''_{\tau_n,c,\varepsilon}(dx, dy) \\ & = (1 - \varepsilon_n(x))\varphi(y - \tau_n x) \max\big\{1, \exp\big(-2\tau_n[xy + b^{\text{fi}}_{c,\varepsilon}(x)|x|]\big)\big\}\lambda(dy)K(dx) \end{aligned} \tag{12.2.3}$$

where

$$b^{\text{fi}}_{c,\varepsilon}(x) = 0 \qquad \text{if} \quad \varepsilon_n(x) \geq 1 - \big[2\Phi(\tau_n|x|)\big]^{-1} \tag{12.2.4}$$

and

$$\frac{\varepsilon_n(x)}{1 - \varepsilon_n(x)} = \exp\big(-2\tau_n|x|b^{\text{fi}}_{c,\varepsilon}(x)\big)\Phi\big(\tau_n|x| - b^{\text{fi}}_{c,\varepsilon}(x)\big) - \Phi\big(-\tau_n|x| - b^{\text{fi}}_{c,\varepsilon}(x)\big) \tag{12.2.5}$$

if $\varepsilon_n(x) < 1 - \big[2\Phi(\tau_n|x|)\big]^{-1}$. Furthermore, condition (10.2.21) reads

$$K\Big(\varepsilon_n(x) < 1 - \big[2\Phi(\tau_n|x|)\big]^{-1}\Big) > 0 \tag{12.2.6}$$

The minimax robust test $\tilde{\phi}_{c,\varepsilon}$ for $\mathcal{U}_{c,\varepsilon}(-\tau_n)$ versus $\mathcal{U}_{c,\varepsilon}(\tau_n)$, by Subsection 10.2.3 and symmetry, is

$$\tilde{\phi}_{c,\varepsilon} = \tfrac{1}{2}\,\mathrm{I}(h_{c,\varepsilon} > 0) + \tfrac{1}{2}\,\mathrm{I}(h_{c,\varepsilon} \geq 0) \tag{12.2.7}$$

where $h_{c,\varepsilon}$ is the loglikelihood of $(R'_{-\tau,c,\varepsilon})^n$ with respect to $(R''_{\tau,c,\varepsilon})^n$ as given in (10.2.23). Thus, the minimax estimator $\tilde{S}^{\text{fi}}_{c,\varepsilon}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\psi}_{c,\varepsilon}(x_i, y_i - x_i \tilde{S}^{\text{fi}}_{c,\varepsilon}) = 0 \qquad \tilde{\psi}_{c,\varepsilon}(x, u) = xu \min\left\{1, \frac{b^{\text{fi}}_{c,\varepsilon}(x)}{|u|}\right\} \tag{12.2.8}$$

with equal randomization between the smallest and the largest solutions. The corresponding IC is

$$\mathrm{IC}_{\tilde{\psi}_{c,\varepsilon}}(x,u) = A^{\mathrm{fi}}_{c,\varepsilon}\tilde{\psi}_{c,\varepsilon}(x,u) \qquad (A^{\mathrm{fi}}_{c,\varepsilon})^{-1} = 2\int x^2\Phi\big(b^{\mathrm{fi}}_{c,\varepsilon}(x)\big)\,K(dx) - \mathcal{K} \quad (12.2.9)$$

where $\mathcal{K} = \int x^2\,K(dx)$.

12.2.1.2 Asymptotic Minimax Estimator

If

$$K\Big(\varepsilon(x) < \sqrt{\tfrac{2}{\pi}}\,\tau|x|\Big) > 0 \quad (12.2.10)$$

the IC of the asymptotic minimax estimator in case of conditional contamination neighborhoods by Theorem 10.3.9 (a) and symmetry is,

$$\tilde{\eta}_{c,\varepsilon}(x,\varepsilon) = A^{\mathrm{as}}_{c,\varepsilon}xu\min\left\{1, \frac{b^{\mathrm{as}}_{c,\varepsilon}(x)}{|x|}\right\} \quad (12.2.11)$$

with

$$(A^{\mathrm{as}}_{c,\varepsilon})^{-1} = 2\int x^2\Phi\big(b^{\mathrm{as}}_{c,\varepsilon}(x)\big)\,K(dx) - \mathcal{K} \quad (12.2.12)$$

where $\mathcal{K} = \int x^2\,K(dx)$. The clipping function $b^{\mathrm{as}}_{c,\varepsilon}(x)$ is determined by

$$b^{\mathrm{as}}_{c,\varepsilon}(x) = 0 \qquad \text{if} \quad \varepsilon(x) \geq \sqrt{\tfrac{2}{\pi}}\,\tau|x| \quad (12.2.13)$$

$$\varepsilon(x) = 2\tau|x|\big[\varphi\big(b^{\mathrm{as}}_{c,\varepsilon}(x)\big) - b^{\mathrm{as}}_{c,\varepsilon}(x)\Phi\big(-b^{\mathrm{as}}_{c,\varepsilon}(x)\big)\big] \qquad \text{if} \quad \varepsilon(x) < \sqrt{\tfrac{2}{\pi}}\,\tau|x| \quad (12.2.14)$$

As in case of unconditional neighborhoods (cf. Section 12.1), it is not clear whether the corresponding asymptotic minimax estimator $\tilde{S}^{\mathrm{as}}_{c,\varepsilon}$ can be constructed as an M estimator. Nevertheless, we can define the M estimator $S^{\mathrm{as}}_{c,\varepsilon}$ satisfying

$$\sum_{i=1}^{n}\tilde{\eta}_{c,\varepsilon}(x_i, y_i - x_i S^{\mathrm{as}}_{c,\varepsilon}) = 0 \quad (12.2.15)$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2 (cf. Theorem 10.2.9 (a)) and serves as a comparative estimator for the finite-sample minimax estimator $\tilde{S}^{\mathrm{fi}}_{c,\varepsilon}$. Moreover, we get the following relations between finite-sample and asymptotic results.

Lemma 12.2.1 *For $\varepsilon_n(x) \in (0,1)$ a.e. $K(dx)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

(a)

$$\sqrt{n}\left(1 - \big[2\Phi(\tau_n|x|)\big]^{-1}\right) < \sqrt{\tfrac{2}{\pi}}\,\tau|x| \qquad \text{a.e. } K(dx) \quad (12.2.16)$$

(b)

$$\frac{\varepsilon(x)}{1-\varepsilon_n(x)} - \sqrt{n}\,\big[\exp\big(-2\tau_n|x|b(x)\big)\Phi\big(\tau_n|x|-b(x)\big) - \Phi\big(-\tau_n|x|-b(x)\big)\big]$$
$$= \varepsilon - 2\tau|x|\big[\varphi\big(b(x)\big) - b(x)\Phi\big(-b(x)\big)\big] + \mathrm{O}(n^{-1/2}) \tag{12.2.17}$$

a.e. $K(dx)$.

(c)

$$b^{\mathrm{fi}}_{c,\varepsilon}(x) = b^{\mathrm{as}}_{c,\varepsilon}(x) + \mathrm{O}(n^{-1/2}) \qquad \text{a.e. } K(dx) \tag{12.2.18}$$

and the $\mathrm{O}(n^{-1/2})$*-corrected asymptotic optimal clipping curve, for all* $x \in \mathbb{R}\setminus\{0\}$*, is*

$$b^{\mathrm{as.c}}_{c,\varepsilon}(x) := b^{\mathrm{as}}_{c,\varepsilon}(x) - \frac{1}{\sqrt{n}}\,\frac{\varepsilon(x)\big[\varepsilon(x) + b^{\mathrm{as}}_{c,\varepsilon}(x)\tau|x|\big]}{2\tau|x|\Phi\big(-b^{\mathrm{as}}_{c,\varepsilon}(x)\big)} \tag{12.2.19}$$

where

$$b^{\mathrm{fi}}_{c,\varepsilon}(x) = b^{\mathrm{as.c}}_{c,\varepsilon}(x) + \mathrm{O}(n^{-1}) \qquad \text{a.e. } K(dx) \tag{12.2.20}$$

Proof Analogously to the proof of Lemma 11.1.3, by identifying $\tau|x|$ with τ and $b(x)$ with b for each $x \in \mathbb{R}\setminus\{0\}$, where we additionally need the assumption (10.1.5), which is $K(x=0) = 0$. ////

12.2.2 Total Variation Neighborhoods

We assume $\varepsilon = 0$; i.e., we consider the following conditional total variation neighborhoods,

$$\mathcal{U}_{v,\delta}(\theta,\delta_n) = \big\{Q \in \mathcal{M}_1(\mathbb{B}^2)\,\big|\,Q(dx) = K(dx),\ d_v\big(Q(dy\,|\,x), \mathcal{N}(x\theta,1)\big) \le \delta_n(x) \text{ a.e. } K(dx)\big\} \tag{12.2.21}$$

with given contamination curve $\delta\colon \mathbb{R} \to [0,\infty)$ such that $\delta_n(x) = \delta(x)/\sqrt{n} \in (0,1)$ a.e. $K(dx)$ and fixed sample size $n \in \mathbb{N}$.

12.2.2.1 Finite-Sample Minimax Estimator

For a given width $\tau_n = \tau/\sqrt{n} \in (0,\infty)$ a least favorable pair $(R'_{-\tau_n,v,\delta}, R''_{\tau_n,v,\delta})$ for $\mathcal{U}_{v,\delta}(-\tau_n,\delta_n)$ versus $\mathcal{U}_{v,\delta}(\tau_n,\delta_n)$ by Subsubsection 10.2.2.2 is

$$R'_{-\tau_n,v,\delta}(dx,dy) \tag{12.2.22}$$
$$= \varphi(y+\tau_n x)\begin{cases} \frac{1+\exp(2\tau_n xy)}{1+\exp(-2\tau_n b^{\mathrm{fi}}_{v,\delta}(x)|x|)} & \mathrm{sign}(x)y < -b^{\mathrm{fi}}_{v,\delta}(x) \\ 1 & |y| \le b^{\mathrm{fi}}_{v,\delta}(x) \\ \frac{1+\exp(2\tau_n xy)}{1+\exp(2\tau_n b^{\mathrm{fi}}_{v,\delta}(x)|x|)} & \mathrm{sign}(x)y > b^{\mathrm{fi}}_{v,\delta}(x) \end{cases} \lambda(dy)K(dx) \tag{12.2.23}$$

$$R''_{\tau_n,v,\delta}(dx,dy) \tag{12.2.24}$$
$$= \varphi(y-\tau_n x)\begin{cases} \frac{1+\exp(-2\tau_n xy)}{1+\exp(2\tau_n b^{\mathrm{fi}}_{v,\delta}(x)|x|)} & \mathrm{sign}(x)y < -b^{\mathrm{fi}}_{v,\delta}(x) \\ 1 & |y| \le b^{\mathrm{fi}}_{v,\delta}(x) \\ \frac{1+\exp(-2\tau_n xy)}{1+\exp(-2\tau_n b^{\mathrm{fi}}_{v,\delta}(x)|x|)} & \mathrm{sign}(x)y > b^{\mathrm{fi}}_{v,\delta}(x) \end{cases} \lambda(dy)K(dx) \tag{12.2.25}$$

where

$$b^{\mathrm{fi}}_{v,\delta}(x) = 0 \qquad \text{if} \quad \delta_n(x) \geq \Phi(\tau_n|x|) - \tfrac{1}{2} \tag{12.2.26}$$

and

$$\begin{aligned}&\big[1 + \exp\big(-2\tau_n|x|b^{\mathrm{fi}}_{v,\delta}(x)\big)\big]\delta_n(x)\\ &\quad = \exp\big(-2\tau_n|x|b^{\mathrm{fi}}_{v,\delta}(x)\big)\Phi\big(\tau_n|x| - b^{\mathrm{fi}}_{v,\delta}(x)\big) - \Phi\big(-\tau_n|x| - b^{\mathrm{fi}}_{v,\delta}(x)\big)\end{aligned} \tag{12.2.27}$$

if $\delta_n(x) < \Phi(\tau_n|x|) - \frac{1}{2}$. Furthermore, condition (10.2.21) reads

$$K\big(\delta_n(x) < \Phi(\tau_n|x|) - \tfrac{1}{2}\big) > 0 \tag{12.2.28}$$

The minimax robust test $\tilde{\phi}_{v,\delta}$ for $\mathcal{U}_{v,\delta}(-\tau_n)$ versus $\mathcal{U}_{v,\delta}(\tau_n)$, by Subsection 10.2.3 and symmetry, is

$$\tilde{\phi}_{v,\delta} = \tfrac{1}{2}\,\mathrm{I}(h_{v,\delta} > 0) + \tfrac{1}{2}\,\mathrm{I}(h_{v,\delta} \geq 0) \tag{12.2.29}$$

where $h_{v,\delta}$ is the loglikelihood of $(R'_{-\tau,v,\delta})^n$ with respect to $(R''_{\tau,v,\delta})^n$ as given in (10.2.23). Thus, the minimax estimator $\tilde{S}^{\mathrm{fi}}_{v,\delta}$ is an M estimator satisfying

$$\sum_{i=1}^{n} \tilde{\psi}_{v,\delta}(x_i, y_i - x_i\tilde{S}^{\mathrm{fi}}_{v,\delta}) = 0 \qquad \tilde{\psi}_{v,\delta}(x,u) = xu\min\left\{1, \frac{b^{\mathrm{fi}}_{v,\delta}(x)}{|u|}\right\} \tag{12.2.30}$$

with equal randomization between the smallest and the largest solutions. The corresponding IC is

$$\mathrm{IC}_{\tilde{\psi}_{v,\delta}}(x,u) = A^{\mathrm{fi}}_{v,\delta}\tilde{\psi}_{v,\delta}(x,u) \qquad (A^{\mathrm{fi}}_{v,\delta})^{-1} = 2\int x^2\Phi\big(b^{\mathrm{fi}}_{v,\delta}(x)\big)\,K(dx) - \mathcal{K} \tag{12.2.31}$$

where $\mathcal{K} = \int x^2\,K(dx)$.

12.2.2.2 Asymptotic Minimax Estimator

If

$$K\Big(\delta(x) < \tfrac{1}{\sqrt{2\pi}}\tau|x|\Big) > 0 \tag{12.2.32}$$

the IC of the asymptotic minimax estimator in case of conditional contamination neighborhoods by Theorem 10.3.9 (a) and symmetry is,

$$\tilde{\eta}_{v,\delta}(x,\varepsilon) = A^{\mathrm{as}}_{v,\delta}xu\min\left\{1, \frac{b^{\mathrm{as}}_{v,\delta}(x)}{|x|}\right\} \tag{12.2.33}$$

with

$$(A^{\mathrm{as}}_{v,\delta})^{-1} = 2\int x^2\Phi\big(b^{\mathrm{as}}_{v,\delta}(x)\big)\,K(dx) - \mathcal{K} \tag{12.2.34}$$

where $\mathcal{K} = \int x^2\,K(dx)$. The clipping function $b^{\mathrm{as}}_{v,\delta}(x)$ is determined by

$$b^{\mathrm{as}}_{v,\delta}(x) = 0 \qquad \text{if} \quad \delta(x) \geq \tfrac{1}{\sqrt{2\pi}}\tau|x| \tag{12.2.35}$$

$$\delta(x) = \tau|x|\big[\varphi\big(b^{\mathrm{as}}_{c,\varepsilon}(x)\big) - b^{\mathrm{as}}_{v,\delta}(x)\Phi\big(-b^{\mathrm{as}}_{v,\delta}(x)\big)\big] \qquad \text{if} \quad \delta(x) < \tfrac{1}{\sqrt{2\pi}}\tau|x| \tag{12.2.36}$$

As in case of conditional contamination neighborhoods (cf. Subsection 12.2.1), it is not clear whether the corresponding asymptotic minimax estimator $\tilde{S}^{\mathrm{as}}_{v,\delta}$ can be constructed as an M estimator. Nevertheless, we can again define the M estimator $S^{\mathrm{as}}_{v,\delta}$ satisfying

$$\sum_{i=1}^{n} \tilde{\eta}_{v,\delta}(x_i, y_i - x_i S^{\mathrm{as}}_{v,\delta}) = 0 \tag{12.2.37}$$

with equal randomization between the smallest and the largest solutions. This estimator is equivariant in sense of Definition 10.2.2 (cf. Theorem 10.2.9 (a)) and serves as a comparative estimator for the finite-sample minimax estimator $\tilde{S}^{\mathrm{fi}}_{v,\delta}$. Moreover, we get the following relations between finite-sample and asymptotic results.

Lemma 12.2.2 *For $\delta_n(x) \in (0,1)$ a.e. $K(dx)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

(a)

$$\sqrt{n}\,\big(\Phi(\tau_n|x|) - \tfrac{1}{2}\big) < \tfrac{1}{\sqrt{2\pi}}\,\tau|x| \qquad \text{a.e. } K(dx) \tag{12.2.38}$$

(b)

$$\begin{aligned}&\big[1 + \exp\big(-2\tau_n|x|b(x)\big)\big]\delta(x)\\ &\qquad -\sqrt{n}\,\big[\exp\big(-2\tau_n|x|b(x)\big)\Phi\big(\tau_n|x| - b(x)\big) - \Phi\big(-\tau_n|x| - b(x)\big)\big]\\ &= \delta(x) - \tau|x|\big[\varphi\big(b(x)\big) - b(x)\Phi\big(-b(x)\big)\big] + \mathrm{O}(n^{-1/2})\end{aligned} \tag{12.2.39}$$

a.e. $K(dx)$.

(c)

$$b^{\mathrm{as}}_{v,\delta}(x) = b^{\mathrm{fi}}_{v,\delta}(x) + \mathrm{O}(n^{-1}) \qquad \text{a.e. } K(dx) \tag{12.2.40}$$

and the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping curve, for all $x \in \mathbb{R}\setminus\{0\}$, is

$$b^{\mathrm{as.c}}_{v,\delta}(x) = b^{\mathrm{as}}_{v,\delta}(x) - \frac{1}{n}\,\frac{\tau\big[2\big(b^{\mathrm{as}}_{v,\delta}(x)\big)^2\delta(x) + \tau\varphi(b^{\mathrm{as}}_{v,\delta}(x))\big]}{6\Phi(-b^{\mathrm{as}}_{v,\delta}(x))} \tag{12.2.41}$$

where

$$b^{\mathrm{fi}}_{v,\delta}(x) = b^{\mathrm{as.c}}_{v,\delta}(x) + \mathrm{O}(n^{-3/2}) \qquad \text{a.e. } K(dx) \tag{12.2.42}$$

Proof Analogously to the proof of Lemma 11.2.3, by identifying $\tau|x|$ with τ and $b(x)$ with b for each $x \in \mathbb{R}\setminus\{0\}$, where we additionally need the assumption (10.1.5), which is $K(x=0) = 0$. ////

12.2.3 Computation of the Finite-Sample Risk

12.2.3.1 Finite-Sample Risk

We fix $n \in \mathbb{N}$ and radius curve $\varepsilon_n(x)\colon \mathbb{R} \to [0,1)$ $(* = c)$, respectively radius curve $\delta_n(x)\colon \mathbb{R} \to [0,1)$ $(* = v)$. Given some clipping function $b(x)\colon \mathbb{R} \to (0,\infty)$ we then want to determine the finite-sample risk of an M estimator S satisfying

$$\sum_{i=1}^{n} \chi_0(x_i, y_i - x_i S) = 0 \qquad \chi_0(x,u) = xu\min\left\{1, \frac{b(x)}{|u|}\right\} \tag{12.2.43}$$

with equal randomization between the smallest and the largest solutions. In particular, we are interested in the finite-sample minimax estimator $\tilde{S}_*^{\mathrm{fi}}$ and the estimator S_*^{as}, respectively. Moreover, we study the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected $(* = c)$, respectively $\mathrm{O}(n^{-1})$-corrected $(* = v)$ asymptotic optimal clipping function. All these estimators are equivariant in sense of Definition 10.2.2; confer Theorem 10.2.9. As already stated in Remark 10.2.3, the finite-sample risk of any equivariant estimator S for given $\tau_n \in (0, \infty)$ reads as given in (12.1.66). Moreover, we have the inclusions (12.1.67) and (12.1.68). Thus, we can proceed analogously to Subsubsection 12.1.3.1 and obtain (12.1.71)–(12.1.76).

By monotonicity of χ_0 the distribution of $\sum_{i=1}^n \chi_0(x_i, y_i)$ is monotone with respect to stochastic order; i.e., the probability of $\sum_{i=1}^n \chi_0(x_i, y_i) > 0$, respectively of $\sum_{i=1}^n \chi_0(x_i, y_i) \geq 0$ under $Q_{-\tau_n} \in \mathcal{U}_{*,\varepsilon}(-\tau_n)$ is maximal if

$$Q_{-\tau_n}\big(\chi_0(x, y) = |x| b(x)\big) = \max! \tag{12.2.44}$$

where $Q_{-\tau_n}(dx) = K(dx)$ and $K(x = 0) = 0$; i.e., (10.2.6) holds. Moreover, we get

$$\begin{aligned} &Q_{-\tau_n}\big(\chi_0(x, y) = |x| b(x)\big) \\ &\quad = Q_{-\tau_n}\big(y \geq b(x),\, x > 0\big) + Q_{-\tau_n}\big(y \leq -b(x),\, x < 0\big) && (12.2.45) \\ &\quad \overset{(10.1.12)}{=} \int_0^\infty \int_{b(x)+\tau_n x}^\infty Q_0(dy \,|\, x)\, K(dx) + \int_{-\infty}^0 \int_{-\infty}^{-b(x)+\tau_n x} Q_0(dy \,|\, x)\, K(dx) && (12.2.46) \end{aligned}$$

Thus, given $x \in \mathbb{R}$, we can argue similar to the location case (cf. Subsection 11.3.1) and obtain that the maximum is achieved for any Markov kernel $H : \mathbb{B} \times \mathbb{R} \to [0, 1]$ such that $H'_{-\tau_n}(dy \,|\, x)$ is concentrated on $[b(x) + \tau_n x, \infty)$ if $x > 0$, respectively on $(-\infty, -b(x) + \tau_n x]$ if $x < 0$. In case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$ this leads us to

$$Q'_{-\tau_n}(dy \,|\, x) = \big(1 - \varepsilon_n(x)\big)\mathcal{N}(-\tau_n x, 1)(dy) + \varepsilon_n(x) H'_{-\tau_n}(dy \,|\, x) \qquad \text{a.e. } K(dx) \tag{12.2.47}$$

In case of conditional total variation neighborhoods $(* = v,\, t = \delta)$ we obtain for all $t \in \mathbb{R}$

$$Q'_{-\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) = \big(\Phi(t + \tau_n x) - \delta_n(x)\big)_+ + \delta_n(x) H'_{-\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) \tag{12.2.48}$$

a.e. $K(dx)$ if $x > 0$ and

$$Q'_{-\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) = \min\big\{\Phi(t + \tau_n x) + \delta_n(x) H'_{-\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big),\, 1\big\} \tag{12.2.49}$$

a.e. $K(dx)$ if $x < 0$.

Remark 12.2.3 If

$$M_1 := \sup{}_K\big\{x > 0 \,\big|\, b(x) + \tau_n x\big\} < \infty \tag{12.2.50}$$

respectively

$$M_2 := \inf{}_K\big\{x < 0 \,\big|\, -b(x) + \tau_n x\big\} > -\infty \tag{12.2.51}$$

we can choose $H'_{\tau_n} : \mathbb{B} \times \mathbb{R} \to [0,1]$ such that $H'_{-\tau_n}(dy \,|\, x)$ is concentrated on $[M_1, \infty)$ if $x > 0$, respectively on $(-\infty, M_2]$ if $x < 0$. ////

Analogously, we obtain the probability of $\sum_{i=1}^n \chi_0(x_i, y_i) < 0$, respectively of $\sum_{i=1}^n \chi_0(x_i, y_i) \le 0$ under $Q_{\tau_n} \in \mathcal{U}_{*,\varepsilon}(\tau_n)$ is maximal if

$$Q_{\tau_n}(\chi_0(x,y) = -|x|b(x)) = \max! \tag{12.2.52}$$

where again $Q_{\tau_n}(dx) = K(dx)$ and (10.2.6) holds. By analogous arguments as in case of $Q_{-\tau_n}$ we get for conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$

$$Q''_{\tau_n}(dy \,|\, x) = (1 - \varepsilon_n(x))\mathcal{N}(\tau_n x, 1)(dy) + \varepsilon_n(x) H''_{\tau_n}(dy \,|\, x) \qquad \text{a.e. } K(dx) \tag{12.2.53}$$

Similarly, we obtain in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$, for all $t \in \mathbb{R}$

$$Q''_{\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) = \min\big\{\Phi(t - \tau_n x) + \delta_n(x) H''_{\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big),\, 1\big\} \tag{12.2.54}$$

a.e. $K(dx)$ if $x > 0$ and

$$Q''_{\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) = \big(\Phi(t - \tau_n x) - \delta_n(x)\big)_+ + \delta_n(x) H''_{\tau_n}\big(dy \in (-\infty, t] \,\big|\, x\big) \tag{12.2.55}$$

a.e. $K(dx)$ if $x < 0$. To achieve the maximum in (12.2.52), $H''_{\tau_n}(dy \,|\, x)$ has to be concentrated on $(-\infty, -b(x) - \tau_n x]$ if $x > 0$, respectively on $[b(x) - \tau_n x, \infty)$ if $x < 0$.

Remark 12.2.4 **(a)** If

$$M_1 := \inf{}_K\big\{x > 0 \,\big|\, -b(x) - \tau_n x\big\} > -\infty \tag{12.2.56}$$

respectively

$$M_2 := \sup{}_K\big\{x < 0 \,\big|\, b(x) - \tau_n x\big\} < \infty \tag{12.2.57}$$

we can choose $H''_{\tau_n} : \mathbb{B} \times \mathbb{R} \to [0,1]$ such that $H''_{\tau_n}(dy \,|\, x)$ is concentrated on $(-\infty, M_1]$ if $x > 0$, respectively on $[M_2, \infty)$ if $x < 0$.

(b) In view of Remark 12.1.5 we additionally have to make sure that $H'_{-\tau_n}(dy \,|\, x)$ and $H''_{\tau_n}(dy \,|\, x)$ are absolutely continuous a.e. $K(dx)$. ////

We may now calculate the finite-sample risk by determining the distribution of $\sum_{i=1}^n \chi_0(x_i, y_i)$ under $\big(Q'_{-\tau_n}\big)^n$ or $\big(Q''_{\tau_n}\big)^n$, respectively. As a first step, we state the conditional distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n}. In case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$ we have, a.e. $K(dx)$

$$Q'_{-\tau_n}\big(\chi_0(x,y) = -|x|b(x) \,\big|\, x\big) = \big(1 - \varepsilon_n(x)\big)\Phi\big(-b(x) + \tau_n|x|\big) \tag{12.2.58}$$

$$\begin{aligned} &Q'_{-\tau_n}\big(-|x|b(x) < \chi_0(x,y) < |x|t \,\big|\, x\big) \\ &\quad = \big(1 - \varepsilon_n(x)\big)\Big[\Phi\big(t + \tau_n|x|\big) - \Phi(-b(x) + \tau_n|x|)\Big] \quad t \in \big(-b(x),\, b(x)\big) \end{aligned} \tag{12.2.59}$$

$$Q'_{-\tau_n}\big(\chi_0(x,y) = |x|b(x) \,\big|\, x\big) = \big(1 - \varepsilon_n(x)\big)\Phi\big(-b(x) - \tau_n|x|\big) + \varepsilon_n(x) \tag{12.2.60}$$

respectively

$$Q''_{\tau_n}\big(\chi_0(x,y) = -|x|b(x) \,\big|\, x\big) = \big(1-\varepsilon_n(x)\big)\Phi\big(-b(x)-\tau_n|x|\big) + \varepsilon_n(x) \tag{12.2.61}$$

$$\begin{aligned} &Q''_{\tau_n}\big(-|x|b(x) < \chi_0(x,y) < |x|t \,\big|\, x\big) \\ &\quad = \big(1-\varepsilon_n(x)\big)\Big[\Phi\big(t-\tau_n|x|\big) - \Phi(-b(x)-\tau_n|x|)\Big] \quad t \in \big(-b(x),\, b(x)\big) \end{aligned} \tag{12.2.62}$$

$$Q''_{\tau_n}\big(\chi_0(x,y) = |x|b(x)\big) = \big(1-\varepsilon_n(x)\big)\Phi\big(-b(x)+\tau_n|x|\big) \tag{12.2.63}$$

In case of conditional total variation neighborhoods ($* = v$, $t = \delta$) we obtain, a.e. $K(dx)$,

$$Q'_{-\tau_n}\big(\chi_0(x,y) = -|x|b(x) \,\big|\, x\big) = \Big(\Phi\big(-b(x)+\tau_n|x|\big) - \delta_n(x)\Big)_+ \tag{12.2.64}$$

$$\begin{aligned} &Q'_{-\tau_n}\big(-|x|b(x) < \chi_0(x,y) < |x|t \,\big|\, x\big) \\ &\quad = \Big(\Phi\big(t+\tau_n|x|\big) - \delta_n(x)\Big)_+ \\ &\qquad - \Big(\Phi\big(-b(x)+\tau_n|x|\big) - \delta_n(x)\Big)_+ \qquad t \in (-b(x),\, b(x)) \end{aligned} \tag{12.2.65}$$

$$Q'_{-\tau_n}\big(\chi_0(x,y) = |x|b(x) \,\big|\, x\big) = 1 - \Big(\Phi\big(b(x)+\tau_n|x|\big) - \delta_n(x)\Big)_+ \tag{12.2.66}$$

respectively

$$Q''_{\tau_n}\big(\chi_0(x,y) = -|x|b(x) \,\big|\, x\big) = \min\Big\{\Phi\big(-b(x)-\tau_n|x|\big) + \delta_n(x),\, 1\Big\} \tag{12.2.67}$$

$$\begin{aligned} &Q''_{\tau_n}\big(-|x|b(x) < \chi_0(x,y) < |x|t \,\big|\, x\big) \\ &\quad = \min\Big\{\Phi\big(t-\tau_n|x|\big) + \delta_n(x),\, 1\Big\} \\ &\qquad - \min\Big\{\Phi\big(-b(x)-\tau_n|x|\big) + \delta_n(x),\, 1\Big\} \quad t \in (-b(x),\, b(x)) \end{aligned} \tag{12.2.68}$$

$$Q''_{\tau_n}\big(\chi_0(x,y) = |x|b(x) \,\big|\, x\big) = 1 - \min\Big\{\Phi\big(b(x)-\tau_n|x|\big) + \delta_n(x),\, 1\Big\} \tag{12.2.69}$$

12.2.3.2 Numerical Algorithms

We determine the finite-sample risk by computing a numerical approximation of the distribution of $\sum_{i=1}^n \chi_0(x_i,y_i)$ under $\big(Q'_{-\tau_n}\big)^n$ and $\big(Q''_{\tau_n}\big)^n$, respectively. We work with an algorithm which is analogous to Algorithm A given in Subsubsection 11.3.2.1; i.e., this algorithm directly uses the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n} where the corresponding conditional distributions, given $x \in \mathbb{R}$, can be read off from (12.2.58)–(12.2.69). The n-fold convolution of the distribution of χ_0 under $Q'_{-\tau_n}$, respectively Q''_{τ_n} is calculated using Algorithm C.2.2 which is based on the fast Fourier transform; confer Appendix C.2.

Remark 12.2.5 We did not implement an algorithm which is the analogon to Algorithm B given in Subsubsection 11.3.2.2 as such an algorithm in the current conditional setup is much more complicated than in the situations before and therefore is probably very slow. ////

We expect the precision of Algorithms A in case of simple linear regression and conditional neighborhoods to be similar to the location case (cf. Subsubsection 11.3.2.3) but we do not have analytical results to check this. Like in case of unconditional neighborhoods (cf. Subsubsection 12.1.3.2) we made several cross checks. We for instance choose $K = \mathrm{I}_{\{1\}}$, respectively $K = \mathrm{Unif}\,([0.95, 1.05])$ such that we could expect results that are identical, respectively close to the location case. In the second case we work with a constant radius curve ε, respectively δ. Moreover, we do some numerical simulations to check the algorithms. That is, we simulate $1\mathrm{e}\,06$ samples of size n and compute the empirical finite-sample risk and the corresponding 95% confidence interval. As it turns out, the results of Algorithms A always lie well within the 95% confidence interval; confer Tables 12.5–12.8. We also give the computation time in seconds T_A of Algorithm A on an Athlon with 2.5 GHz and 512 MB RAM. To restrict the amount of results we consider only the finite-sample risk of the estimator $S^{\mathrm{as}}_{*,\varepsilon}$, respectively $S^{\mathrm{as}}_{*,\delta}$ and choose one "typical" situation. That is, we fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon(x) \equiv 0.2$, respectively $\delta(x) \equiv 0.1$. $\mathrm{Risk}_{\mathrm{A}}$ denotes the finite-sample risk of $S^{\mathrm{as}}_{*,\varepsilon}$, respectively $S^{\mathrm{as}}_{*,\delta}$ computed with Algorithm A and 2^q stands for the number of lattice points used in step 2 of Algorithm C.2.2. The first two tables (see Tables 12.5 and 12.6) show the results for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and conditional contamination, respectively total variation neighborhoods. The remaining two tables (see Tables 12.7 and 12.8) contain the results for $K = \mathrm{Unif}\,([-1, 2])$ and conditional contamination, respectively total variation neighborhoods.

n	emp. risk	95% conf. int.	q	$\mathrm{Risk_A}$	T_A
2	0.1721	[0.1714, 0.1729]	10	0.171588	1.80
			12	0.171645	5.94
			14	0.171648	22.60
3	0.1188	[0.1182, 0.1195]	10	0.118494	1.81
			12	0.118521	6.06
			14	0.118522	22.61
4	0.0917	[0.0911, 0.0923]	10	0.091565	1.81
			12	0.091586	5.96
			14	0.091592	22.14
5	0.0728	[0.0723, 0.0733]	10	0.072981	1.79
			12	0.073001	5.77
			14	0.073006	22.33
10	0.0374	[0.0370, 0.0378]	10	0.037282	1.82
			12	0.037297	5.77
			14	0.037300	22.71

Table 12.5: A comparison between Algorithm A and empirical results for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ in case of conditional contamination neighborhoods $* = c$, $t = \varepsilon$). [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$, and $\varepsilon(x) \equiv 0.2$. Risk_A denotes the finite-sample risk of $S^{\mathrm{as}}_{c,\varepsilon}$ computed with Algorithm A, 2^q the number of lattice points used in Algorithm C.2.2 and T_A the computation time in seconds for Algorithm A.]

n	emp. risk	95% conf. int.	q	$\mathrm{Risk_A}$	T_A
2	0.0979	[0.0973, 0.0985]	10	0.097985	1.89
			12	0.098013	5.58
			14	0.098015	21.62
3	0.0653	[0.0648, 0.0658]	10	0.065130	1.70
			12	0.065143	5.64
			14	0.065144	21.78
4	0.0485	[0.0481, 0.0489]	10	0.048696	1.88
			12	0.048704	5.67
			14	0.048708	21.77
5	0.0383	[0.0379, 0.0387]	10	0.038228	1.70
			12	0.038235	5.70
			14	0.038239	21.74
10	0.0198	[0.0195, 0.0201]	10	0.019758	1.72
			12	0.019763	5.80
			14	0.019765	22.46

Table 12.6: A comparison between Algorithm A and empirical results for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$, and $\delta(x) \equiv 0.1$. Risk_A denotes the finite-sample risk of $S^{\mathrm{as}}_{v,\delta}$ computed with Algorithm A, 2^q the number of lattice points used in Algorithm C.2.2 and T_A the computation time in seconds for Algorithm A.]

n	emp. risk	95% conf. int.	q	Risk_A	T_A
2	0.1957	[0.1948, 0.1964]	10	0.195295	17.24
			12	0.195272	67.51
			14	0.195274	270.41
3	0.1510	[0.1503, 0.1517]	10	0.151148	16.66
			12	0.151142	65.86
			14	0.151152	265.21
4	0.1230	[0.1223, 0.1236]	10	0.122928	16.81
			12	0.122912	67.66
			14	0.122918	260.55
5	0.1036	[0.1030, 0.1042]	10	0.103321	16.28
			12	0.103302	63.87
			14	0.103314	257.29

Table 12.7: A comparison between Algorithm A and empirical results for $K = \text{Unif}([-1, 2])$ in case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$, and $\varepsilon(x) \equiv 0.2$. Risk_A denotes the finite-sample risk of $S^{\text{as}}_{c,\varepsilon}$ computed with Algorithm A, 2^q the number of lattice points used in Algorithm C.2.2 and T_A the computation time in seconds for Algorithm A.]

n	emp. risk	95% conf. int.	q	Risk_A	T_A
2	0.1294	[0.1287, 0.1300]	10	0.129224	24.28
			12	0.129201	97.81
			14	0.129186	381.60
3	0.0961	[0.0955, 0.0966]	10	0.095924	23.21
			12	0.095917	91.81
			14	0.095915	370.74
4	0.0763	[0.0758, 0.0769]	10	0.076152	22.87
			12	0.076154	90.61
			14	0.076148	365.84
5	0.0629	[0.0625, 0.0634]	10	0.063095	22.35
			12	0.063096	88.53
			14	0.063091	364.61

Table 12.8: A comparison between Algorithm A and empirical results for $K = \text{Unif}([-1, 2])$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$. [We fix $\tau = \Phi^{-1}(0.995) \approx 2.576$, and $\delta(x) \equiv 0.1$. Risk_A denotes the finite-sample risk of $S^{\text{as}}_{v,\delta}$ computed with Algorithm A, 2^q the number of lattice points used in Algorithm C.2.2 and T_A the computation time in seconds for Algorithm A.]

12.3 Numerical Comparisons

In this section we present various numerical comparisons between finite-sample and asymptotic results; i.e., we numerically check the asymptotics against finite-sample results obtained for fixed unconditional, respectively conditional regression neighborhoods. To restrict the amount of results, we choose $\tau = \Phi^{-1}(0.95)$, $\Phi^{-1}(0.975)$, $\Phi^{-1}(0.995)$ such that 2τ corresponds to the width of 90%, 95%, 99%-confidence intervals in case of the standard normal distribution. Moreover, in most cases we use $q = 12$ in Algorithms A and B. As it turns out, the results are very similar to the results obtained in normal location; confer Section 11.4.

12.3.1 Unconditional Contamination Neighborhoods

12.3.1.1 Optimal Clipping Bound

We fix radius $\varepsilon = 0.1, 0.5$ and compute $b_{c,0}^{\mathrm{fi}}$, $b_{c,0}^{\mathrm{as}}$ and $b_{c,0}^{\mathrm{as,c}}$ by (12.1.3), (12.1.10) and (12.1.16) for sample size $n \leq 25$; confer Figures 12.3 and 12.4. In all cases considered the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is too small compared to the finite-sample optimal clipping bound. For small n the $\mathrm{O}(n^{-1/2})$-correction can even lead to irregular (negative) clipping bounds; confer also Tables 12.9 and 12.10. However, for not too large width τ and moderate sample sizes n it comes very close to the finite-sample one. In particular, if the radius is not too large.

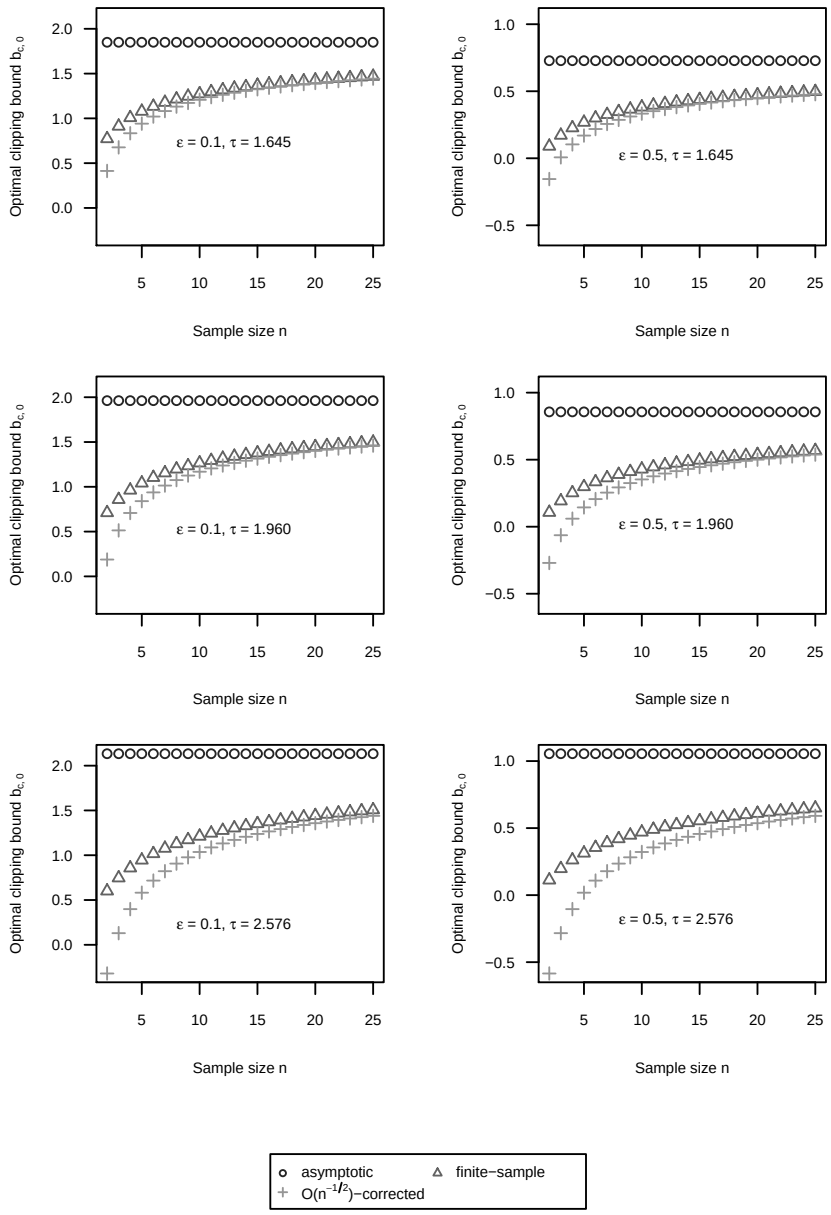

Figure 12.3: Optimal clipping bounds for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \leq 25$, $\varepsilon = 0.1, 0.5$ and $\tau = 1.645, 1.960, 2.576$ in case of unconditional contamination neighborhoods $(* = c, t = 0)$.

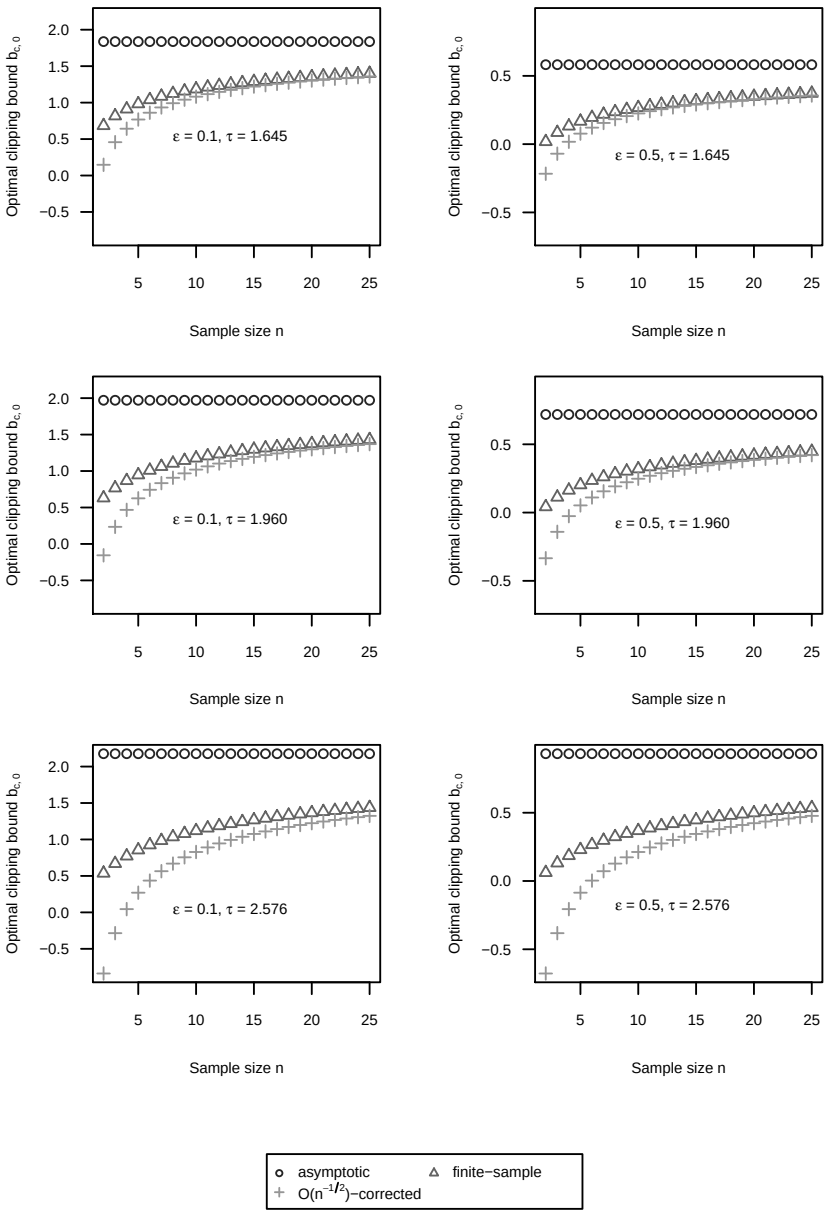

Figure 12.4: Optimal clipping bounds for $K = \text{Unif}([-1,2])$, $n \leq 25$, $\varepsilon = 0.1, 0.5$ and $\tau = 1.645, 1.960, 2.576$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

12.3.1.2 Finite-Sample Risk

We fix radius $\varepsilon = 0.1, 0.5$ and determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_{c,0}^{\mathrm{fi}}$, the asymptotic minimax estimator $\tilde{S}_{c,0}^{\mathrm{as}}$ and the estimator $S_{c,0}^{\mathrm{as,c}}$ which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound. Although there are clear differences between the clipping bounds of the finite-sample and the asymptotic minimax estimator, the differences (in absolute values) concerning the corresponding finite-sample risks are only small; see Figures 12.5 and 12.6 and Tables 12.9 and 12.10. Moreover, for moderate sample sizes n the finite-sample risk of the estimator which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is very close to the finite-sample risk of the finite-sample minimax estimator.
Figures 12.7 and 12.9 show the speed of convergence for the finite-sample risks. It seems to be of order $n^{-1/2}$ as in case of the optimal clipping bounds; confer Lemma 12.1.1. To get a better impression, we consider $y = \mathrm{Risk}_{\mathrm{as}}^{\natural} - \mathrm{Risk}_{\mathrm{fi}}^{\natural}$ and apply the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002); i.e., we estimate lambda by means of maximum likelihood such that $y^{\mathrm{lambda}} \approx n$. That is, lambda ≈ -2 indicates $y = \mathrm{O}(n^{-1/2})$. Indeed, the estimated values of lambda are close to -2 which confirms our conjecture that we have a convergence of order $n^{-1/2}$; see Figures 12.8 and 12.10. Moreover, the speed seems to depend on τ and ε. A larger τ leads to a slightly faster convergence whereas a larger ε slightly decreases the speed of convergence.

So far, we compared absolute values for given radius ε and given width τ. Now, we study relative risks; i.e., how much efficiency do we lose if we take the asymptotic optimal clipping bound, respectively the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound instead of the finite-sample optimal. More precisely, given some estimator S, we consider

$$\mathrm{relRisk}(S; c, 0) = \frac{\mathrm{Risk}(S; c, 0)}{\mathrm{Risk}(\tilde{S}_c^{\mathrm{fi}}; c, 0)} \tag{12.3.1}$$

and determine the radius $\varepsilon_{\mathrm{as}}$ and $\varepsilon_{\mathrm{as.c}}$, respectively, such that the relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound is maximal, respectively. To obtain valid results in case of the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound, we replace negative values of $b_{c,0}^{\mathrm{as,c}}$ by 0. As the numerical results show (see Tables 12.11 and 12.12), the maximum relative risks of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound are very small. Moreover, if we choose the width τ not too large and if we have a sample of moderate size (about $20 - 50$), there is no big difference between the finite-sample risks of all three estimators. These results are important if we know the radius.

But, in most applications the neighborhood radius ε is unknown ($\rho = 0$) or unknown except to belong to some radius interval ($\rho = 3, 2$), respectively. Thus, we also determine the least favorable radii for the finite-sample minimax estimator; confer Tables 12.13 and 12.14. The corresponding radii are defined analogously to Section 2.2. Since the calculations of ε_3 and ε_2 are quite time consuming, we choose $q = 10$ in these two cases, for ε_0 we can use $q = 12$ as in the computa-

tions before. For $n < 10$ the least favorable radii and the corresponding relative risks behave different for even and odd sample size, respectively. This is probably caused by the fact that there is mass at zero in case of an even sample size; confer Remark 11.3.3. However, this mass decays exponentially in n and therefore the effect disappears very fast for increasing sample size. Moreover, in all cases ε_0 is small compared to $\varepsilon_{\max}^{\mathrm{fi}}$. The relative risks for $\rho = 0$ in all cases considered stay well below 60%. In most cases ε_3 and ε_2 are close to $\varepsilon_{\max}^{\mathrm{fi}}/3$ and $\varepsilon_{\max}^{\mathrm{fi}}/2$, respectively. But, the corresponding relative risks ($\rho = 3, 2$) are small and in all cases stay well below 26% and 13%, respectively. Hence, in case that the true radius is completely unknown, respectively unknown except to belong to some radius interval, we strongly recommend to use the finite-sample minimax estimator which is given by the corresponding least favorable radius.

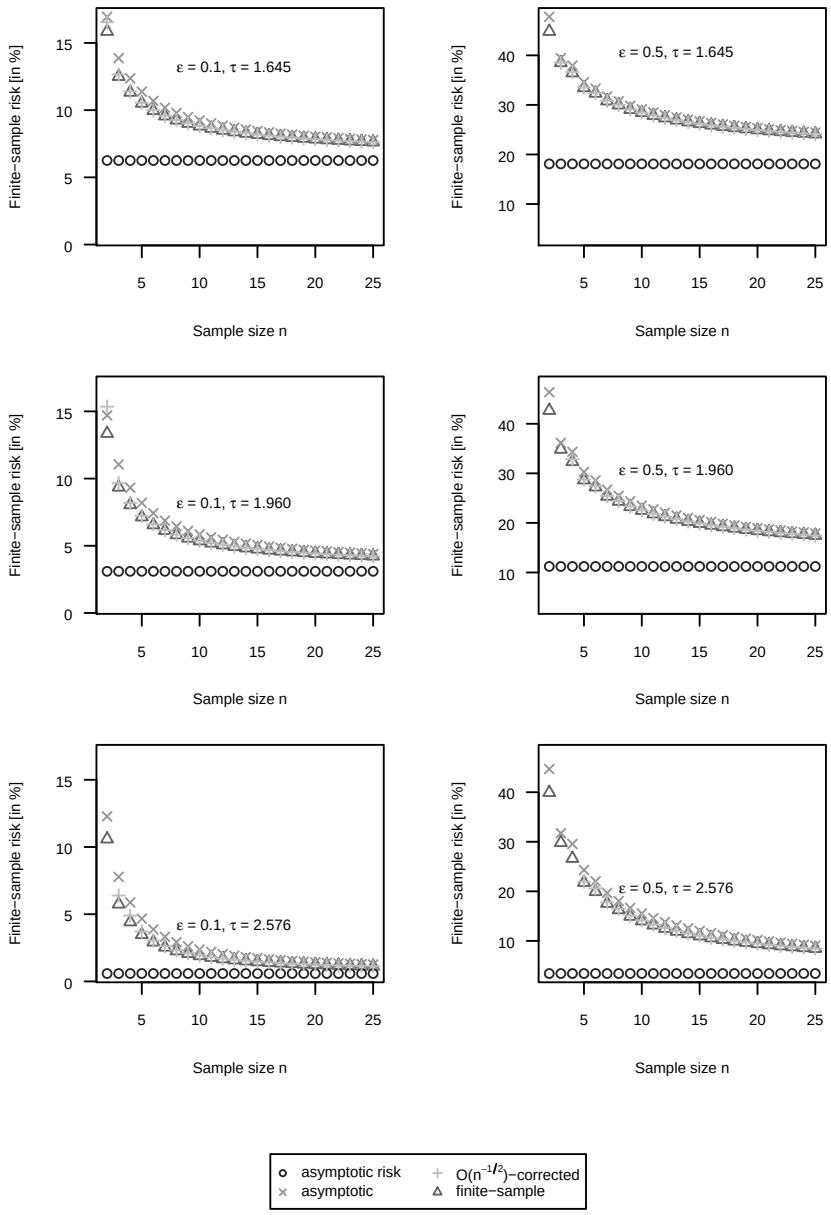

Figure 12.5: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \le 25$, $\varepsilon = 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

ε	τ	n	$b_{c,0}^{\mathrm{fi}}$	$b_{c,0}^{\mathrm{as.c}}$	$b_{c,0}^{\mathrm{as}}$	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.1	1.645	2	0.773	0.413	1.850	15.856	16.536	16.943
		3	0.912	0.677		12.501	12.655	13.866
		4	1.008	0.834		11.325	11.388	12.365
		5	1.079	0.941		10.510	10.550	11.374
		10	1.274	1.207		8.810	8.817	9.231
		50	1.575	1.562		7.138	7.138	7.207
		100	1.653	1.647		6.838	6.838	6.870
	1.960	2	0.711	0.188	1.962	13.356	15.351	14.700
		3	0.858	0.513		9.358	9.674	11.048
		4	0.962	0.708		8.049	8.181	9.319
		5	1.041	0.840		7.141	7.222	8.191
		10	1.266	1.169		5.361	5.373	5.848
		50	1.626	1.607		3.803	3.804	3.874
		100	1.720	1.711		3.553	3.553	3.584
	2.576	2	0.600	−0.320	2.134	10.591	—	12.278
		3	0.746	0.130		5.746	6.387	7.777
		4	0.857	0.398		4.424	4.898	5.873
		5	0.945	0.582		3.478	3.712	4.662
		10	1.211	1.036		1.905	1.929	2.377
		50	1.676	1.643		0.894	0.894	0.941
		100	1.803	1.787		0.773	0.773	0.792
0.5	1.645	2	0.089	−0.155	0.728	44.818	—	47.771
		3	0.171	0.007		38.501	38.515	39.433
		4	0.227	0.104		36.453	36.642	37.942
		5	0.268	0.169		33.461	33.496	34.565
		10	0.382	0.333		28.445	28.458	29.108
		50	0.561	0.551		22.152	22.152	22.303
		100	0.608	0.603		20.860	20.860	20.935
	1.960	2	0.107	−0.270	0.857	42.676	—	46.361
		3	0.192	−0.064		34.839	—	36.166
		4	0.252	0.060		32.347	32.857	34.345
		5	0.298	0.144		28.656	28.737	30.271
		10	0.429	0.353		22.508	22.542	23.481
		50	0.646	0.631		15.308	15.309	15.516
		100	0.705	0.697		13.944	13.944	14.045
	2.576	2	0.112	−0.585	1.055	39.922	—	44.670
		3	0.198	−0.284		29.830	—	31.730
		4	0.262	−0.104		26.634	—	29.565
		5	0.313	0.018		21.783	21.954	24.310
		10	0.470	0.322		14.024	14.139	15.534
		50	0.756	0.727		6.440	6.442	6.698
		100	0.837	0.823		5.310	5.310	5.424

Table 12.9: Comparison of the optimal clipping bounds and the corresponding finite-sample risks for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ in case of unconditional contamination neighborhoods $(* = c, t = 0)$. [The finite-sample risks are given in percent.]

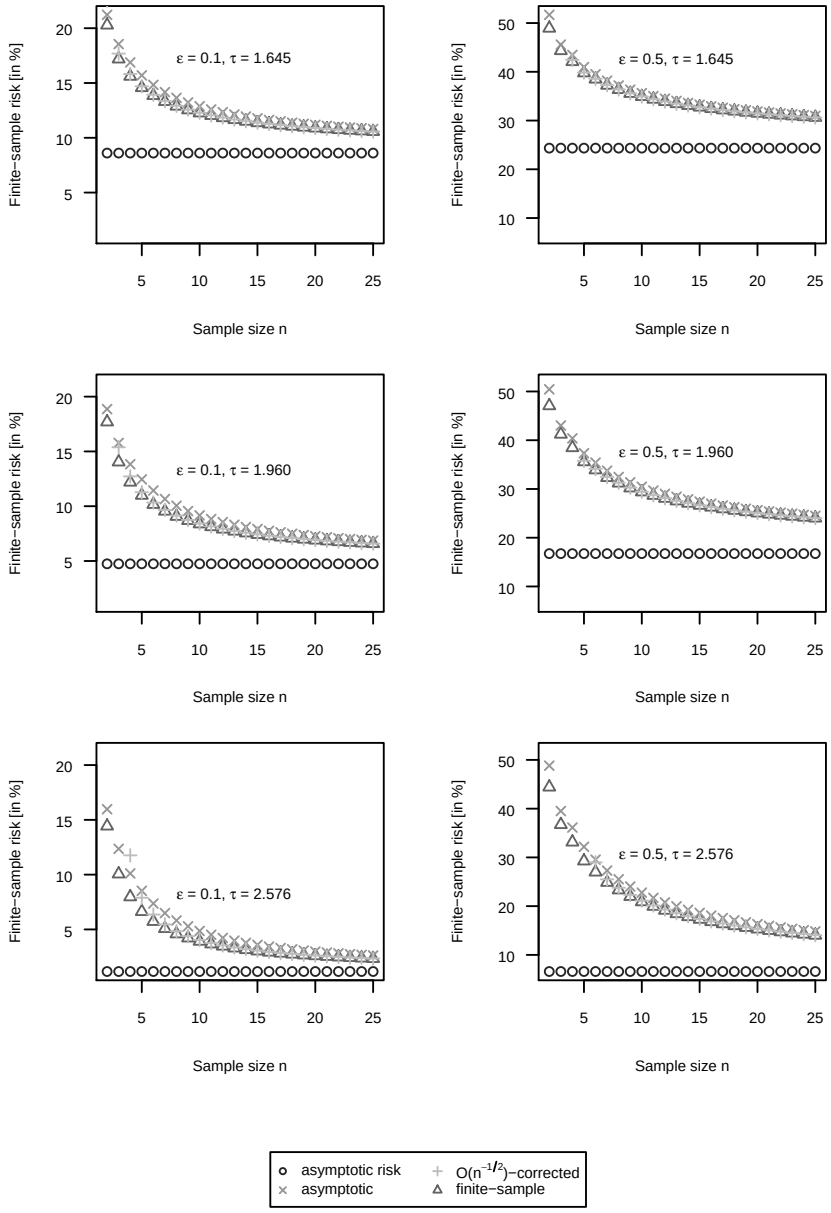

Figure 12.6: Finite-sample risk for $K = \text{Unif}([-1,2])$, $n \leq 25$, $\varepsilon = 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

ε	τ	n	$b_{c,0}^{\mathrm{fi}}$	$b_{c,0}^{\mathrm{as.c}}$	$b_{c,0}^{\mathrm{as}}$	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.1	1.645	2	0.684	0.147	1.837	20.284	22.521	21.222
		3	0.817	0.457		17.189	17.684	18.550
		4	0.912	0.642		15.611	15.820	16.878
		5	0.983	0.768		14.577	14.697	15.698
		10	1.187	1.081		12.266	12.287	12.878
		50	1.519	1.499		9.866	9.867	9.976
		100	1.608	1.598		9.432	9.432	9.484
	1.960	2	0.631	−0.156	1.971	17.690	—	14.700
		3	0.768	0.234		14.034	15.383	15.778
		4	0.869	0.467		12.193	12.720	13.832
		5	0.947	0.626		10.993	11.283	12.450
		10	1.178	1.020		8.382	8.426	9.163
		50	1.576	1.545		5.883	5.884	6.010
		100	1.685	1.670		5.471	5.471	5.529
	2.576	2	0.537	−0.838	2.177	14.431	—	15.972
		3	0.670	−0.285		10.057	—	12.357
		4	0.773	0.045		7.970	11.760	10.112
		5	0.856	0.270		6.628	7.892	8.524
		10	1.120	0.828		3.917	4.039	4.860
		50	1.629	1.574		1.826	1.828	1.939
		100	1.777	1.750		1.558	1.558	1.605
0.5	1.645	2	0.018	−0.217	0.582	48.937	—	51.686
		3	0.085	−0.070		44.331	—	45.567
		4	0.132	0.017		42.080	42.451	43.466
		5	0.167	0.077		39.776	39.846	40.966
		10	0.268	0.225		34.924	34.940	35.625
		50	0.430	0.422		28.614	28.614	28.774
		100	0.473	0.469		27.263	27.263	27.343
	1.960	2	0.042	−0.334	0.718	47.070	—	50.440
		3	0.112	−0.141		41.227	—	43.021
		4	0.164	−0.026		38.465	—	40.381
		5	0.203	0.053		35.573	35.780	37.331
		10	0.320	0.248		29.387	29.435	30.462
		50	0.521	0.508		21.573	21.574	21.815
		100	0.576	0.570		19.987	19.988	20.107
	2.576	2	0.062	−0.677	0.932	44.421	—	48.818
		3	0.132	−0.382		36.727	—	39.503
		4	0.186	−0.206		33.137	—	36.123
		5	0.230	−0.086		29.261	—	32.180
		10	0.369	0.212		20.888	21.124	22.771
		50	0.639	0.610		11.203	11.207	11.582
		100	0.719	0.705		9.531	9.532	9.708

Table 12.10: Comparison of the optimal clipping bounds and the corresponding finite-sample risks for $K = \mathrm{Unif}([-1, 2])$ in case of unconditional contamination neighborhoods ($* = c$, $t = 0$). [The risks are given in percent.]

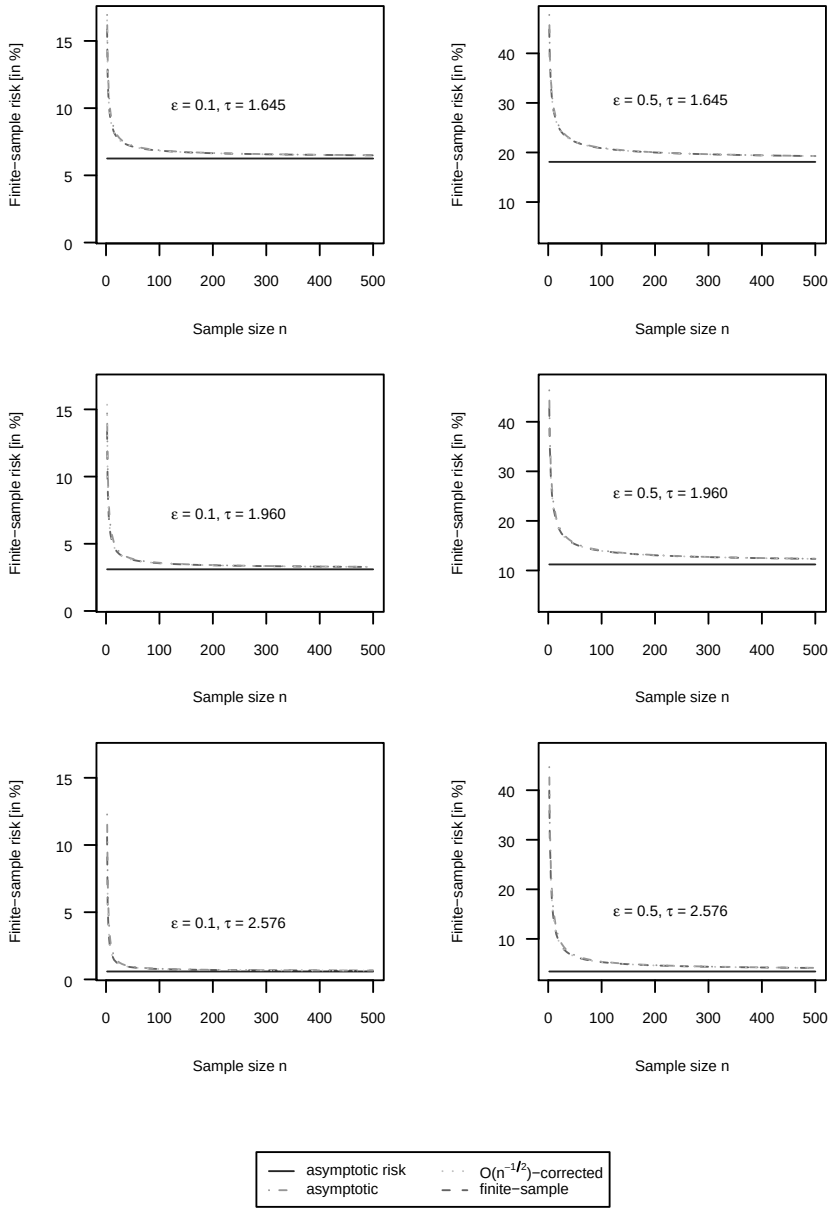

Figure 12.7: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, increasing sample size n, $\varepsilon = 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional contamination neighborhoods ($* = c$, $t = 0$).

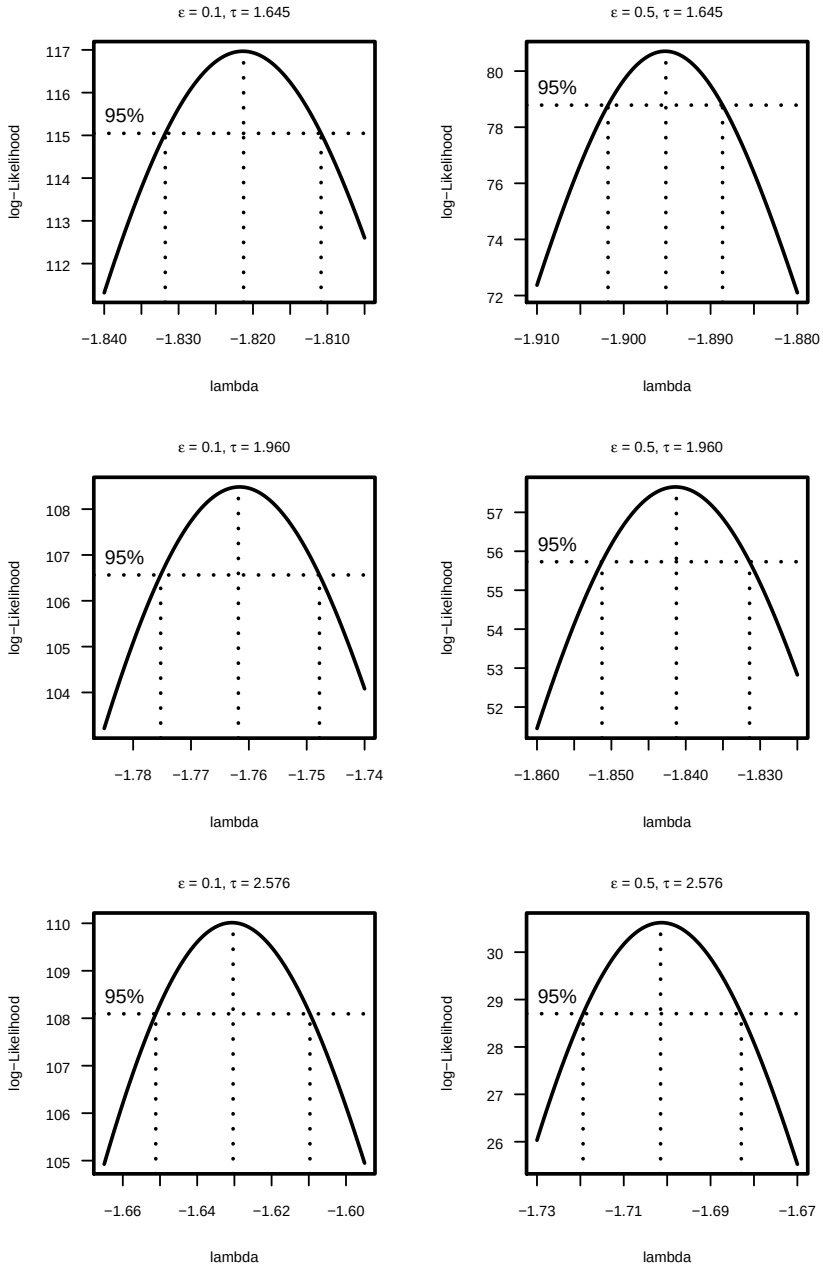

Figure 12.8: Results of the Box-Cox power transformation for the speed of convergence of the finite-sample risk in case of the finite-sample minimax estimator, $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional contamination neighborhoods $(* = c,\, t = 0)$.

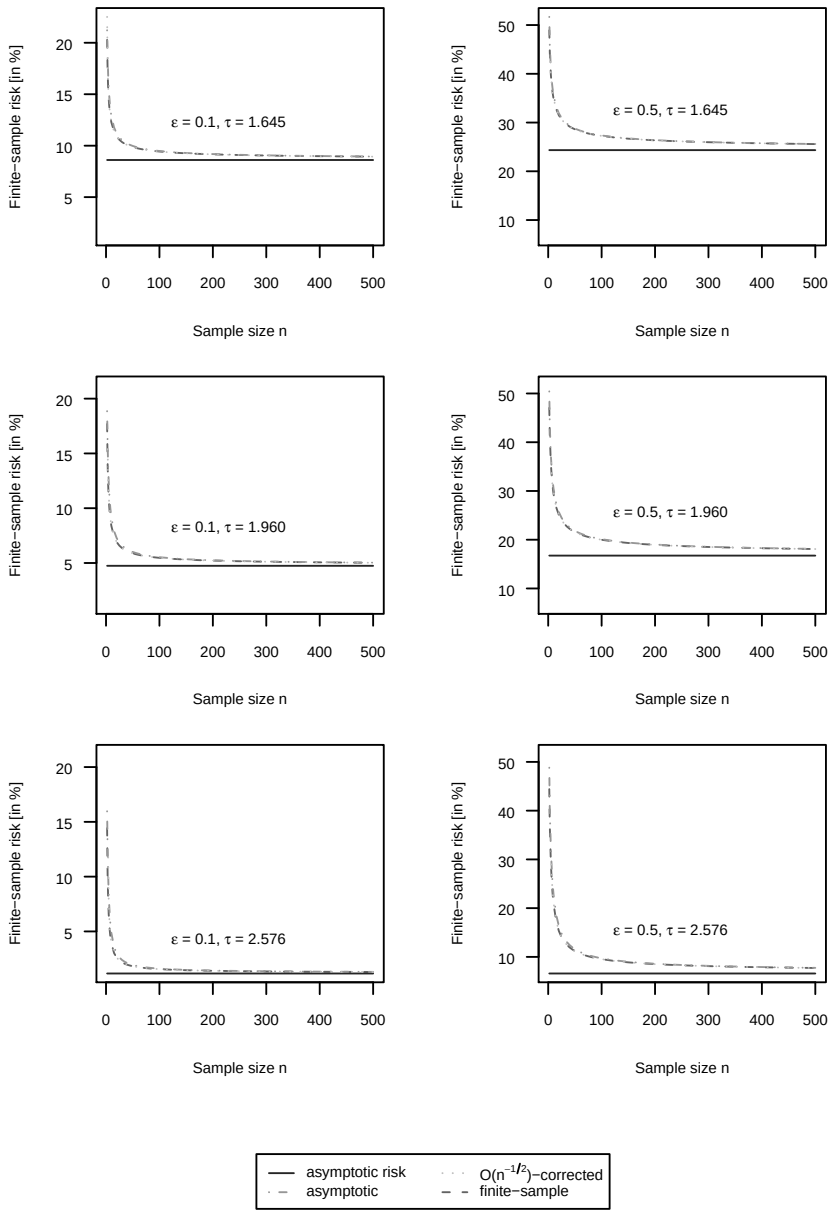

Figure 12.9: Finite-sample risk for $K = \text{Unif}([-1, 2])$, increasing sample size n, $\varepsilon = 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional contamination neighborhoods ($* = c$, $t = 0$).

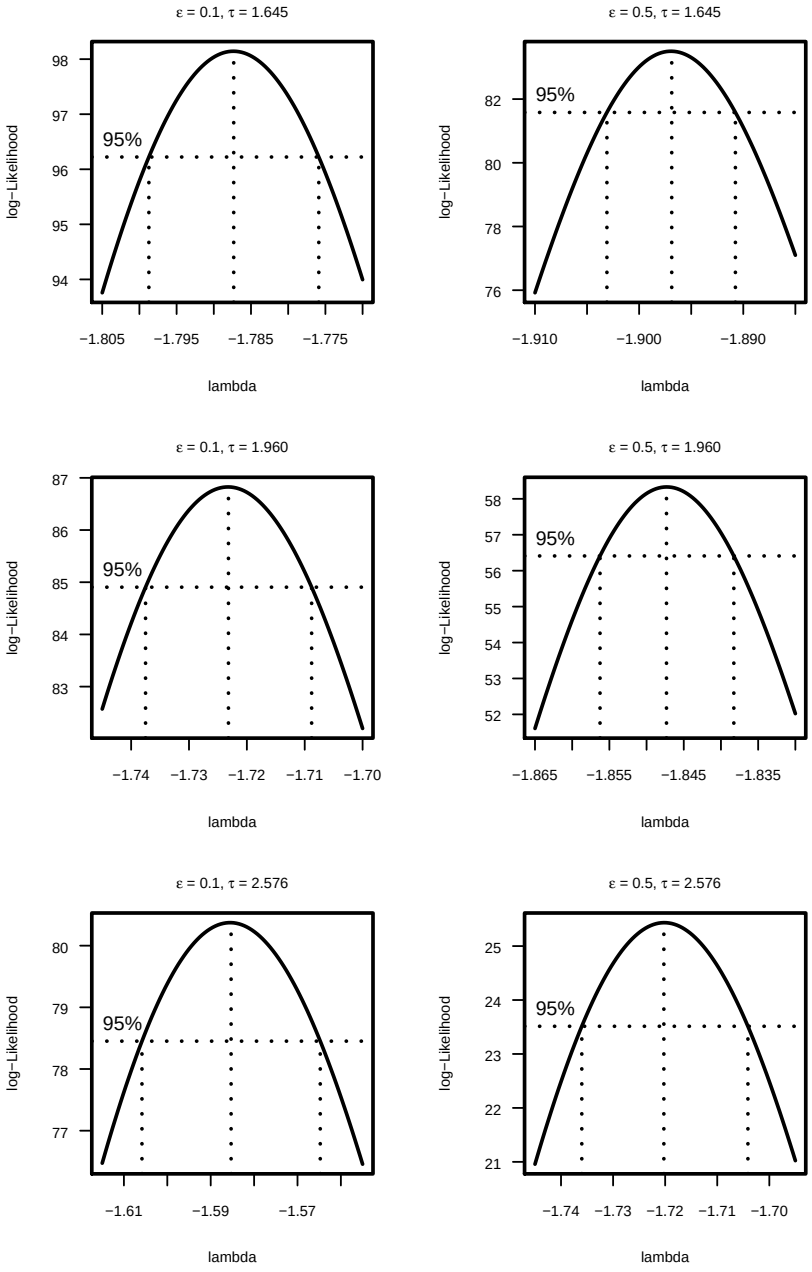

Figure 12.10: Results of the Box-Cox power transformation for the speed of convergence of the finite-sample risk in case of the finite-sample minimax estimator, $K = \mathrm{Unif}([-1, 2])$ and unconditional contamination neighborhoods ($* = c$, $t = 0$).

τ	n	ε_{as}	$\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{c,0})$	$\text{relRisk}^{\natural}_{\text{as}}$	$\varepsilon_{\text{as.c}}$	$\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{c,0})$	$\text{relRisk}^{\natural}_{\text{as.c}}$
1.645	2	0.248	27.76	1.078	0.149	19.99	1.046
	3	0.092	11.97	1.109	0.078	11.06	1.013
	4	0.067	9.40	1.096	0.323	25.10	1.006
	5	0.068	8.79	1.085	0.168	14.22	1.004
	10	0.085	8.15	1.048	0.218	14.09	1.001
	20	0.109	8.17	1.025	0.264	14.26	1.000
	50	0.140	8.42	1.010	0.298	13.95	1.000
	100	0.160	8.62	1.005	0.311	13.66	1.000
1.960	2	0.205	21.90	1.107	0.093	12.78	1.150
	3	0.077	8.04	1.184	0.058	6.93	1.038
	4	0.057	5.84	1.172	0.318	20.64	1.019
	5	0.058	5.24	1.156	0.139	8.95	1.012
	10	0.077	4.61	1.092	0.210	9.29	1.002
	20	0.107	4.60	1.048	0.283	9.90	1.001
	50	0.146	4.79	1.019	0.335	9.76	1.000
	100	0.172	4.97	1.009	0.357	9.53	1.000
2.576	2	0.127	12.79	1.160	0.024	4.00	2.218
	3	0.052	3.57	1.389	0.029	2.55	1.219
	4	0.038	2.15	1.413	0.233	10.59	1.124
	5	0.038	1.72	1.401	0.090	3.17	1.067
	10	0.059	1.26	1.259	0.159	3.02	1.013
	20	0.096	1.20	1.137	0.284	3.96	1.003
	50	0.150	1.28	1.054	0.388	4.26	1.000
	100	0.188	1.36	1.027	0.433	4.18	1.000

Table 12.11: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound in case of $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional contamination neighborhoods $(* = c,\, t = 0)$. [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by ε_{as} and $\varepsilon_{\text{as.c}}$, respectively, and $\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{c,0})$ is the finite-sample risk of $\tilde{S}^{\text{fi}}_{c,0}$ evaluated for the radius ε_{as} and $\varepsilon_{\text{as.c}}$, respectively. $\text{relRisk}^{\natural}_{\text{as}}$ and $\text{relRisk}^{\natural}_{\text{as.c}}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b^{\text{as,c}}_{c,0}$ by 0.]

τ	n	ε_{as}	$\text{Risk}^{\natural}(\tilde{S}_{c,0}^{\text{fi}})$	$\text{relRisk}_{\text{as}}^{\natural}$	$\varepsilon_{\text{as.c}}$	$\text{Risk}^{\natural}(\tilde{S}_{c,0}^{\text{fi}})$	$\text{relRisk}_{\text{as.c}}^{\natural}$
1.645	2	0.276	34.37	1.057	0.205	28.99	1.120
	3	0.113	18.13	1.080	0.073	15.14	1.030
	4	0.083	14.42	1.082	0.068	13.33	1.014
	5	0.073	12.80	1.079	0.073	12.81	1.009
	10	0.063	10.21	1.052	0.084	11.38	1.002
	20	0.071	9.47	1.028	0.094	10.60	1.000
	50	0.085	9.22	1.011	0.104	10.05	1.000
	100	0.094	9.20	1.005	0.109	9.79	1.000
1.960	2	0.240	29.21	1.078	0.040	12.17	1.691
	3	0.106	14.45	1.124	0.061	11.21	1.107
	4	0.078	10.76	1.137	0.464	36.13	1.036
	5	0.069	9.13	1.138	0.073	9.40	1.027
	10	0.060	6.56	1.098	0.088	7.84	1.005
	20	0.071	5.84	1.054	0.103	7.00	1.001
	50	0.090	5.59	1.022	0.117	6.43	1.000
	100	0.103	5.56	1.011	0.125	6.19	1.000
2.576	2	0.199	22.69	1.116	0.001	5.16	2.792
	3	0.090	9.41	1.229	0.019	4.70	1.812
	4	0.065	6.09	1.282	0.127	9.49	1.592
	5	0.057	4.67	1.308	0.077	5.57	1.194
	10	0.051	2.53	1.261	0.091	3.64	1.031
	20	0.067	1.97	1.152	0.113	2.87	1.006
	50	0.096	1.77	1.062	0.138	2.38	1.001
	100	0.116	1.74	1.030	0.152	2.20	1.000

Table 12.12: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound in case of $K = \text{Unif}([-1,2])$ and unconditional contamination neighborhoods ($* = c$, $t = 0$=. [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by ε_{as} and $\varepsilon_{\text{as.c}}$, respectively, and $\text{Risk}^{\natural}(\tilde{S}_{c,0}^{\text{fi}})$ is the finite-sample risk of $\tilde{S}_c^{\text{fi}}$ evaluated for the radius ε_{as} and $\varepsilon_{\text{as.c}}$, respectively. $\text{relRisk}_{\text{as}}^{\natural}$ and $\text{relRisk}_{\text{as.c}}^{\natural}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b_{c,0}^{\text{as,c}}$ by 0.]

τ	n	$\varepsilon^{\text{fi}}_{\max}$	ε_0	ε_3	ε_2	$\rho=0$	$\rho=3$	$\rho=2$
1.645	2	0.584	0.085	0.195	0.292	1.091	1.046	1.025
	3	0.665	0.042	0.098	0.121	1.082	1.034	1.016
	4	0.721	0.064	0.240	0.360	1.136	1.057	1.031
	5	0.763	0.063	0.171	0.231	1.139	1.050	1.024
	10	0.886	0.096	0.295	0.443	1.185	1.069	1.034
	11	0.901	0.100	0.268	0.382	1.190	1.072	1.035
	20	0.989	0.131	0.330	0.494	1.217	1.085	1.041
	50	1.094	0.175	0.365	0.513	1.245	1.101	1.049
	100	1.153	0.201	0.384	0.551	1.259	1.110	1.054
	∞	1.312	0.279	0.437	0.646	1.291	1.131	1.064
1.960	2	0.615	0.065	0.205	0.307	1.095	1.043	1.022
	3	0.709	0.028	0.079	0.096	1.082	1.036	1.016
	4	0.777	0.047	0.259	0.389	1.151	1.058	1.032
	5	0.830	0.046	0.154	0.210	1.153	1.054	1.025
	10	0.984	0.078	0.250	0.492	1.216	1.075	1.037
	11	1.004	0.083	0.264	0.384	1.224	1.081	1.039
	20	1.119	0.118	0.306	0.434	1.262	1.098	1.047
	50	1.261	0.169	0.374	0.531	1.304	1.121	1.059
	100	1.341	0.203	0.410	0.578	1.325	1.134	1.065
	∞	1.564	0.305	0.509	0.702	1.376	1.167	1.082
2.576	2	0.652	0.036	0.182	0.292	1.094	1.034	1.016
	3	0.766	0.012	0.048	0.057	1.071	1.036	1.016
	4	0.853	0.025	0.284	0.426	1.167	1.056	1.030
	5	0.921	0.023	0.114	0.156	1.163	1.058	1.026
	10	1.135	0.047	0.195	0.271	1.262	1.088	1.041
	11	1.163	0.052	0.236	0.350	1.274	1.096	1.045
	20	1.333	0.086	0.292	0.421	1.339	1.125	1.059
	50	1.553	0.146	0.382	0.550	1.414	1.166	1.079
	100	1.682	0.189	0.432	0.619	1.454	1.189	1.091
	∞	2.055	0.337	0.577	0.811	1.556	1.251	1.122

Table 12.13: Least favorable radii for the finite-sample minimax estimator in case of $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional contamination neighborhoods $(* = c, t = 0)$. [The least favorable radii $\varepsilon_0, \varepsilon_3, \varepsilon_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\varepsilon^{\text{fi}}_{\max}$ denotes the upper bound on the radius given by the disjointness condition (12.1.4). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

τ	n	$\varepsilon_{\max}^{\mathrm{fi}}$	ε_0	ε_3	ε_2	$\rho = 0$	$\rho = 3$	$\rho = 2$
1.645	2	0.519	0.114	0.173	0.259	1.066	1.040	1.025
	3	0.586	0.067	0.155	0.213	1.076	1.033	1.016
	4	0.634	0.074	0.211	0.317	1.113	1.052	1.029
	5	0.669	0.073	0.219	0.322	1.129	1.051	1.026
	10	0.770	0.091	0.257	0.385	1.182	1.072	1.037
	11	0.782	0.094	0.261	0.391	1.189	1.074	1.038
	20	0.853	0.117	0.284	0.426	1.221	1.088	1.045
	50	0.935	0.151	0.312	0.467	1.255	1.106	1.054
	100	0.979	0.173	0.326	0.490	1.271	1.115	1.059
	∞	1.094	0.239	0.365	0.547	1.305	1.139	1.071
1.960	2	0.549	0.101	0.183	0.275	1.069	1.040	1.024
	3	0.629	0.054	0.142	0.195	1.076	1.032	1.016
	4	0.686	0.062	0.229	0.343	1.122	1.054	1.030
	5	0.730	0.060	0.216	0.327	1.140	1.052	1.026
	10	0.859	0.077	0.286	0.430	1.211	1.078	1.040
	11	0.876	0.081	0.292	0.438	1.219	1.081	1.041
	20	0.969	0.106	0.323	0.484	1.265	1.101	1.051
	50	1.081	0.147	0.360	0.540	1.314	1.126	1.064
	100	1.142	0.174	0.381	0.571	1.338	1.139	1.071
	∞	1.303	0.261	0.434	0.652	1.393	1.176	1.089
2.576	2	0.589	0.080	0.196	0.294	1.069	1.037	1.022
	3	0.686	0.035	0.114	0.156	1.071	1.030	1.014
	4	0.759	0.043	0.253	0.380	1.130	1.054	1.031
	5	0.817	0.040	0.191	0.304	1.149	1.052	1.025
	10	0.997	0.052	0.269	0.499	1.250	1.083	1.043
	11	1.021	0.055	0.295	0.502	1.263	1.089	1.045
	20	1.162	0.079	0.323	0.581	1.338	1.119	1.060
	50	1.339	0.127	0.388	0.669	1.425	1.161	1.080
	100	1.439	0.162	0.426	0.720	1.470	1.186	1.093
	∞	1.713	0.287	0.541	0.856	1.581	1.257	1.128

Table 12.14: Least favorable radii for the finite-sample minimax estimator in case of $K = \mathrm{Unif}([-1, 2])$ and unconditional contamination neighborhoods ($* = c$, $t = 0$). [The least favorable radii $\varepsilon_0, \varepsilon_3, \varepsilon_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\varepsilon_{\max}^{\mathrm{fi}}$ denotes the upper bound on the radius given by the disjointness condition (12.1.4). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

12.3.1.3 Finite-Sample Distribution

We use Algorithm A (cf. Subsubsection 12.1.3.2) to compute the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{c,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance. By symmetry it suffices to consider the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{c,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$. Moreover, we give only the two "extreme" situations $\varepsilon = 0.1$, $\tau = 1.645$ (see Figures 12.11 and 12.13) and $\varepsilon = 0.5$, $\tau = 2.576$ (see Figures 12.12 and 12.14).

Remark 12.3.1 If $H''_{\tau_n}(dy \,|\, x)$ is absolutely continuous a.e. $H''_{\tau_n}(dx)$, then by Remark 12.1.5 the distribution of $\tilde{S}^{\mathrm{fi}}_{c,0}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^{n} \tilde{\psi}_{c,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. ////

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^{n} \tilde{\psi}_{c,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ and of $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the `R` function `optim`; confer R Development Core Team (2005).
As we see, in all cases about 10 observations are enough to get already quite close to a normal distribution (i.e., $d_\kappa < 0.02$ for $n \geq 10$) where the jumps included in the cumulative distribution functions decay exponentially in n; confer also Remark 11.3.3.

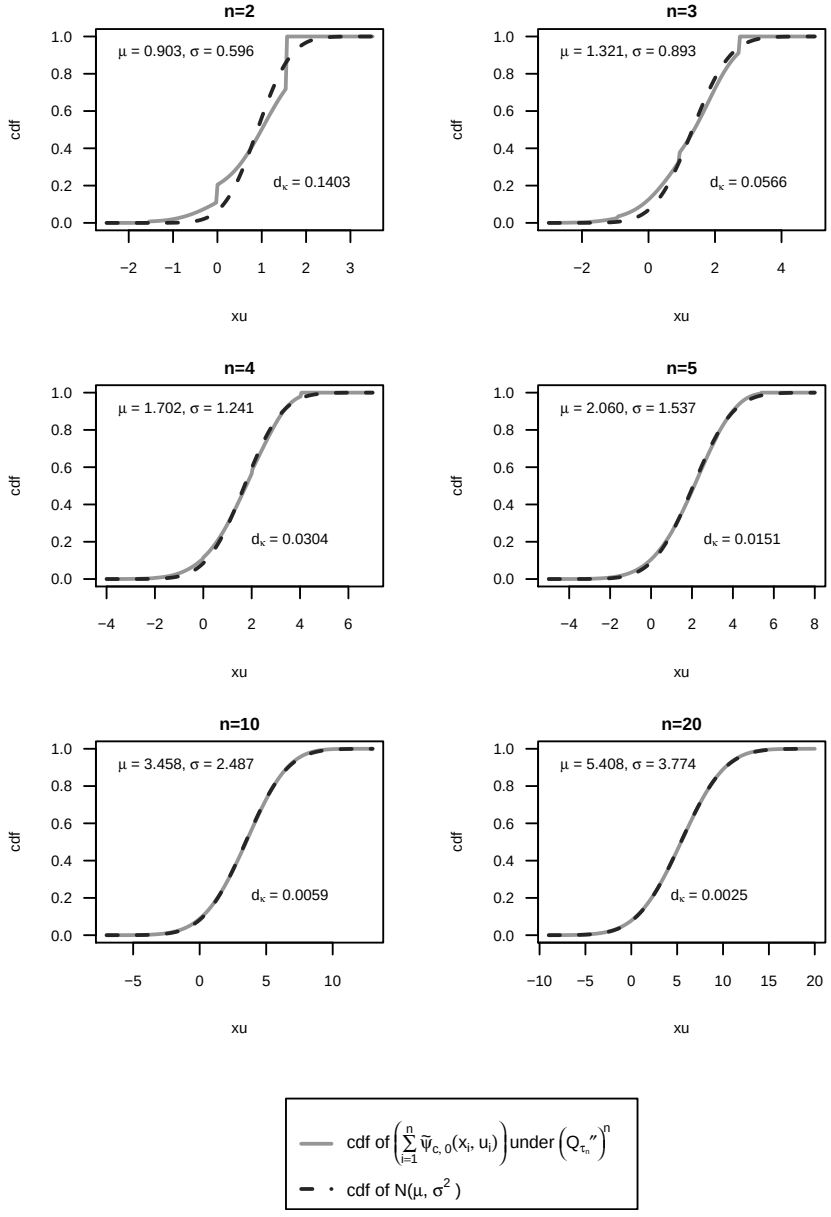

Figure 12.11: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\varepsilon = 0.1$ and $\tau = 1.645$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

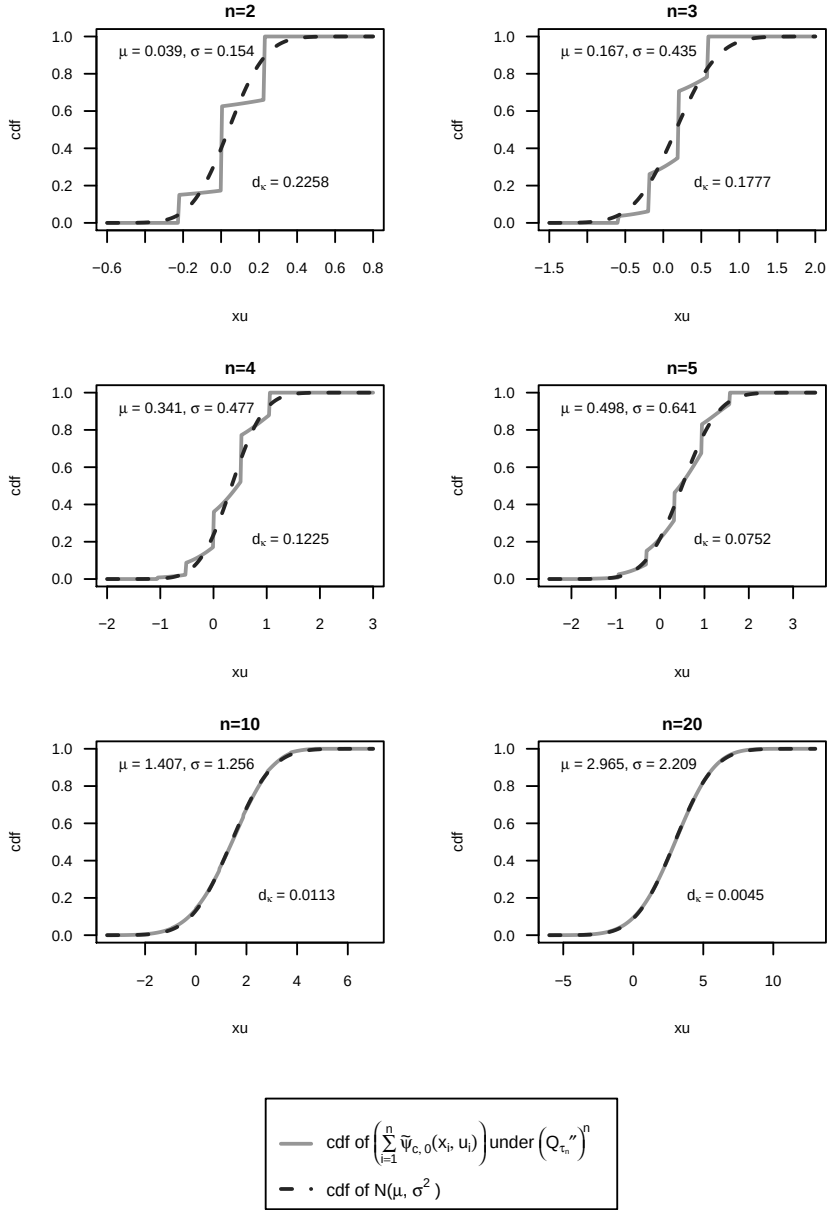

Figure 12.12: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\varepsilon = 0.5$ and $\tau = 2.576$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

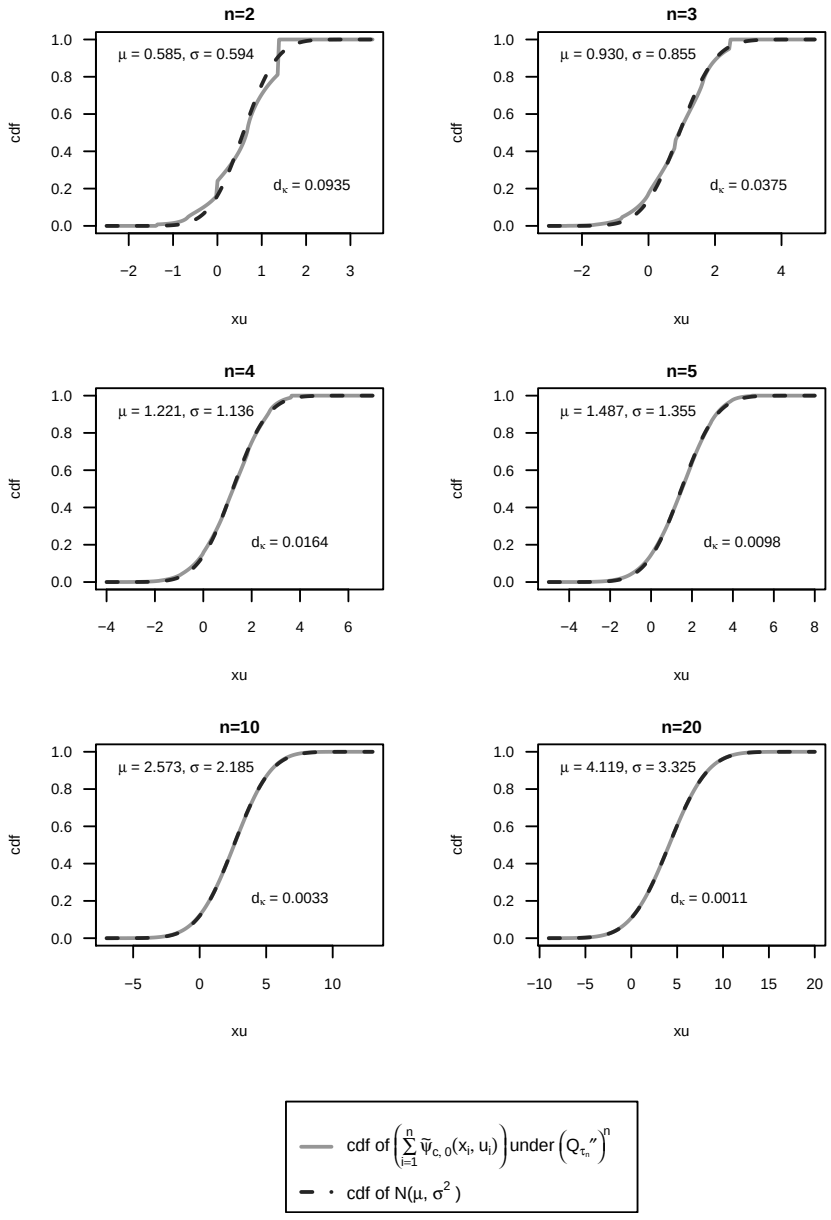

Figure 12.13: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\varepsilon = 0.1$ and $\tau = 1.645$ in case of unconditional contamination neighborhoods $(* = c,\, t = 0)$.

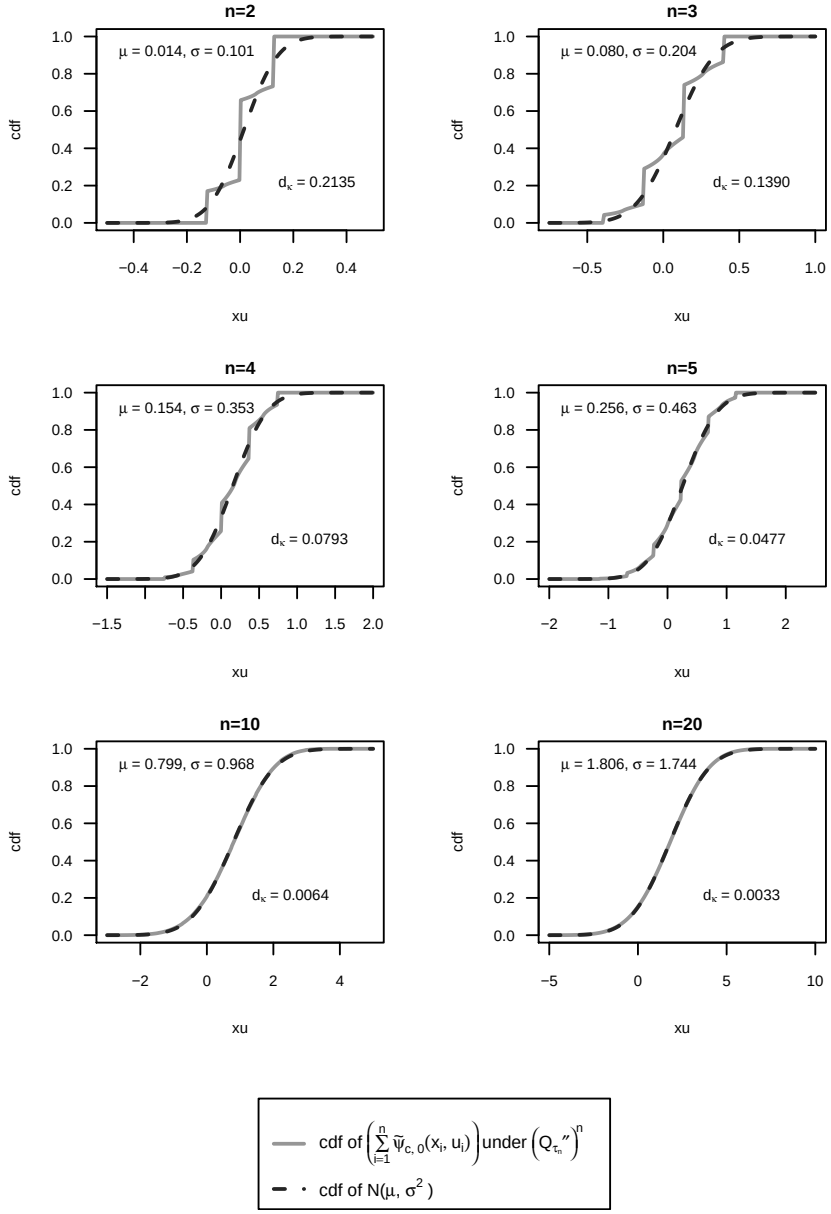

Figure 12.14: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\varepsilon = 0.5$ and $\tau = 2.576$ in case of unconditional contamination neighborhoods $(* = c,\ t = 0)$.

12.3.2 Unconditional Total Variation Neighborhoods

12.3.2.1 Optimal Clipping Bound

We fix radius $\delta = 0.05, 0.25$. That is, $\delta = \varepsilon/2$, which by (12.1.10) and (12.1.44) leads to $b_{c,0}^{\mathrm{as}} = b_{v,0}^{\mathrm{as}}$. Thus, we have well comparable results for unconditional contamination and total variation neighborhoods. We compute $b_{v,0}^{\mathrm{fi}}$, $b_{v,0}^{\mathrm{as}}$ and $b_{v,0}^{\mathrm{as.c}}$ by (12.1.37), (12.1.44) and (12.1.50). Figures 12.15 and 12.16 show the results for sample size $n \leq 25$. In all cases considered the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is too small compared with the finite-sample optimal clipping bound. For small n the $\mathrm{O}(n^{-1})$-correction can even lead to irregular (negative) clipping bounds; confer also Tables 12.15 and 12.16. But, already for moderate sample sizes n between 10 and 20 they are very close to each other where the approximation is worse for larger values of τ.

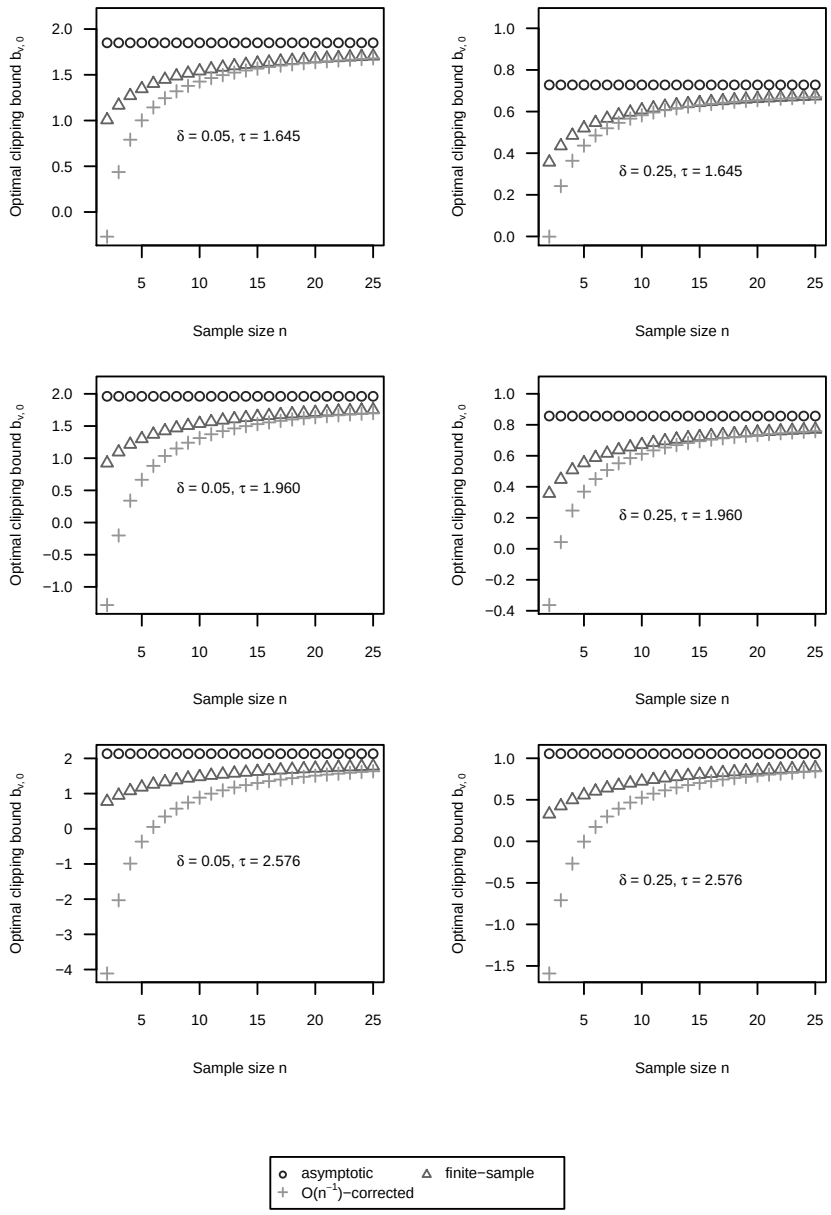

Figure 12.15: Optimal clipping bounds for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \le 25$, $\delta = 0.05, 0.25$ and $\tau = 1.645, 1.960, 2.576$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

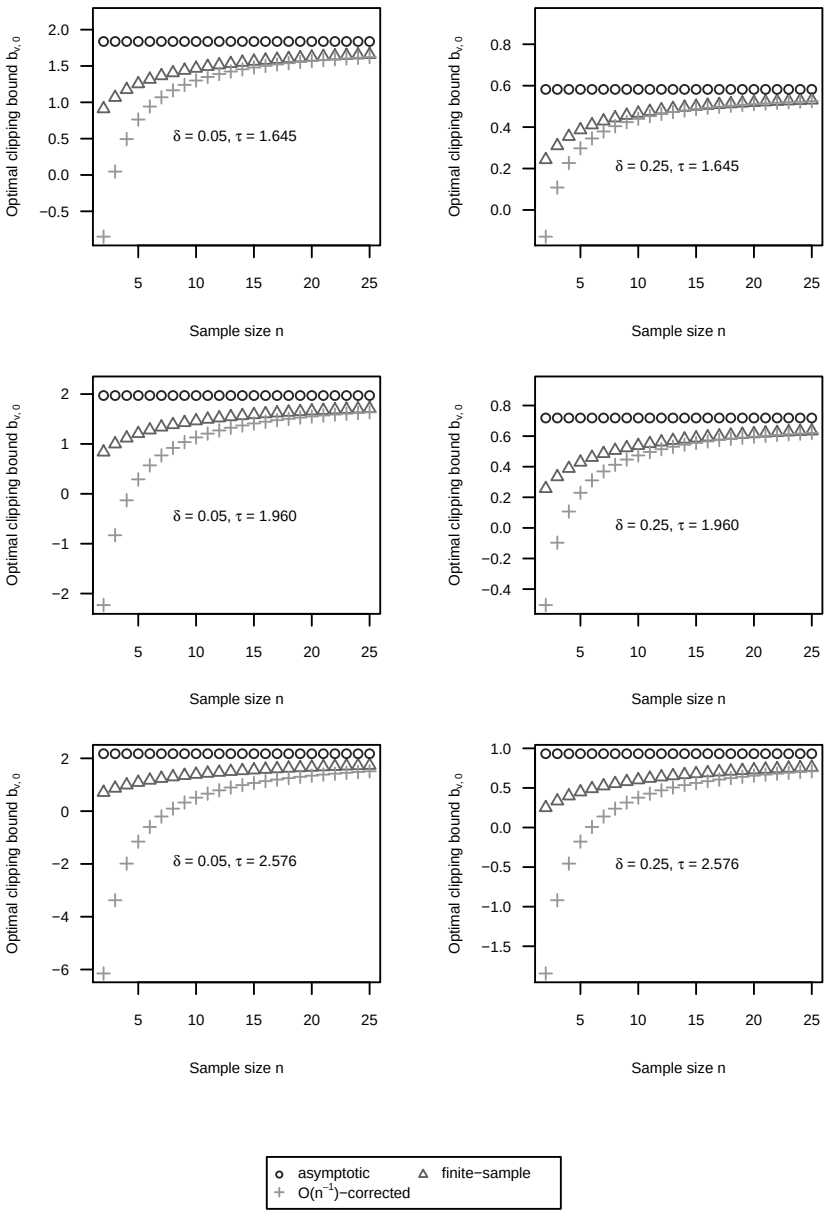

Figure 12.16: Optimal clipping bounds for $K = \text{Unif}([-1,2])$, $n \leq 25$, $\delta = 0.05, 0.25$ and $\tau = 1.645, 1.960, 2.576$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

12.3.2.2 Finite-Sample Risk

We fix radius $\delta = 0.05, 0.25$ and determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}^{\text{fi}}_{v,0}$, the asymptotic minimax estimator $\tilde{S}^{\text{as}}_{v,0}$ and the estimator $S^{\text{as.c}}_{v,0}$ which is based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound. Although there are clear differences between the clipping bounds of the finite-sample and the asymptotic minimax estimator, the differences (in absolute values) concerning the corresponding finite-sample risks are only small; see Figures 12.17 and 12.18 and Tables 12.15 and 12.16. Moreover, for moderate sample sizes n the finite-sample risk of the estimator which is based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is very close to the finite-sample risk of the finite-sample minimax estimator.

Figures 12.19 and 12.21 show the speed of convergence for the finite-sample risks. It seems to be of order n^{-1} as in case of the optimal clipping bounds; confer Lemma 12.1.2. To get a better impression, we consider $y = \text{Risk}^{\natural}_{\text{as}} - \text{Risk}^{\natural}_{\text{fi}}$ and apply the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002); i.e., we estimate lambda by means of maximum likelihood such that $y^{\text{lambda}} \approx n$. That is, lambda ≈ -1 indicates $y = \mathrm{O}(n^{-1})$. Indeed, the estimated values of lambda are very close to -1 which confirms our conjecture that we have a convergence of order n^{-1}; see Figures 12.20 and 12.22. Moreover, the results are almost independent from τ and δ.

So far, we compared absolute values for given radius δ and given width τ. But, now we again study relative risks; i.e., how much efficiency do we lose if we take the asymptotic optimal clipping bound, respectively the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound instead of the finite-sample optimal. More precisely, given some estimator S, we consider

$$\text{relRisk}(S; v, 0) = \frac{\text{Risk}(S; v, 0)}{\text{Risk}(\tilde{S}^{\text{fi}}_{c}; v, 0)} \tag{12.3.2}$$

and determine the radius δ_{as} and $\delta_{\text{as.c}}$, respectively, such that the relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is maximal, respectively. To obtain valid results in case of the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound, we replace negative values of $b^{\text{as.c}}_{v,0}$ by 0. As the numerical results show (see Tables 12.17 and 12.18), the maximum relative risks of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound may be large for small sample sizes n where $b^{\text{as.c}}_{v,0}$ is set to 0. However, if we obtain regular values for $b^{\text{as.c}}_{v,0}$ the maximum relative risks are small, respectively very small. Moreover, if we choose the width τ not too large and if we have a sample of moderate size (about 20), there is no big difference between the finite-sample risks of all three estimators.

Next, we determine the least favorable radii for the finite-sample minimax estimator; confer Tables 12.19 and 12.20. The corresponding radii are defined analogously to Section 2.2. Since the calculations of δ_3 and δ_2 are quite time consuming, we choose $q = 10$ in these two cases, for δ_0 we can use $q = 12$ as in the computations before. For $n < 10$ the least favorable radii and the corresponding relative risks behave different for even and odd sample size, respectively. This is probably

caused by the fact that there is mass at zero in case of an even sample size; confer Remark 11.3.3. However, this mass decays exponentially in n and therefore the effect disappears very fast for increasing sample size. Moreover, in all cases δ_0 is small compared to $\delta^{\mathrm{fi}}_{\max}$. The relative risks for $\rho = 0$ in all cases considered stay well below 60%. In most cases δ_3 and δ_2 are close to $\delta^{\mathrm{fi}}_{\max}/3$ and $\delta^{\mathrm{fi}}_{\max}/2$, respectively. But, the corresponding relative risks ($\rho = 3, 2$) are small and in most cases stay well below 26% and 13%, respectively. Thus, as in case of unconditional contamination neighborhoods we strongly recommend to use the finite-sample minimax estimator which is given by the corresponding least favorable radius if the true radius is completely unknown, respectively unknown except to belong to some radius interval.

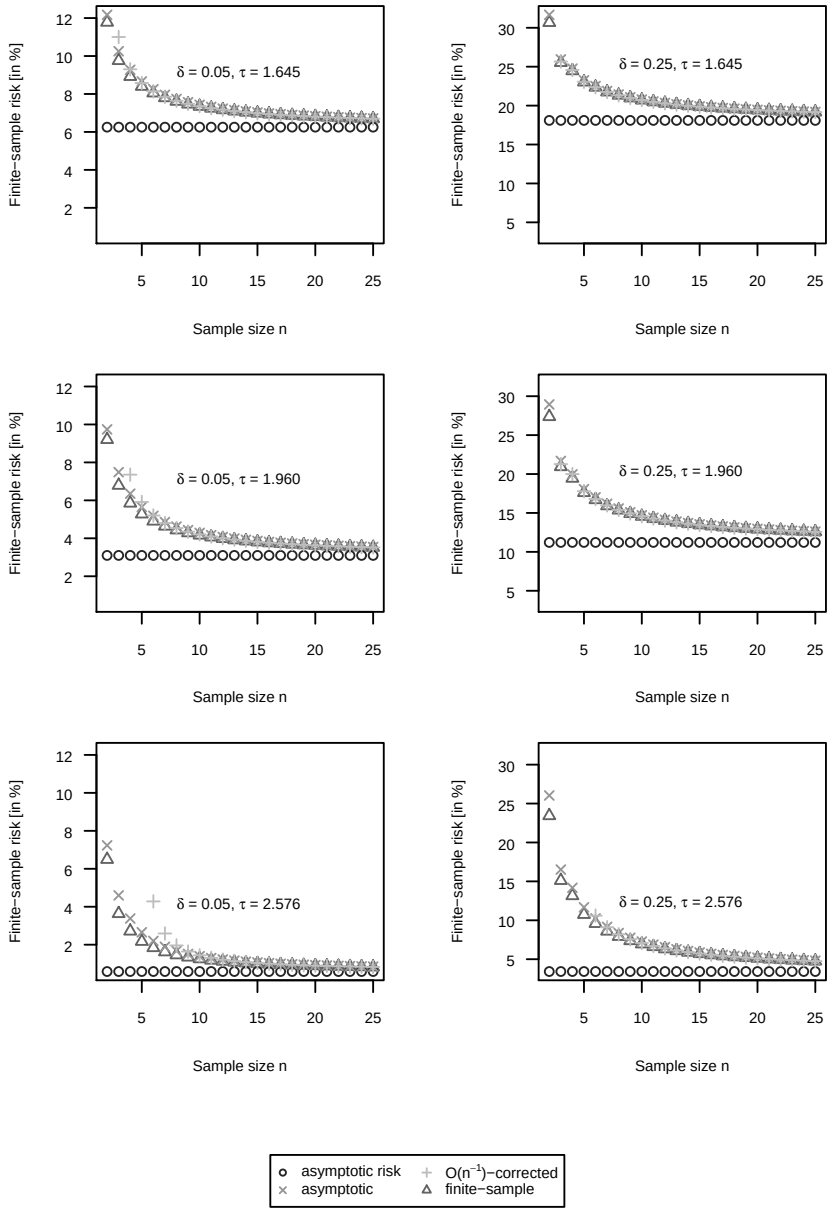

Figure 12.17: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \leq 25$, $\delta = 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

δ	τ	n	$b_{v,0}^{\mathrm{fi}}$	$b_{v,0}^{\mathrm{as.c}}$	$b_{v,0}^{\mathrm{as}}$	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.05	1.645	2	1.007	−0.269	1.850	11.778	—	12.170
		3	1.166	0.437		9.758	11.001	10.252
		4	1.272	0.790		8.920	9.301	9.275
		5	1.347	1.002		8.400	8.578	8.662
		10	1.541	1.426		7.346	7.359	7.431
		50	1.771	1.765		6.476	6.476	6.480
		100	1.809	1.807		6.365	6.365	6.366
	1.960	2	0.924	−1.282	1.962	9.203	—	9.739
		3	1.095	−0.201		6.791	—	7.486
		4	1.214	0.340		5.851	7.355	6.354
		5	1.302	0.664		5.279	5.914	5.649
		10	1.539	1.313		4.167	4.209	4.286
		50	1.847	1.832		3.309	3.309	3.315
		100	1.901	1.897		3.204	3.204	3.206
	2.576	2	0.775	−4.113	2.134	6.492	—	7.233
		3	0.949	−2.031		3.637	—	4.598
		4	1.080	−0.989		2.715	—	3.380
		5	1.181	−0.365		2.173	—	2.653
		10	1.478	0.884		1.254	1.439	1.393
		50	1.929	1.884		0.700	0.700	0.706
		100	2.022	2.009		0.643	0.643	0.644
0.25	1.645	2	0.357	−0.001	0.728	30.670	—	31.663
		3	0.435	0.242		25.534	25.662	25.937
		4	0.485	0.364		24.475	24.574	24.754
		5	0.521	0.436		23.026	23.066	23.256
		10	0.608	0.582		20.710	20.714	20.782
		50	0.700	0.699		18.642	18.642	18.646
		100	0.714	0.713		18.373	18.373	18.373
	1.960	2	0.357	−0.363	0.857	27.373	—	28.941
		3	0.447	0.043		20.941	21.275	21.683
		4	0.509	0.247		19.450	19.991	19.963
		5	0.554	0.369		17.611	17.795	18.053
		10	0.672	0.613		14.559	14.575	14.706
		50	0.811	0.808		11.900	11.900	11.908
		100	0.833	0.832		11.559	11.559	11.561
	2.576	2	0.327	−1.591	1.055	23.437	—	26.037
		3	0.426	−0.709		15.112	—	16.515
		4	0.500	−0.268		13.158	—	14.176
		5	0.557	−0.003		10.746	—	11.667
		10	0.723	0.526		6.956	7.087	7.269
		50	0.961	0.949		4.071	4.071	4.087
		100	1.005	1.002		3.741	3.741	3.745

Table 12.15: Comparison of the optimal clipping bounds and the corresponding finite-sample risks for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ in case of unconditional total variation neighborhoods ($* = v$, $t = 0$). [The risks are given in percent.]

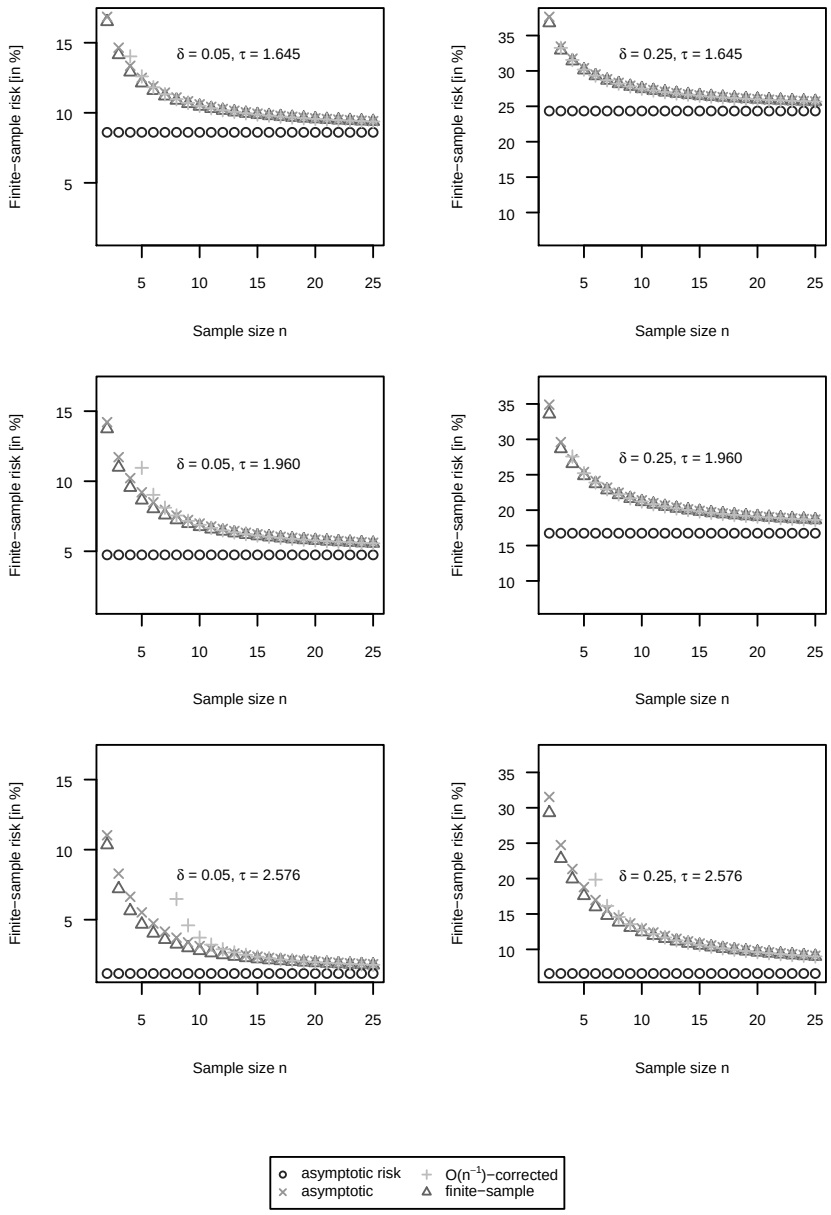

Figure 12.18: Finite-sample risk for $K = \text{Unif}([-1,2])$, $n \leq 25$, $\delta = 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

δ	τ	n	$b^{\mathrm{fi}}_{v,0}$	$b^{\mathrm{as.c}}_{v}$	$b^{\mathrm{as}}_{v,0}$	$\mathrm{Risk}^{\natural}_{\mathrm{fi}}$	$\mathrm{Risk}^{\natural}_{\mathrm{as.c}}$	$\mathrm{Risk}^{\natural}_{\mathrm{as}}$
0.05	1.645	2	0.910	−0.849	1.837	16.502	—	16.850
		3	1.067	0.047		14.144	18.075	14.649
		4	1.175	0.494		12.904	14.028	13.355
		5	1.254	0.763		12.125	12.598	12.495
		10	1.468	1.300		10.461	10.497	10.603
		50	1.740	1.730		8.997	8.997	9.005
		100	1.786	1.783		8.804	8.804	8.806
	1.960	2	0.836	−2.231	1.971	13.737	—	14.214
		3	0.999	−0.830		10.991	—	11.718
		4	1.117	−0.130		9.552	—	10.220
		5	1.207	0.290		8.650	10.955	9.209
		10	1.463	1.131		6.757	6.888	6.977
		50	1.825	1.803		5.156	5.157	5.169
		100	1.893	1.887		4.951	4.951	4.954
	2.576	2	0.706	−6.151	2.177	10.339	—	11.029
		3	0.866	−3.375		7.206	—	8.289
		4	0.989	−1.987		5.631	—	6.639
		5	1.087	−1.154		4.671	—	5.522
		10	1.393	0.511		2.791	3.721	3.117
		50	1.915	1.844		1.451	1.453	1.468
		100	2.031	2.010		1.303	1.303	1.307
0.25	1.645	2	0.243	−0.129	0.582	36.784	—	37.639
		3	0.310	0.108		32.930	33.252	33.439
		4	0.355	0.227		31.399	31.550	31.728
		5	0.386	0.298		30.123	30.190	30.384
		10	0.467	0.440		27.487	27.492	27.574
		50	0.555	0.554		25.018	25.018	25.023
		100	0.568	0.568		24.683	24.683	24.684
	1.960	2	0.257	−0.504	0.718	33.565	—	34.896
		3	0.334	−0.097		28.667	—	29.610
		4	0.388	0.107		26.563	27.573	27.203
		5	0.429	0.229		24.859	25.225	25.382
		10	0.538	0.474		21.195	21.222	21.385
		50	0.673	0.670		17.705	17.706	17.716
		100	0.695	0.694		17.229	17.229	17.232
	2.576	2	0.250	−1.844	0.932	29.299	—	31.527
		3	0.334	−0.919		22.826	—	24.731
		4	0.398	−0.456		19.936	—	21.353
		5	0.449	−0.178		17.587	—	18.795
		10	0.601	0.377		12.475	12.783	12.963
		50	0.835	0.821		7.798	7.798	7.827
		100	0.881	0.877		7.196	7.196	7.203

Table 12.16: Comparison of the optimal clipping bounds and the corresponding finite-sample risks for $K = \mathrm{Unif}([-1, 2])$ in case of unconditional total variation neighborhoods ($* = v$, $t = 0$). [The risks are given in percent.]

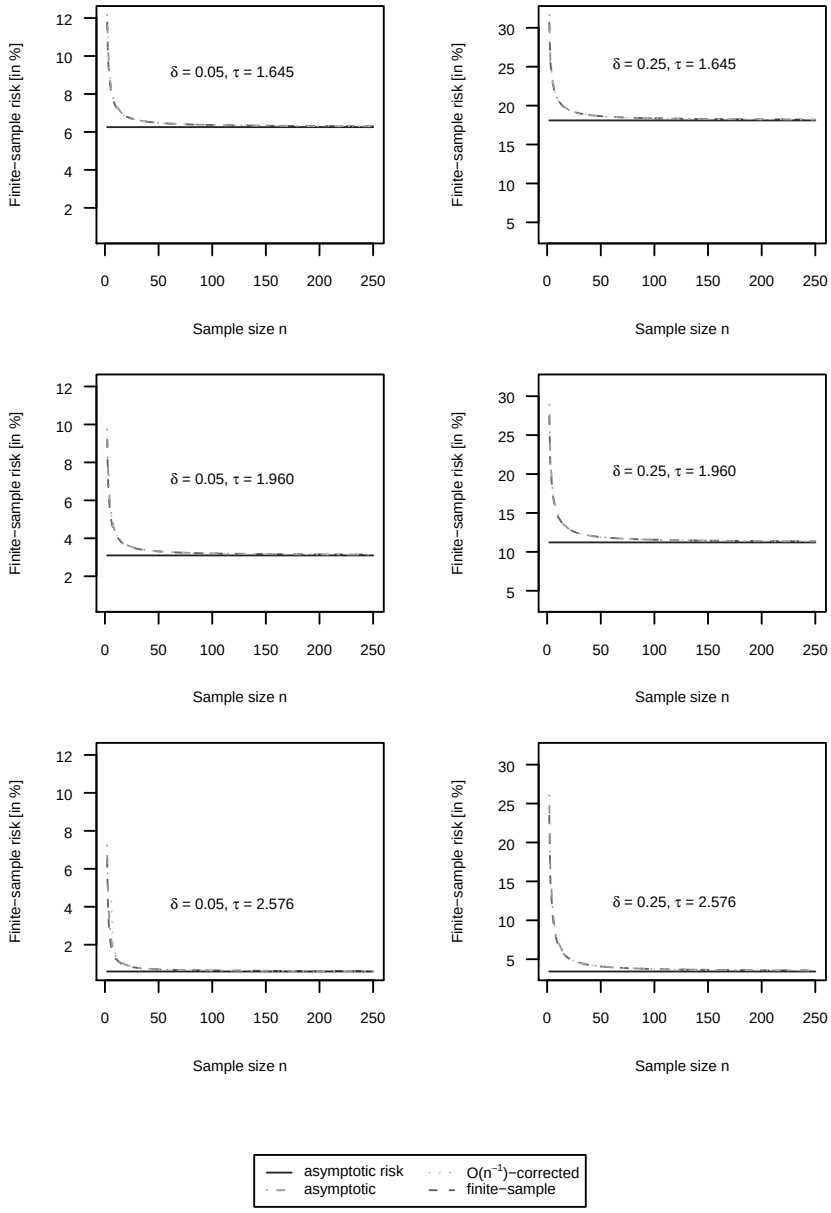

Figure 12.19: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, increasing sample size n, $\delta = 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

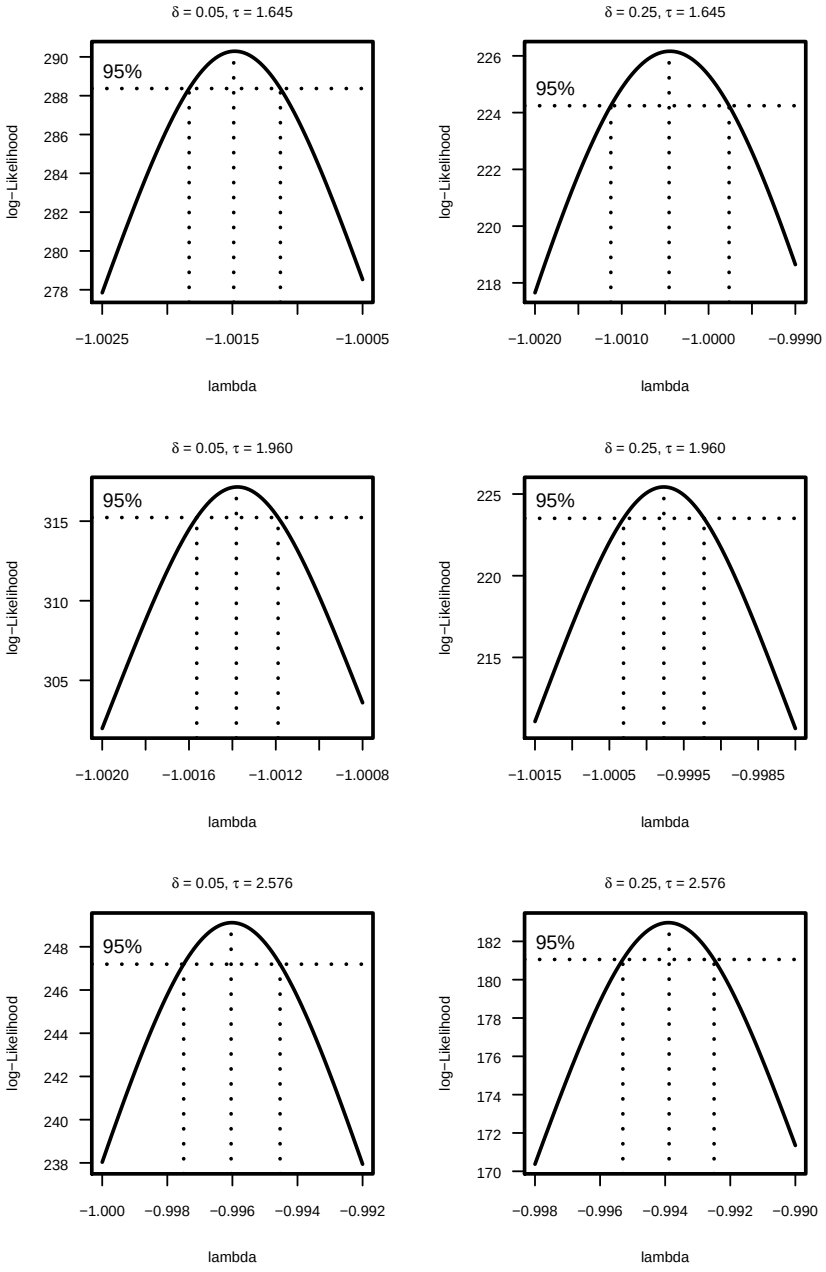

Figure 12.20: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional total variation neighborhoods $(* = v,\, t = 0)$.

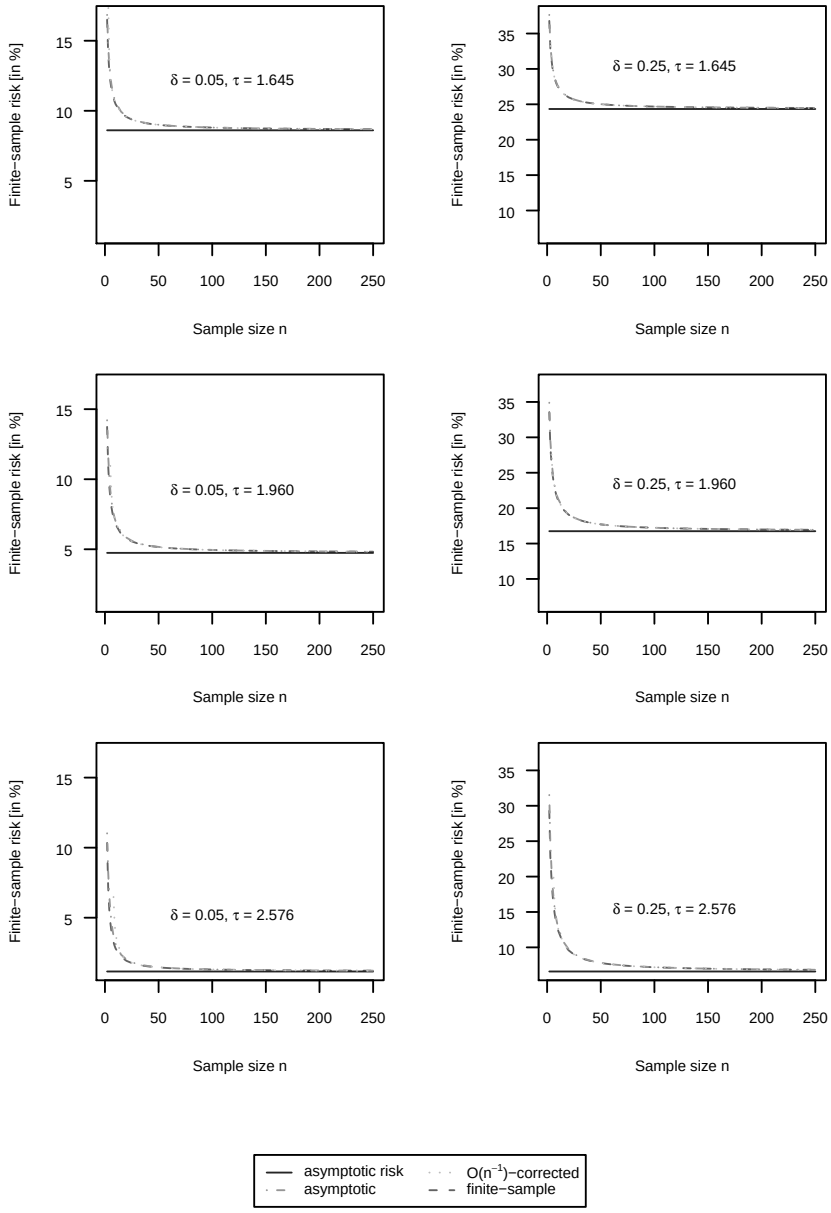

Figure 12.21: Finite-sample risk for $K = \text{Unif}([-1, 2])$, increasing sample size n, $\delta = 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of unconditional total variation neighborhoods ($* = v$, $t = 0$).

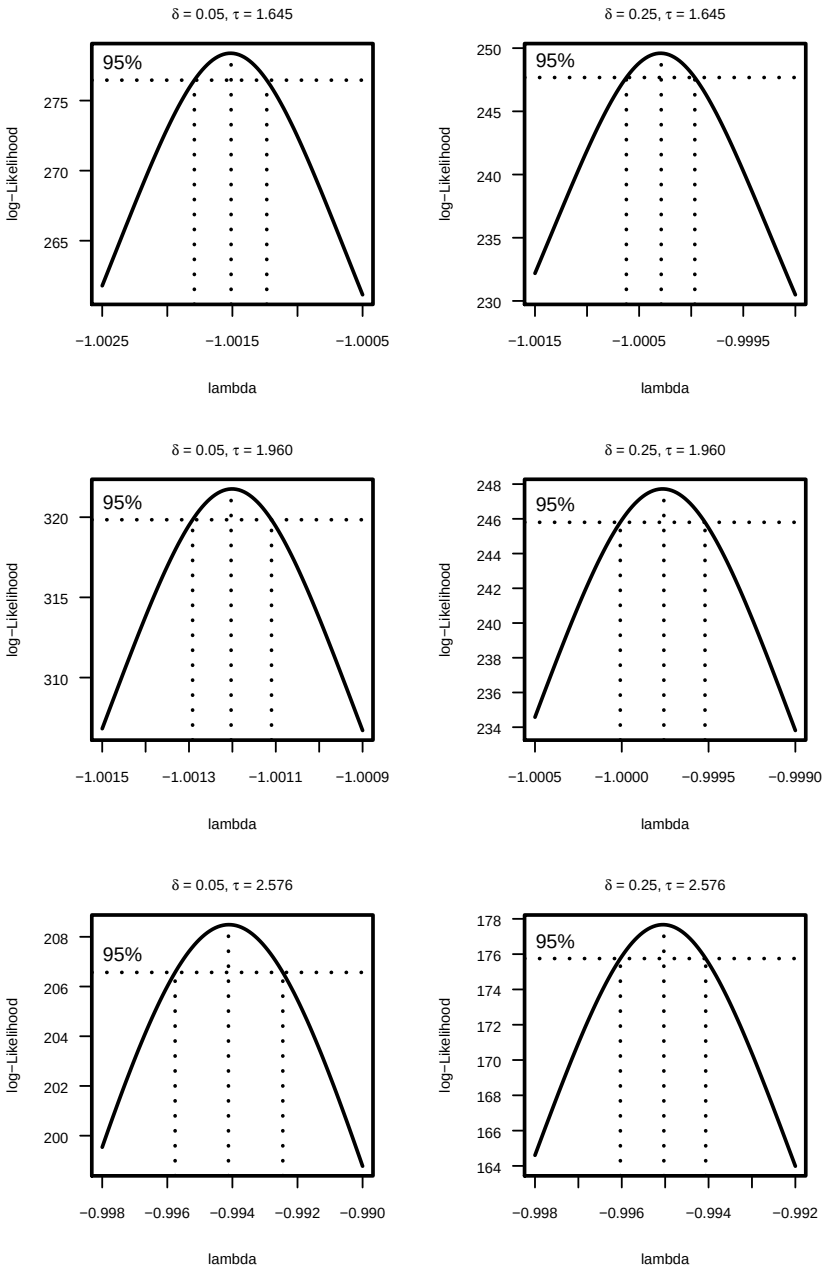

Figure 12.22: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \mathrm{Unif}([-1, 2])$ and unconditional total variation neighborhoods ($* = v$, $t = 0$).

τ	n	δ_{as}	$\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{v,0})$	$\text{relRisk}^{\natural}_{\text{as}}$	$\delta_{\text{as.c}}$	$\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{v,0})$	$\text{relRisk}^{\natural}_{\text{as.c}}$
1.645	2	0.126	19.441	1.039	0.000	6.196	2.385
	3	0.054	10.090	1.051	0.000	5.307	1.701
	4	0.038	8.063	1.041	0.003	5.200	1.128
	5	0.034	7.315	1.032	0.005	5.112	1.050
	10	0.031	6.196	1.012	0.007	4.720	1.004
	20	0.030	5.737	1.004	0.007	4.457	1.000
	50	0.030	5.455	1.001	0.007	4.288	1.000
	100	0.029	5.353	1.000	0.008	4.270	1.000
1.960	2	0.128	16.689	1.065	0.000	3.915	2.945
	3	0.051	6.860	1.102	0.000	3.070	2.058
	4	0.036	5.037	1.088	0.000	2.675	3.269
	5	0.033	4.379	1.073	0.001	2.565	1.390
	10	0.032	3.433	1.030	0.006	2.373	1.021
	20	0.034	3.062	1.010	0.008	2.183	1.002
	50	0.034	2.832	1.002	0.008	2.046	1.000
	100	0.034	2.748	1.000	0.008	2.010	1.000
2.576	2	0.125	13.119	1.121	0.000	1.674	4.353
	3	0.040	3.153	1.267	0.000	1.066	2.796
	4	0.028	1.894	1.265	0.000	0.799	5.668
	5	0.026	1.470	1.240	0.000	0.658	4.022
	10	0.030	0.924	1.116	0.003	0.489	1.357
	20	0.037	0.740	1.041	0.007	0.429	1.021
	50	0.042	0.632	1.008	0.009	0.374	1.001
	100	0.044	0.592	1.002	0.010	0.354	1.000

Table 12.17: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound in case of $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional total variation neighborhoods $(* = v,\, t = 0)$. [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by δ_{as} and $\delta_{\text{as.c}}$, respectively, and $\text{Risk}^{\natural}(\tilde{S}^{\text{fi}}_{v,0})$ is the finite-sample risk of $\tilde{S}^{\text{fi}}_{v,0}$ evaluated for the radius δ_{as} and $\delta_{\text{as.c}}$, respectively. $\text{relRisk}^{\natural}_{\text{as}}$ and $\text{relRisk}^{\natural}_{\text{as.c}}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b^{\text{as.c}}_{v}$ by 0.]

τ	n	δ_{as}	$\mathrm{Risk}^{\natural}(\tilde{S}^{\mathrm{fi}}_{v,0})$	$\mathrm{relRisk}^{\natural}_{\mathrm{as}}$	$\delta_{\mathrm{as.c}}$	$\mathrm{Risk}^{\natural}(\tilde{S}^{\mathrm{fi}}_{v,0})$	$\mathrm{relRisk}^{\natural}_{\mathrm{as.c}}$
1.645	2	0.124	24.512	1.026	0.000	10.643	1.977
	3	0.063	15.467	1.036	0.000	8.702	1.719
	4	0.046	12.511	1.035	0.000	7.717	2.295
	5	0.038	11.031	1.031	0.000	7.133	1.379
	10	0.026	8.461	1.015	0.004	6.499	1.010
	20	0.022	7.444	1.005	0.004	5.953	1.001
	50	0.021	6.879	1.001	0.004	5.580	1.000
	100	0.020	6.688	1.000	0.004	5.472	1.000
1.960	2	0.133	22.314	1.044	0.000	8.093	2.256
	3	0.066	12.428	1.068	0.000	6.104	1.944
	4	0.048	9.354	1.070	0.000	5.108	2.815
	5	0.040	7.856	1.066	0.000	4.525	2.465
	10	0.027	5.348	1.035	0.002	3.644	1.060
	20	0.025	4.407	1.013	0.004	3.253	1.004
	50	0.024	3.897	1.003	0.004	2.941	1.000
	100	0.023	3.723	1.001	0.004	2.835	1.000
2.576	2	0.137	18.909	1.081	0.000	5.051	2.842
	3	0.065	8.316	1.153	0.000	3.237	2.427
	4	0.046	5.394	1.179	0.000	2.378	4.135
	5	0.039	4.067	1.185	0.000	1.897	3.557
	10	0.028	2.044	1.125	0.000	1.070	6.162
	20	0.028	1.398	1.052	0.003	0.837	1.062
	50	0.029	1.085	1.012	0.005	0.689	1.002
	100	0.030	0.982	1.003	0.005	0.631	1.000

Table 12.18: Maximum relative risk of the asymptotic minimax estimator and of the estimator based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound in case of $K = \mathrm{Unif}([-1,2])$ and unconditional total variation neighborhoods $(* = v,\, t = 0)$. [The finite-sample risk is given in percent. The radius which leads to the maximum relative risk is denoted by δ_{as} and $\delta_{\mathrm{as.c}}$, respectively, and $\mathrm{Risk}^{\natural}(\tilde{S}^{\mathrm{fi}}_{v,0})$ is the finite-sample risk of $\tilde{S}^{\mathrm{fi}}_{v,0}$ evaluated for the radius δ_{as} and $\delta_{\mathrm{as.c}}$, respectively. $\mathrm{relRisk}^{\natural}_{\mathrm{as}}$ and $\mathrm{relRisk}^{\natural}_{\mathrm{as.c}}$ are the corresponding maximum relative risks. To obtain valid results we replace negative values of $b^{\mathrm{as.c}}_{v}$ by 0.]

τ	n	$\delta^{\mathrm{fi}}_{\max}$	δ_0	δ_3	δ_2	$\rho=0$	$\rho=3$	$\rho=2$
1.645	2	0.498	0.099	0.166	0.249	1.126	1.067	1.036
	3	0.539	0.064	0.137	0.152	1.137	1.047	1.021
	4	0.564	0.077	0.188	0.282	1.176	1.074	1.041
	5	0.580	0.082	0.189	0.241	1.194	1.072	1.034
	10	0.615	0.103	0.205	0.307	1.239	1.096	1.047
	11	0.619	0.105	0.206	0.300	1.243	1.098	1.048
	20	0.635	0.119	0.212	0.317	1.264	1.111	1.054
	50	0.647	0.131	0.216	0.324	1.280	1.123	1.060
	100	0.652	0.135	0.217	0.326	1.286	1.127	1.062
	∞	0.656	0.140	0.219	0.323	1.291	1.131	1.064
1.960	2	0.544	0.082	0.181	0.272	1.136	1.064	1.033
	3	0.601	0.049	0.114	0.127	1.148	1.045	1.020
	4	0.636	0.062	0.212	0.318	1.202	1.078	1.043
	5	0.659	0.068	0.194	0.248	1.226	1.075	1.034
	10	0.714	0.095	0.238	0.357	1.292	1.108	1.052
	11	0.720	0.099	0.236	0.334	1.299	1.112	1.054
	20	0.746	0.119	0.249	0.373	1.332	1.132	1.064
	50	0.767	0.137	0.256	0.350	1.358	1.151	1.074
	100	0.774	0.145	0.254	0.351	1.367	1.159	1.077
	∞	0.782	0.152	0.255	0.351	1.376	1.167	1.082
2.576	2	0.604	0.051	0.201	0.302	1.144	1.052	1.025
	3	0.687	0.025	0.063	0.070	1.149	1.041	1.018
	4	0.743	0.036	0.248	0.371	1.232	1.078	1.042
	5	0.783	0.041	0.169	0.216	1.268	1.073	1.033
	10	0.885	0.070	0.295	0.442	1.383	1.121	1.057
	11	0.896	0.075	0.257	0.371	1.397	1.130	1.061
	20	0.950	0.104	0.269	0.379	1.462	1.168	1.080
	50	0.995	0.138	0.282	0.398	1.516	1.212	1.102
	100	1.011	0.152	0.285	0.402	1.536	1.230	1.111
	∞	1.028	0.169	0.289	0.405	1.556	1.251	1.122

Table 12.19: Least favorable radii for the finite-sample minimax estimator in case of $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and unconditional total variation neighborhoods $(* = v,\, t = 0)$. [The least favorable radii $\delta_0, \delta_3, \delta_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\delta^{\mathrm{fi}}_{\max}$ denotes the upper bound on the radius given by the disjointness condition (12.1.4). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

τ	n	$\delta_{\max}^{\mathrm{fi}}$	δ_0	δ_3	δ_2	$\rho = 0$	$\rho = 3$	$\rho = 2$
1.645	2	0.410	0.118	0.136	0.205	1.085	1.055	1.035
	3	0.443	0.086	0.148	0.222	1.115	1.055	1.028
	4	0.464	0.085	0.155	0.232	1.149	1.071	1.039
	5	0.478	0.086	0.159	0.239	1.173	1.077	1.040
	10	0.509	0.094	0.170	0.254	1.232	1.100	1.052
	11	0.512	0.096	0.171	0.256	1.238	1.103	1.053
	20	0.527	0.104	0.176	0.263	1.267	1.116	1.060
	50	0.539	0.113	0.179	0.269	1.289	1.129	1.066
	100	0.543	0.116	0.181	0.271	1.297	1.134	1.069
	∞	0.547	0.119	0.182	0.273	1.305	1.139	1.071
1.960	2	0.449	0.113	0.150	0.224	1.091	1.056	1.035
	3	0.494	0.078	0.165	0.247	1.124	1.055	1.027
	4	0.523	0.077	0.174	0.261	1.167	1.075	1.041
	5	0.542	0.078	0.181	0.271	1.197	1.082	1.042
	10	0.590	0.090	0.197	0.295	1.279	1.113	1.058
	11	0.595	0.092	0.198	0.297	1.288	1.117	1.060
	20	0.619	0.105	0.206	0.309	1.332	1.138	1.071
	50	0.638	0.119	0.213	0.319	1.368	1.159	1.081
	100	0.645	0.124	0.215	0.322	1.380	1.167	1.085
	∞	0.652	0.130	0.217	0.326	1.393	1.176	1.089
2.576	2	0.504	0.097	0.168	0.252	1.096	1.054	1.033
	3	0.568	0.060	0.189	0.283	1.128	1.049	1.023
	4	0.612	0.060	0.204	0.306	1.185	1.076	1.043
	5	0.644	0.060	0.215	0.322	1.226	1.083	1.042
	10	0.728	0.073	0.243	0.364	1.356	1.129	1.066
	11	0.738	0.076	0.246	0.369	1.372	1.135	1.069
	20	0.785	0.095	0.262	0.392	1.454	1.173	1.087
	50	0.826	0.119	0.275	0.413	1.526	1.215	1.108
	100	0.841	0.130	0.280	0.420	1.553	1.234	1.117
	∞	0.856	0.143	0.270	0.428	1.581	1.257	1.128

Table 12.20: Least favorable radii for the finite-sample minimax estimator in case of $K = \mathrm{Unif}([-1, 2])$ and unconditional total variation neighborhoods ($* = v$, $t = 0$). [The least favorable radii $\delta_0, \delta_3, \delta_2$ and the corresponding relative risks are defined analogously to Section 2.2. $\delta_{\max}^{\mathrm{fi}}$ denotes the upper bound on the radius given by the disjointness condition (12.1.4). The results for $n = \infty$ are based on the asymptotic risk of the asymptotic minimax estimator.]

12.3.2.3 Finite-Sample Distribution

We once more use Algorithm A (cf. Subsubsection 12.1.3.2) to compute the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{v,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance. By symmetry it suffices to consider only the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{v,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$. Moreover, we give only the two "extreme" situations $\delta = 0.05$, $\tau = 1.645$ (see Figures 12.23 and 12.25) and $\delta = 0.25$, $\tau = 2.576$ (see Figures 12.24 and 12.26).

Remark 12.3.2 If $H''_{\tau_n}(dy \,|\, x)$ is absolutely continuous a.e. $H''_{\tau_n}(dx)$, then by Remark 12.1.5 the distribution of $\tilde{S}^{\text{fi}}_{v,0}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^{n} \tilde{\psi}_{v,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. ////

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^{n} \tilde{\psi}_{v,0}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ and $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the `R` function `optim`; confer R Development Core Team (2005). As we see, in all cases about 10 observations are enough to get already quite close to a normal distribution (i.e., $d_\kappa \leq 0.01$ for $n \geq 10$) where the jumps included in the cumulative distribution functions decay exponentially in n; confer also Remark 11.3.3.

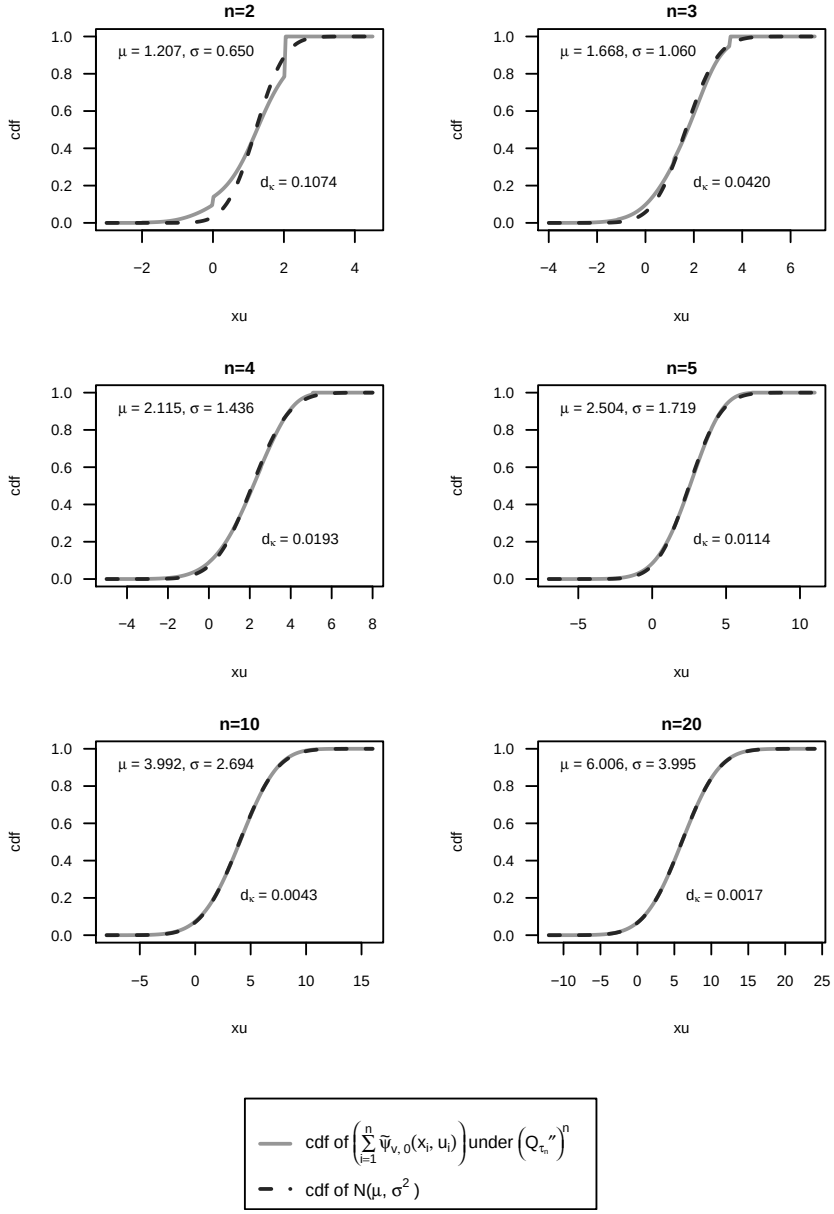

Figure 12.23: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\delta = 0.05$ and $\tau = 1.645$ in case of unconditional total variation neighborhoods ($* = v$, $t = 0$).

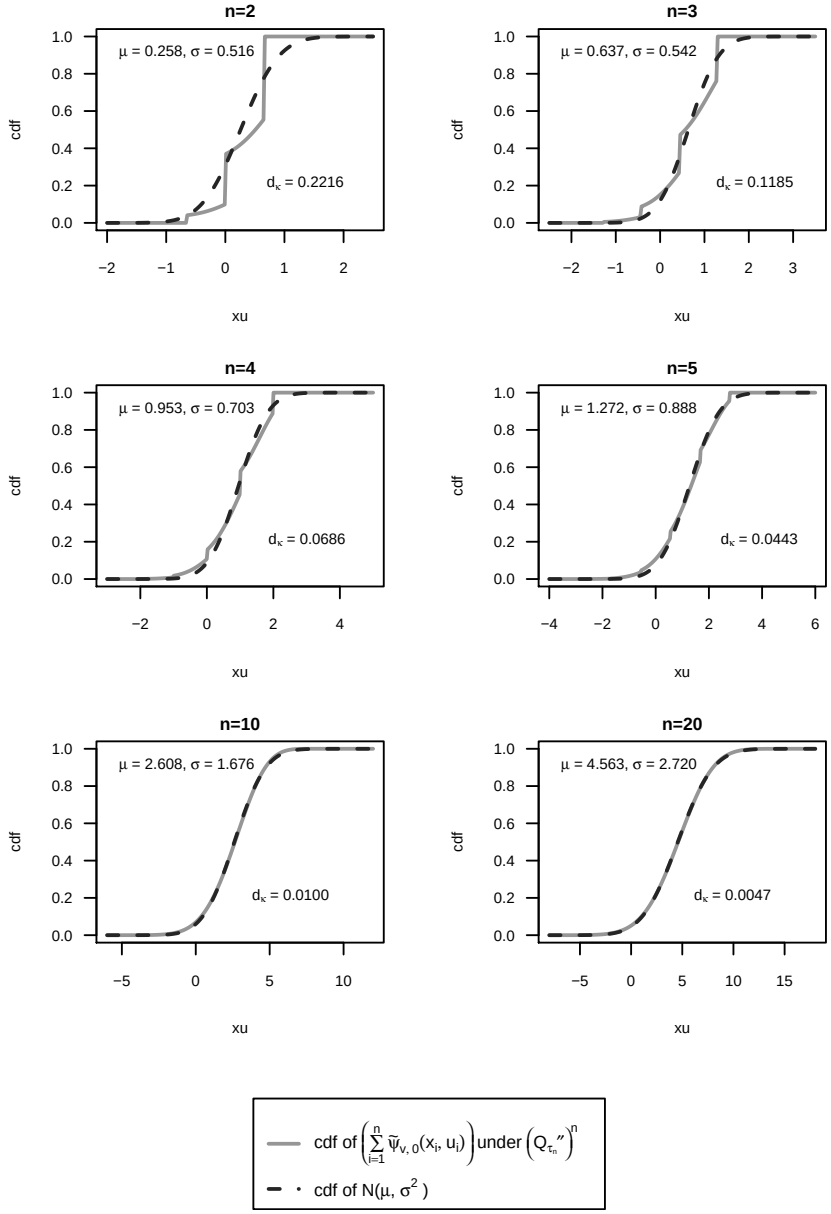

Figure 12.24: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\delta = 0.25$ and $\tau = 2.576$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

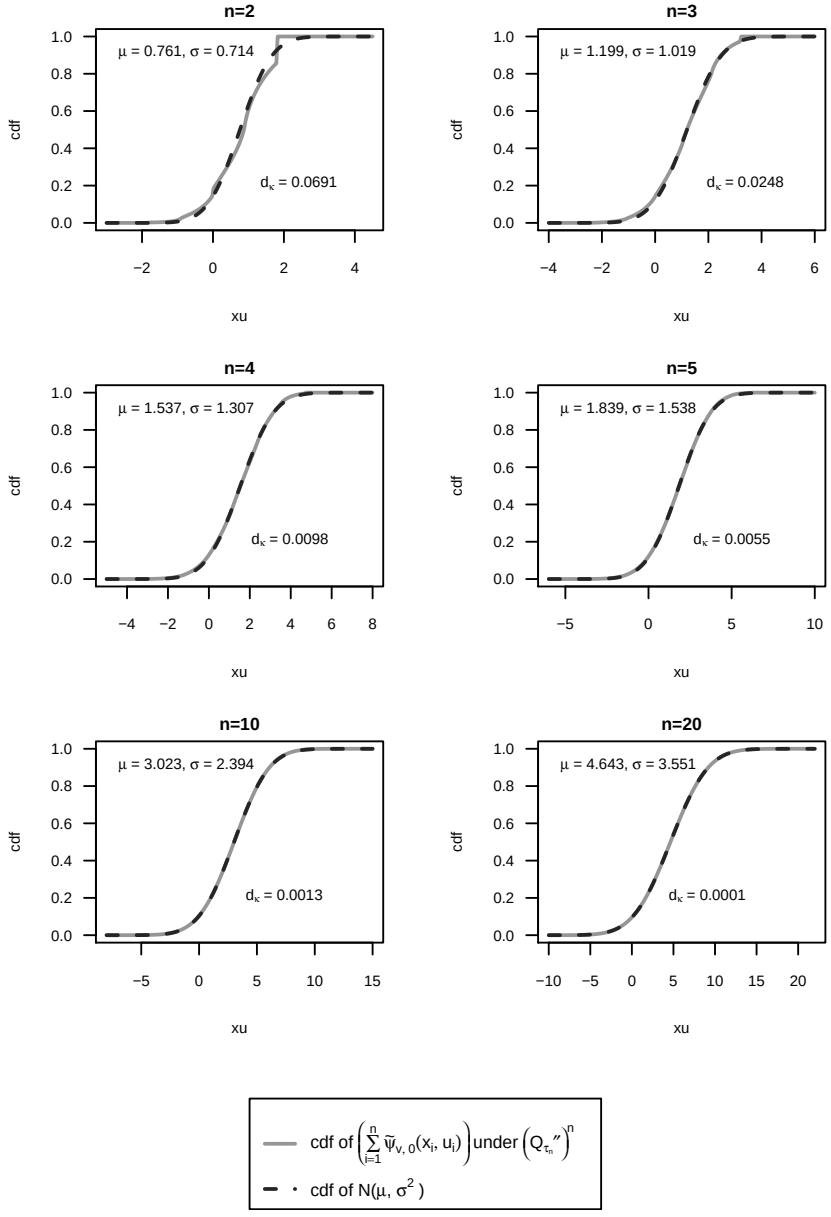

Figure 12.25: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\delta = 0.05$ and $\tau = 1.645$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

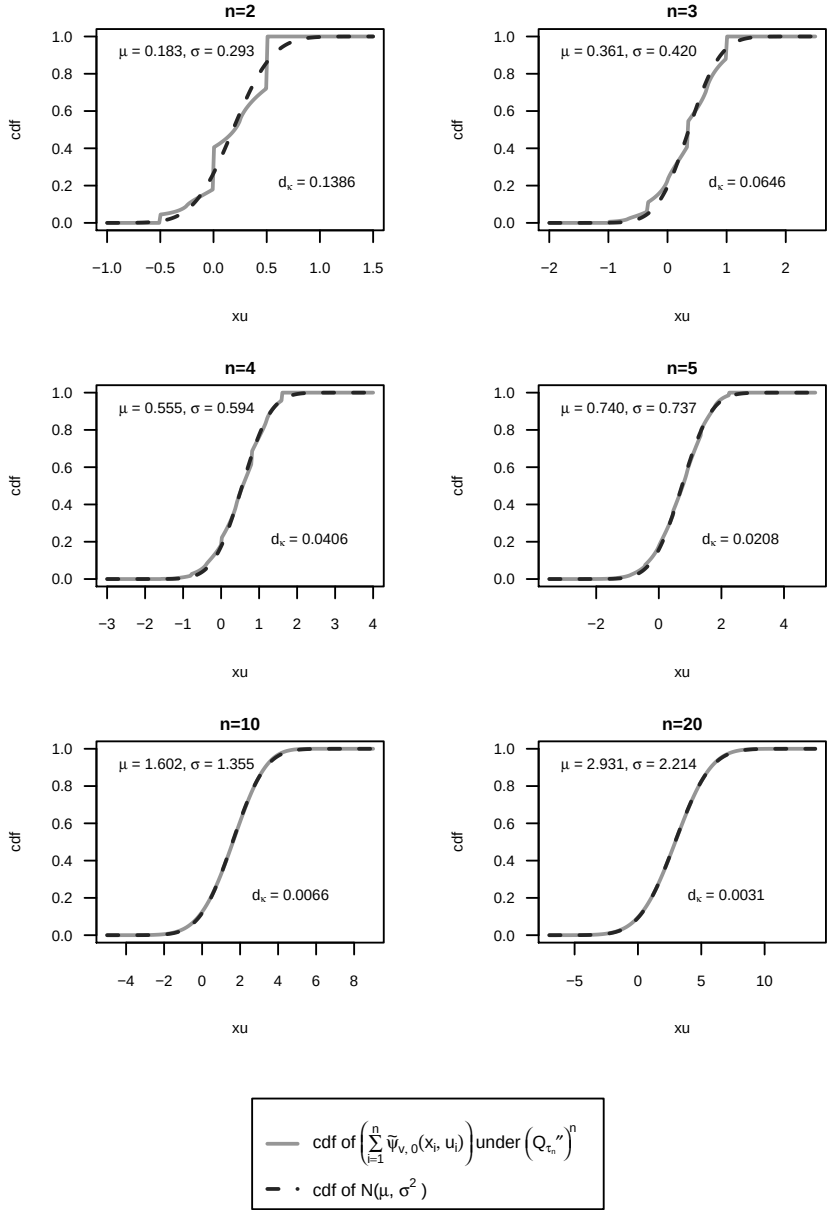

Figure 12.26: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\delta = 0.25$ and $\tau = 2.576$ in case of unconditional total variation neighborhoods $(* = v,\, t = 0)$.

12.3.3 Conditional Contamination Neighborhoods

12.3.3.1 Hampel-Krasker Estimator

The so-called Hampel-Krasker bounded influence estimator in case of normal linear regression has an IC of form

$$Axu \min\left\{1, \frac{b}{|xu|}\right\} \tag{12.3.3}$$

That is, its clipping function $b(x)$ is $b(x) = b/|x|$. This estimator arises as optimal in two ways. First, it is the finite-sample and asymptotic minimax estimator for unconditional contamination neighborhoods; confer Subsubsections 12.1.1.1 and 12.1.1.2. This result contradicts the conjecture of Huber (1983) (cf. Sections 1, 7 and Rejoinder) that the Hampel-Krasker estimator is not concerned with errors-in-variables. Second, in case of conditional contamination neighborhoods it is the finite-sample minimax estimator for contamination curve $\varepsilon_{n,\mathrm{HK}}(x)$ determined by

$$\frac{\varepsilon_{n,\mathrm{HK}}(x)}{1-\varepsilon_{n,\mathrm{HK}}(x)} = \exp\big(-2\tau_n b\big)\Phi\bigg(\tau_n|x| - \frac{b}{|x|}\bigg) - \Phi\bigg(-\tau_n|x| - \frac{b}{|x|}\bigg) \tag{12.3.4}$$

(cf. equation (12.2.5)), respectively the asymptotic minimax estimator for contamination curve $\varepsilon_{\mathrm{HK}}(x)$ determined by

$$\varepsilon_{\mathrm{HK}}(x) = 2\tau\bigg[|x|\varphi\bigg(\frac{b}{|x|}\bigg) - b\Phi\bigg(-\frac{b}{|x|}\bigg)\bigg] \tag{12.3.5}$$

(cf. equation (12.2.14)). Using Taylor expansions one obtains

$$\varepsilon_{n,\mathrm{HK}}(x) \sim \sqrt{2/\pi}\,\tau_n|x|\exp\bigg(-\frac{b^2}{2|x|^2}\bigg) \qquad \text{as } |x| \to 0 \tag{12.3.6}$$

$$\varepsilon_{n,\mathrm{HK}}(x) \sim \big(1+\exp(2\tau_n b)\big)^{-1} \qquad \text{as } |x| \to \infty \tag{12.3.7}$$

confer Section 5 (b) of Rieder (1989). The second approximation is also given in equation (5.6) of Huber (1983). Furthermore, we obtain by Taylor expansions

$$\varepsilon_{\mathrm{HK}}(x) \sim \sqrt{2/\pi}\,\tau|x|\exp\bigg(-\frac{b^2}{2|x|^2}\bigg) \qquad \text{as } |x| \to 0 \tag{12.3.8}$$

$$\varepsilon_{\mathrm{HK}}(x) \sim \sqrt{2/\pi}\,\tau|x| \qquad \text{as } |x| \to \infty \tag{12.3.9}$$

where the second approximation is also given on p 70 of Huber (1983). As these approximations indicate and as we see in Figures 12.27 and 12.28, the contamination curves are very similar for small $|x|$. However, for large $|x|$ an enormous contamination is needed to obtain the Hampel-Krasker estimator in the asymptotic setup. In the finite-sample setup this huge amount of contamination can only be realized for (very) large sample sizes n. A more detailed discussion is given in Section 6 of Huber (1983). We also refer to the various comments on pages 72–80 following Huber (1983).

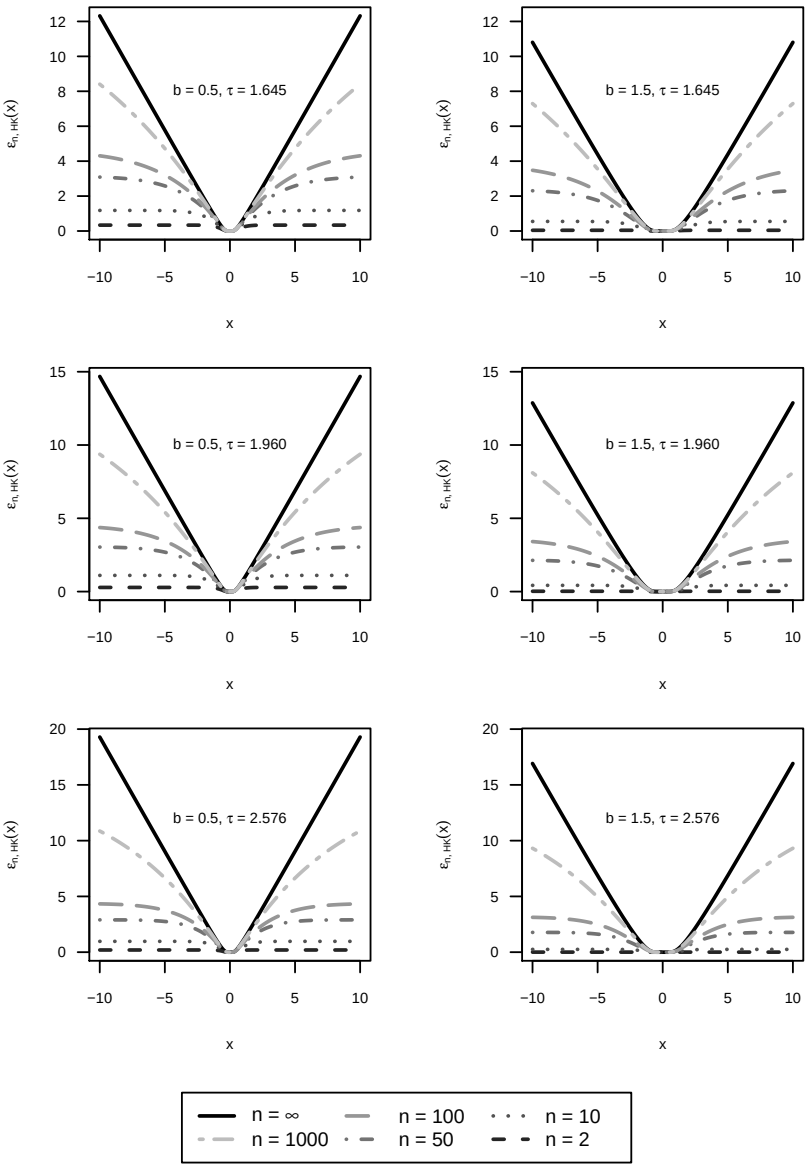

Figure 12.27: Contamination curve of the Hampel-Krasker estimator for given $b(x) = b/|x| = 0.5/|x|, 1.5/|x|$ $(x \in [-10, 10])$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods $(* = c, t = \varepsilon)$.

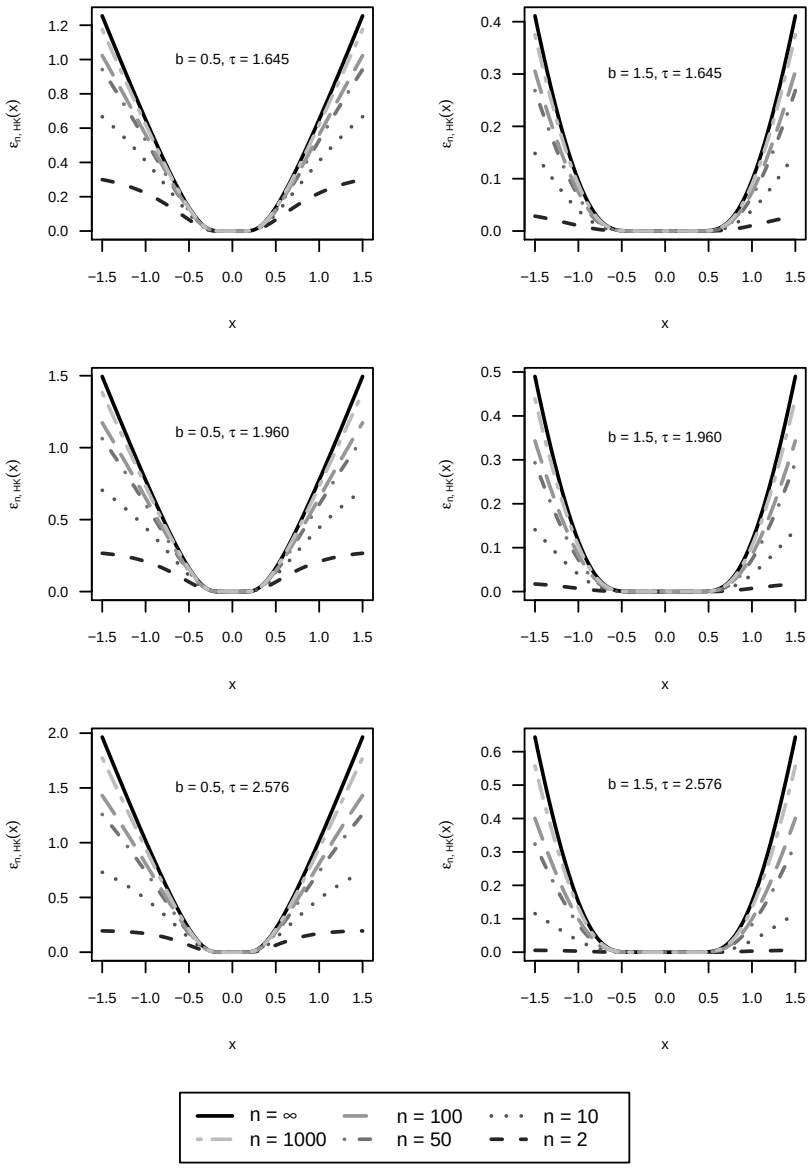

Figure 12.28: Contamination curve of the Hampel-Krasker estimator for given $b(x) = b/|x| = 0.5/|x|, 1.5/|x|$ $(x \in [-1.5, 1.5])$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$.

12.3.3.2 Huber Estimator

The so-called Huber estimator in case of normal linear regression has an IC of form

$$Axu \min\left\{1, \frac{b}{|u|}\right\} \tag{12.3.10}$$

i.e., its clipping function $b(x)$ is $b(x) \equiv b$. Hence, the Huber estimator bounds only the residuals irrespectively of the value of the regressor x. In case of conditional contamination neighborhoods it is the finite-sample minimax estimator for contamination curve $\varepsilon_{n,\mathrm{HU}}(x)$ determined by

$$\frac{\varepsilon_{n,\mathrm{HK}}(x)}{1-\varepsilon_{n,\mathrm{HK}(x)}} = \exp\big(-2\tau_n|x|b\big)\Phi\big(\tau_n|x|-b\big) - \Phi\big(-\tau_n|x|-b\big) \tag{12.3.11}$$

(cf. equation (12.2.5)), respectively the asymptotic minimax estimator for contamination curve $\varepsilon_{\mathrm{HU}}(x)$ determined by

$$\varepsilon_{\mathrm{HU}}(x) = 2\tau|x|\big[\varphi\big(b\big) - b\Phi\big(-b\big)\big] \tag{12.3.12}$$

(cf. equation (12.2.14)). That is, we obtain the Huber estimator in the asymptotic setup if most of the contamination is put on large values of $|x|$. Furthermore, using Taylor expansions one obtains in the finite-sample setup

$$\varepsilon_{n,\mathrm{HU}}(x) \sim \sqrt{2/\pi}\,\tau_n|x| \exp\left(-\frac{b^2}{2}\right) \qquad \text{as } |x| \to 0 \tag{12.3.13}$$

$$\varepsilon_{n,\mathrm{HU}}(x) \sim \exp\big(-2\tau_n|x|b\big) \qquad \text{as } |x| \to \infty \tag{12.3.14}$$

confer Section 5 (c) of Rieder (1989) where the second approximation is also given in equation (5.7) of Huber (1983). As the approximation for small $|x|$ indicates and as we see in Figures 12.27 and 12.28, the contamination curves are very similar for small $|x|$. However, they behave very different for large $|x|$. That is, we obtain the Huber estimator in the finite-sample setup if the amount of contamination decreases exponentially for sufficiently large $|x|$ which is contrary to the asymptotic setup where an enormous amount of contamination is needed (proportional to $|x|$). Again, we refer to the discussion given in Section 6 of Huber (1983) as well as to the various comments on pages 72–80 following Huber (1983). Since minimizing the asymptotic under-/overshoot confidence risk is equivalent to minimizing the asymptotic variance subject to a bias bound (cf. Lemma 10.3.5), the results about contamination saddle points given in Subsection 7.5.2 of Rieder (1994) apply to our setup, too. Hence, we now quote Remark 7.5.17 (ibid.); in our special case $\Gamma(z) = 2\tau(\varphi(z) - z\Phi(-z))$ and we have to replace $|Ax|$ by $|x|$ and $\Lambda_f(u)$ by u.

Remark 12.3.3 The Huber estimator, when compared with the Hampel-Krasker estimator, is minimax in a similar though slightly smaller model: square, as opposed to mean, conditional contamination. The difference is reflected by their least favorable contamination curves, but only after standardization by the linear rate: Thus one gets a constant, respectively $\Gamma(b/|x|)$ increasing from 0 to its finite maximum $\Gamma(0) = \tau\,\mathrm{E}\,|u|$ as $|x|$ goes from 0 to ∞. In this relative sense, Hampel-Krasker safeguards against contamination at leverage points, more than Huber does. ////

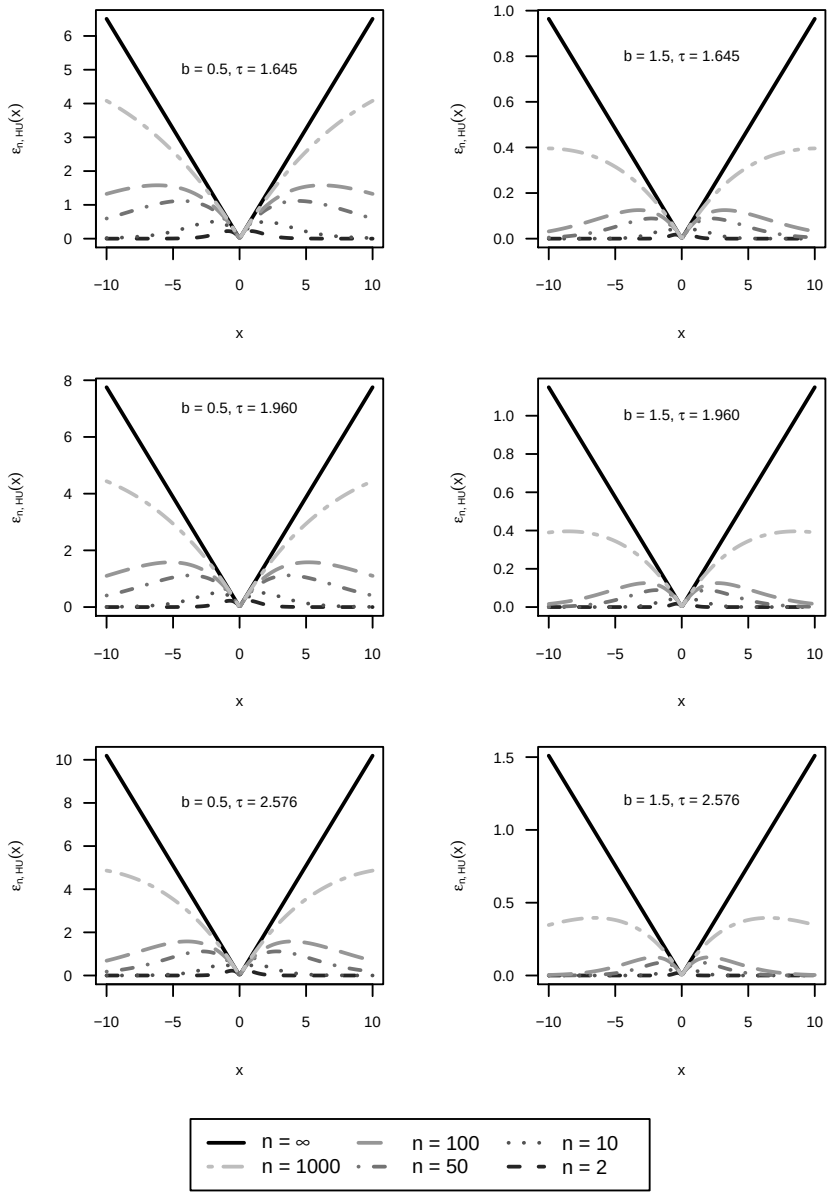

Figure 12.29: Contamination curve of the Huber estimator for $b(x) \equiv 0.5, 1.5$ ($x \in [-10, 10]$) and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

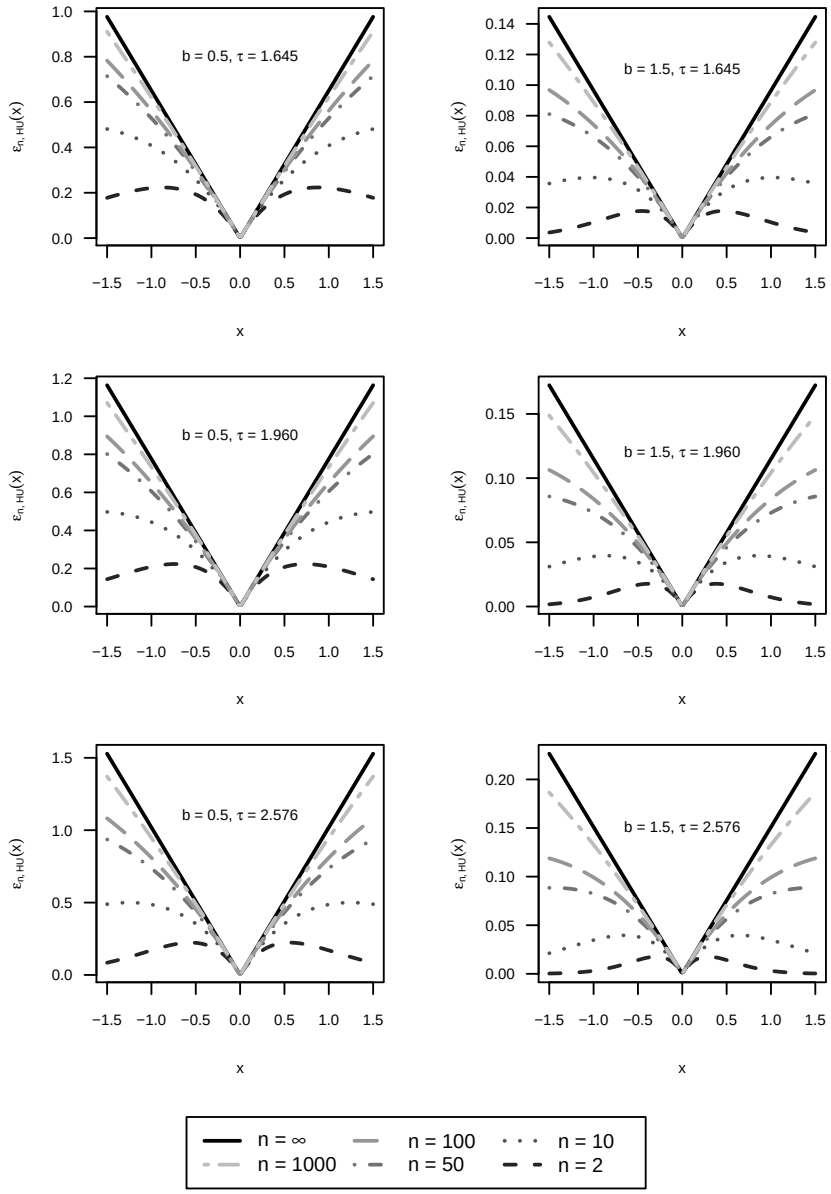

Figure 12.30: Contamination curve of the Huber estimator for $b(x) \equiv 0.5, 1.5$ ($x \in [-1.5, 1.5]$) and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

12.3.3.3 Optimal Clipping Function

We now consider constant contamination curves. In the finite-sample setup ε_n is characterized by a clipping function $b^{\mathrm{fi}}_{c,\varepsilon}(x)$ which is determined by (12.2.4) and (12.2.5). In particular, one obtains

$$b^{\mathrm{fi}}_{c,\varepsilon}(x) = 0 \qquad \text{if } |x| \le \tau_n^{-1}\Phi^{-1}\left(\frac{1}{2(1-\varepsilon_n)}\right) \tag{12.3.15}$$

$$b^{\mathrm{fi}}_{c,\varepsilon}(x) \sim \frac{1}{2\tau_n|x|}\log\left(\frac{1-\varepsilon_n}{\varepsilon_n}\right) \qquad \text{as } |x| \to \infty \tag{12.3.16}$$

confer Section 5 (a) of Rieder (1989) and equations (5.3) and (5.4) of Huber (1983), respectively. In the asymptotic setup a constant contamination curve ε corresponds to a clipping function $b^{\mathrm{as}}_{c,\varepsilon}(x)$ which is determined by (12.2.13) and (12.2.14). That is,

$$b^{\mathrm{as}}_{c,\varepsilon}(x) = 0 \qquad \text{if } |x| \le \sqrt{\frac{\pi}{2}}\,\frac{\varepsilon}{\tau} \tag{12.3.17}$$

$$\frac{\varepsilon}{2\tau|x|} = \varphi\big(b^{\mathrm{as}}_{c,\varepsilon}(x)\big) - b^{\mathrm{as}}_{c,\varepsilon}(x)\Phi\big(-b^{\mathrm{as}}_{c,\varepsilon}(x)\big) \qquad \text{if } |x| > \sqrt{\frac{\pi}{2}}\,\frac{\varepsilon}{\tau} \tag{12.3.18}$$

Consequently, $b^{\mathrm{as}}_{c,\varepsilon}$ increases with increasing $|x|$ and ε and may become arbitrarily large; confer also Figure 12.31. As in Subsubsections 12.3.3.1 and 12.3.3.2, we refer to the discussion given in Section 6 of Huber (1983) as well as to the various comments on pages 72–80 following Huber (1983).

In Figure 12.32 one can additionally find the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping function which can be determined via (12.2.19). In all cases considered $b^{\mathrm{as.c}}_{c,\varepsilon}$ lies below the finite-sample optimal clipping function and is even negative for small and large values of $|x|$, respectively. However, for moderate values of $|x|$ and not too small sample sizes n the approximation seems to work quite well.

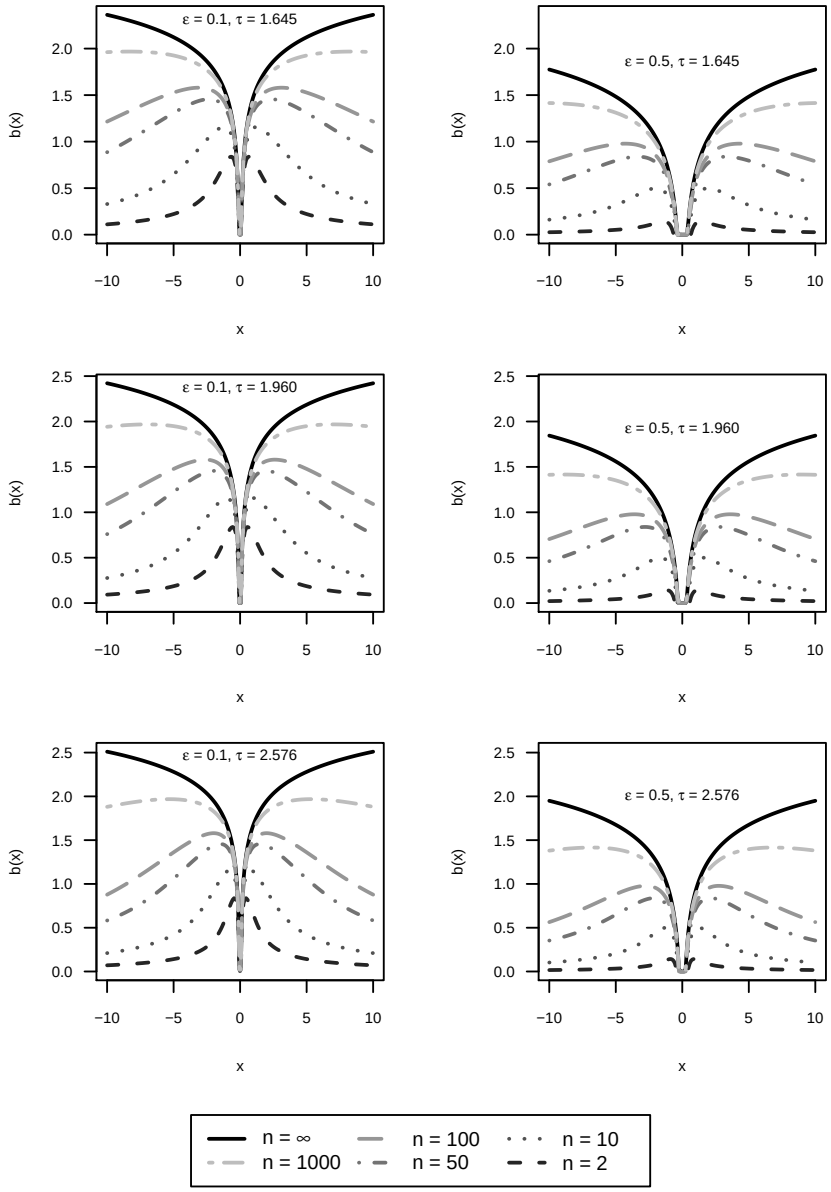

Figure 12.31: Finite-sample and asymptotic optimal clipping function for $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

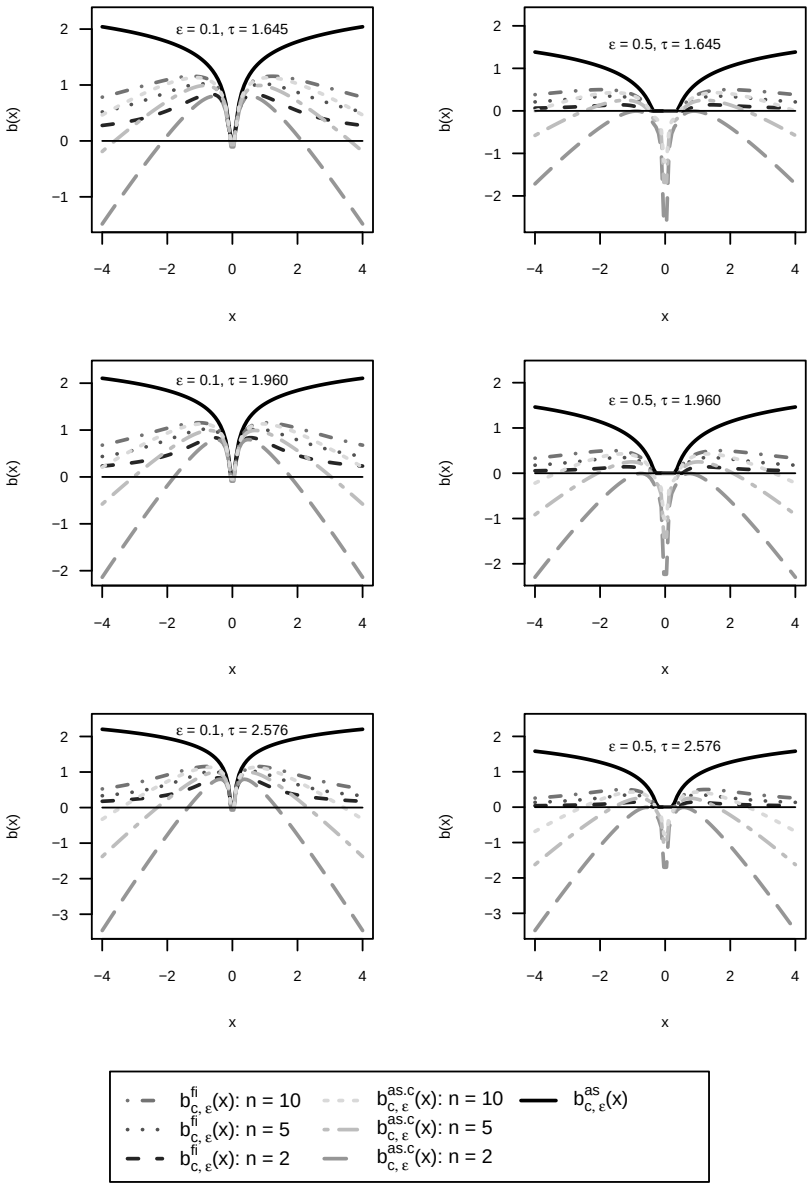

Figure 12.32: Finite-sample, $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal and asymptotic optimal clipping function for $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

12.3.3.4 Finite-Sample Risk

We consider constant contamination curves; i.e., $\varepsilon(x) \equiv 0.1$ and $\varepsilon(x) \equiv 0.5$, respectively. Given these contamination curves we determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_{c,\varepsilon}^{\text{fi}}$, the asymptotic minimax estimator $\tilde{S}_{c,\varepsilon}^{\text{as}}$ and the estimator $S_{c,\varepsilon}^{\text{as.c}}$ which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping function. Since clipping functions have to be non-negative, negative values of $b_{c,\varepsilon}^{\text{as.c}}$ are set to 0.

As we see in Figure 12.32 there are clear differences between the optimal clipping functions, in particular, for large values of $|x|$. However, in case of the regressor distributions considered we have $|x| \leq 2$. Thus, not surprisingly, the differences concerning the corresponding finite-sample risks are only small; see Figures 12.33 and 12.36. The only exceptions are the finite-sample risks of the estimator which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound for small sample sizes n.

Figures 12.34 and 12.37 show the speed of convergence for the finite-sample risks. It seems to be of order $n^{-1/2}$ as in case of the optimal clipping functions; confer Lemma 12.2.1. To get a better impression, we consider $y = \text{Risk}_{\text{as}}^{\natural} - \text{Risk}_{\text{fi}}^{\natural}$ and apply the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002); i.e., we estimate lambda by means of maximum likelihood such that $y^{\text{lambda}} \approx n$. That is, lambda ≈ -2 indicates $y = \mathrm{O}(n^{-1/2})$. Indeed, the estimated values of lambda are close to -2 which indicates a convergence of order $n^{-1/2}$; see Figures 12.35 and 12.38. We conjecture that the relatively slow convergence in case $K = \text{Unif}([-1, 2])$ and $\varepsilon(x) \equiv 0.5$ is caused by the fact that the finite-sample optimal clipping bound is equal to 0 for $|x| \leq \tau_n^{-1}\Phi^{-1}\big(1/(2(1-\varepsilon_n(x)))\big)$.

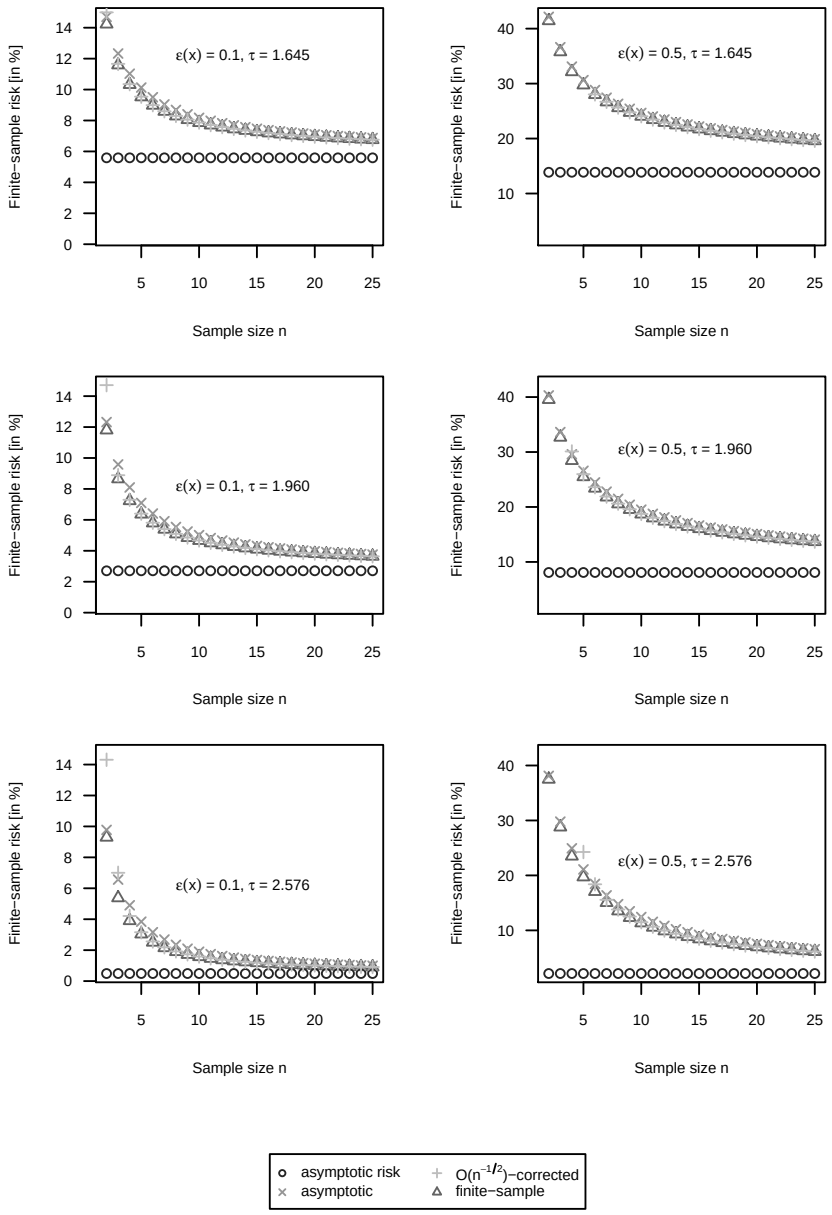

Figure 12.33: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \le 25$, $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$.

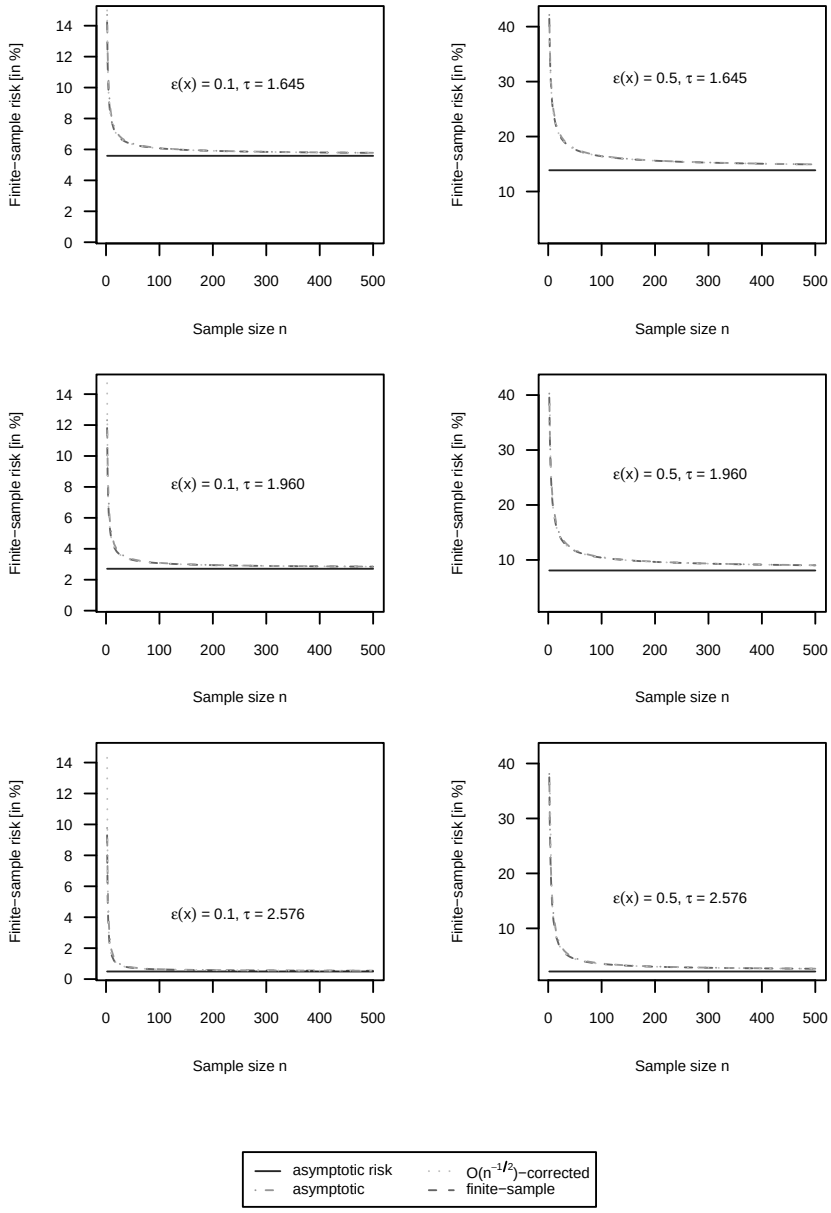

Figure 12.34: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, increasing sample size n, $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$.

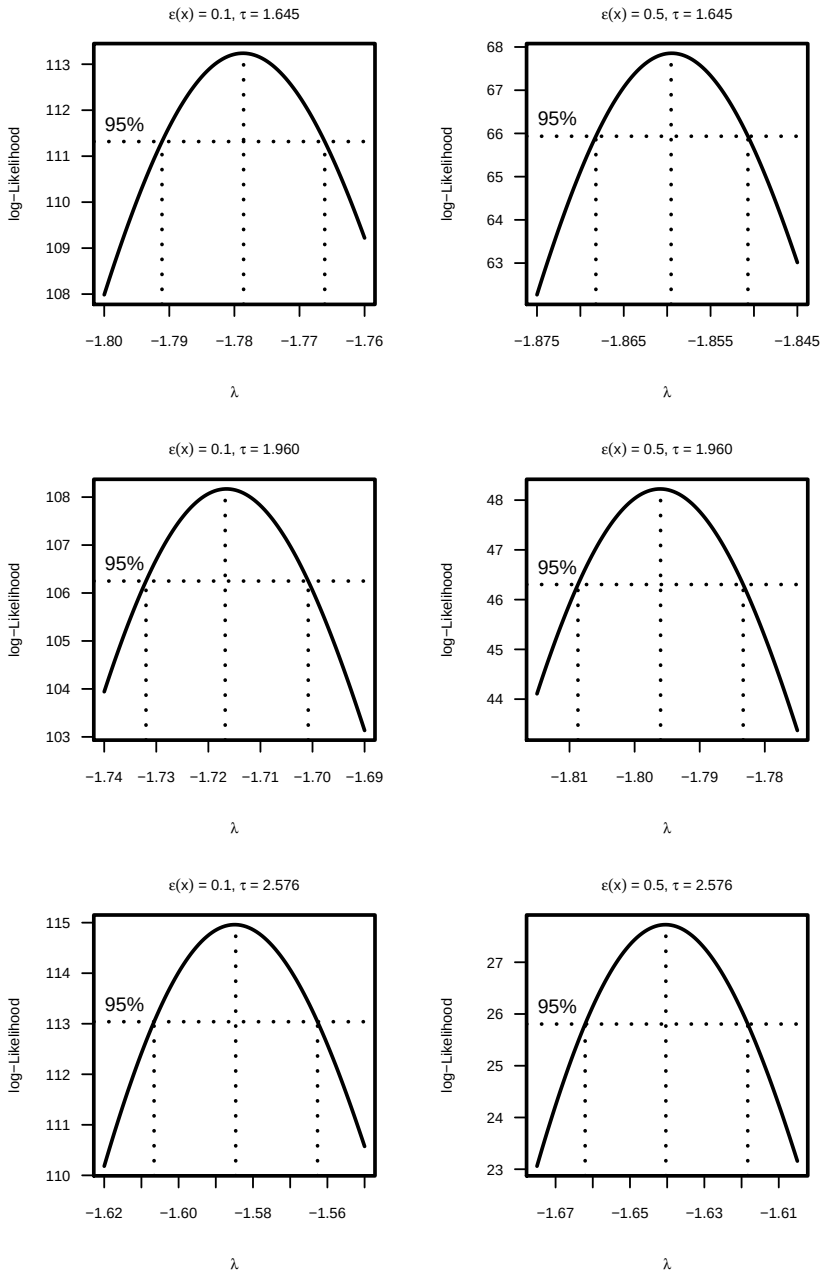

Figure 12.35: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$.

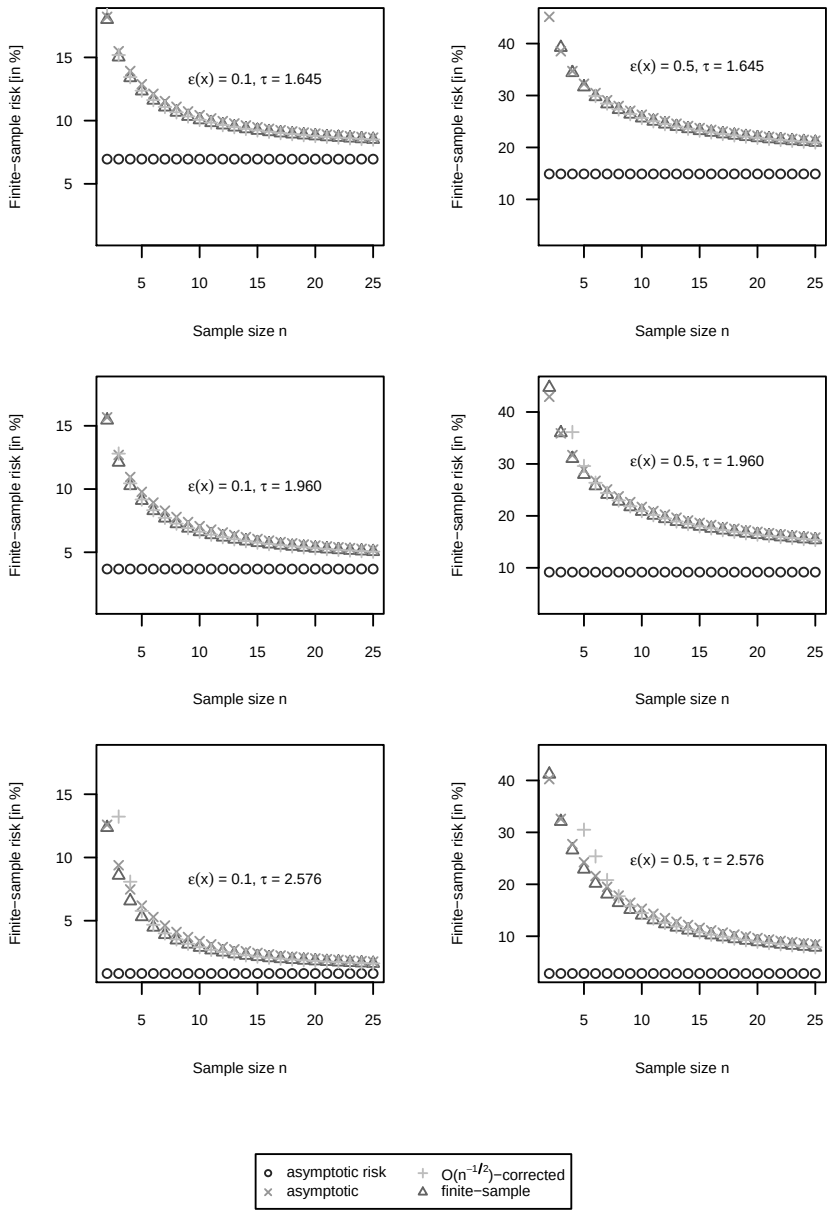

Figure 12.36: Finite-sample risk for $K = \text{Unif}([-1,2])$, $n \le 25$, $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

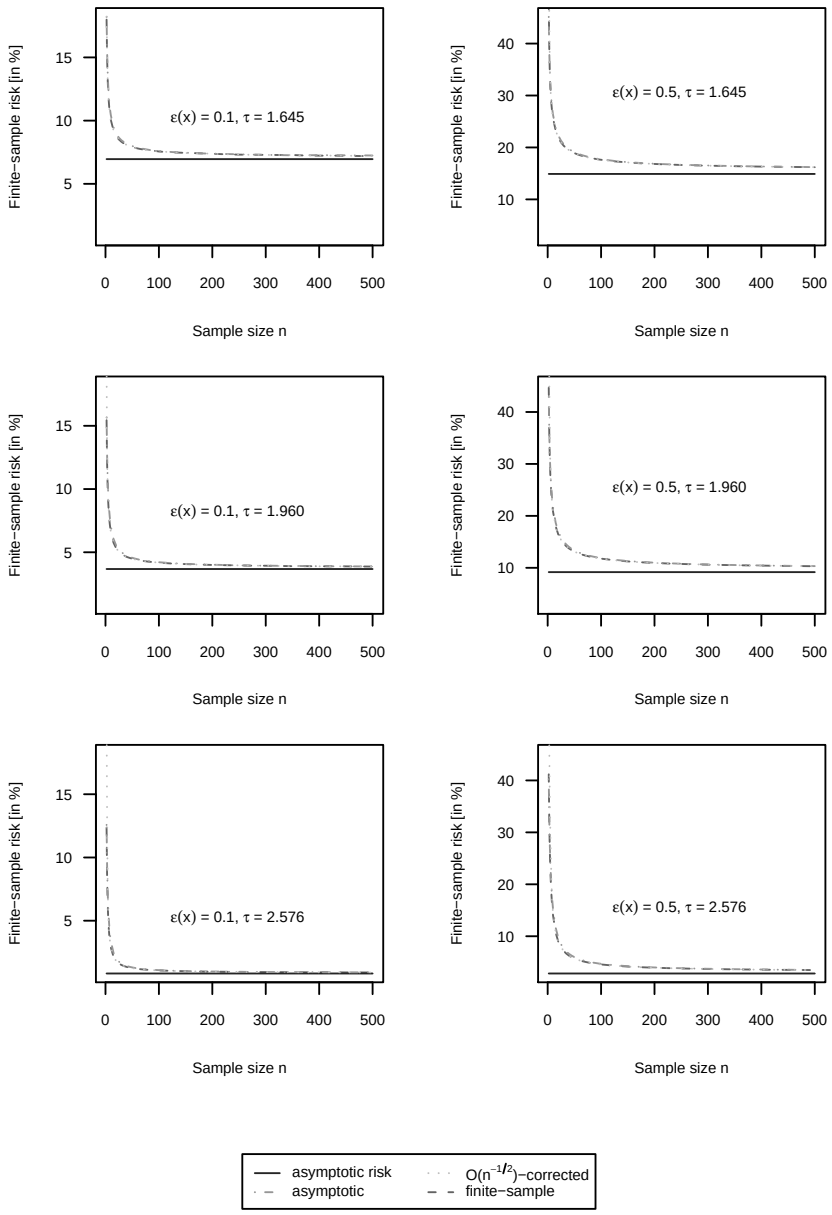

Figure 12.37: Finite-sample risk for $K = \text{Unif}([-1,2])$, increasing sample size n, $\varepsilon(x) \equiv 0.1, 0.5$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

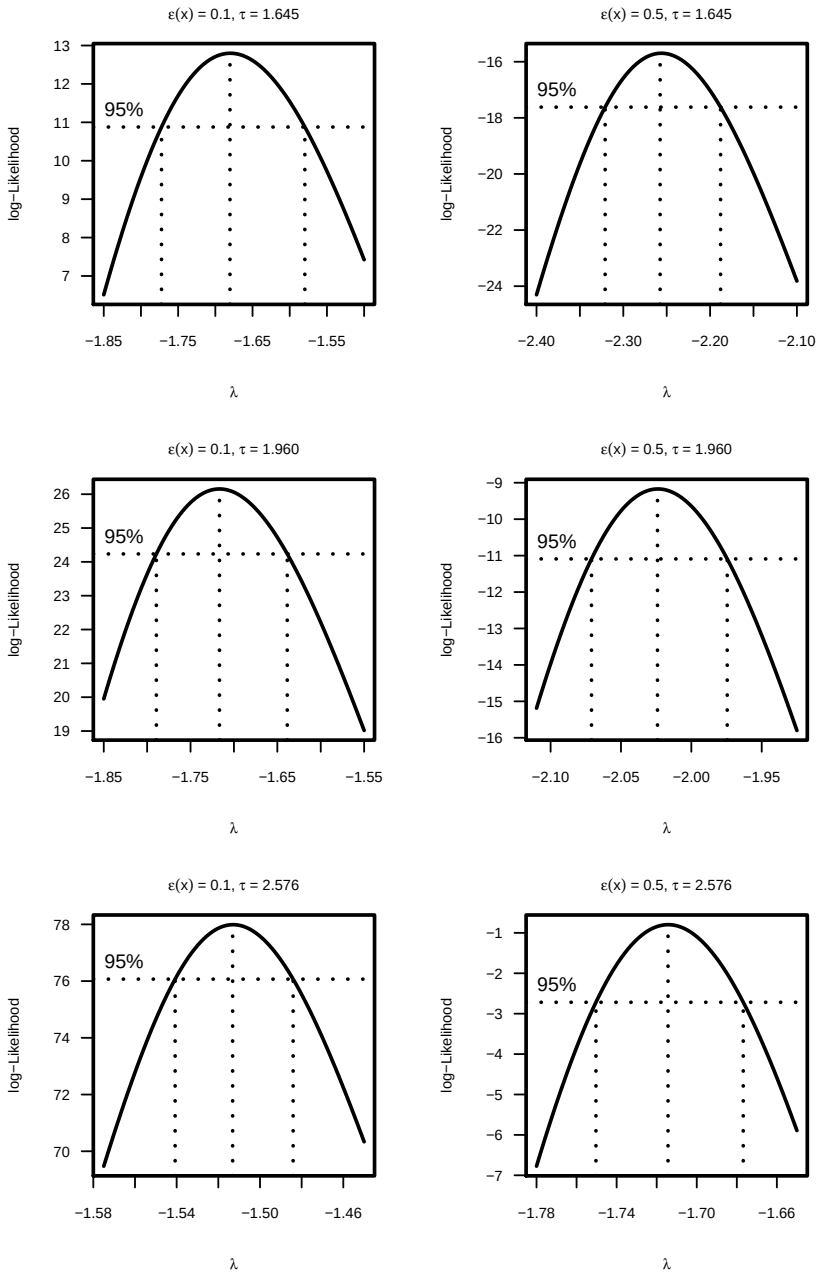

Figure 12.38: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \mathrm{Unif}([-1, 2])$ and conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

12.3.3.5 Finite-Sample Distribution

As before, we use Algorithm A (cf. Subsubsection 12.2.3.2) to compute the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{c,\varepsilon}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance. By symmetry it suffices to consider the cumulative distribution function of $\sum_{i=1}^{n} \tilde{\psi}_{c,\varepsilon}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$. Moreover, we choose constant contamination curves and consider only the two "extreme" situations $\varepsilon(x) \equiv 0.1\,,\ \tau = 1.645$ (see Figures 12.39 and 12.41) and $\varepsilon(x) \equiv 0.5\,,\ \tau = 2.576$ (see Figures 12.40 and 12.42).

Remark 12.3.4 If $H''_{\tau_n}(dy \,|\, x)$ is absolutely continuous a.e. $H''_{\tau_n}(dx)$, then by Remark 12.2.4 (b) the distribution of $\tilde{S}^{\text{fi}}_{c,\varepsilon}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^{n} \tilde{\psi}_{c,\varepsilon}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. ////

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^{n} \tilde{\psi}_{c,\varepsilon}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ and of $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the `R` function `optim`; confer R Development Core Team (2005).
As we see, in all cases about 5 observations are enough to get already quite close to a normal distribution (i.e., $d_\kappa < 0.02$ for $n \geq 5$) where the jumps included in the cumulative distribution functions decay exponentially in n; confer also Remark 11.3.3. In particular, the Kolmogorov distances are slightly smaller than in all cases considered so far.
In Figure 12.41 the Kolmogorov distance is not monotone decreasing. Since the corresponding values are very small, we conjecture this is just a matter of numerical precision.

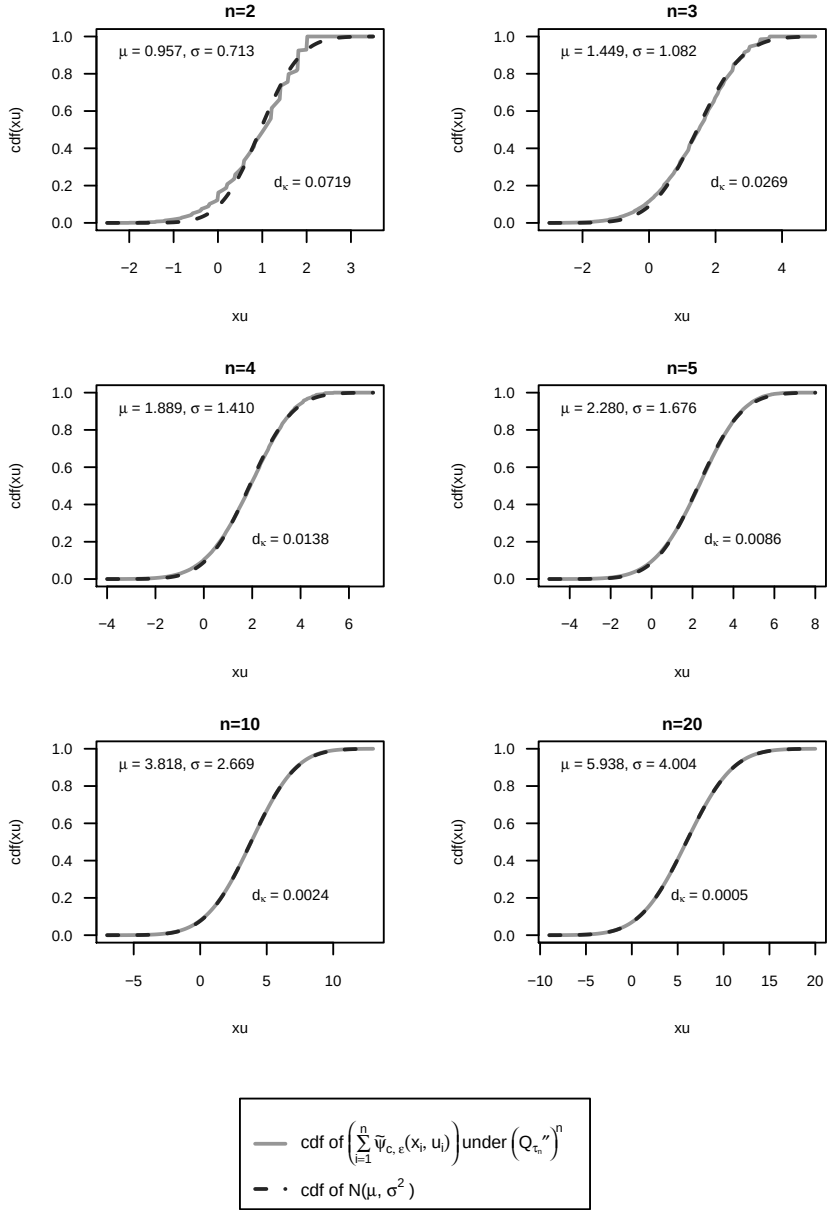

Figure 12.39: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\varepsilon(x) \equiv 0.1$ and $\tau = 1.645$ in case of conditional contamination neighborhoods $(* = c,\, t = \varepsilon)$.

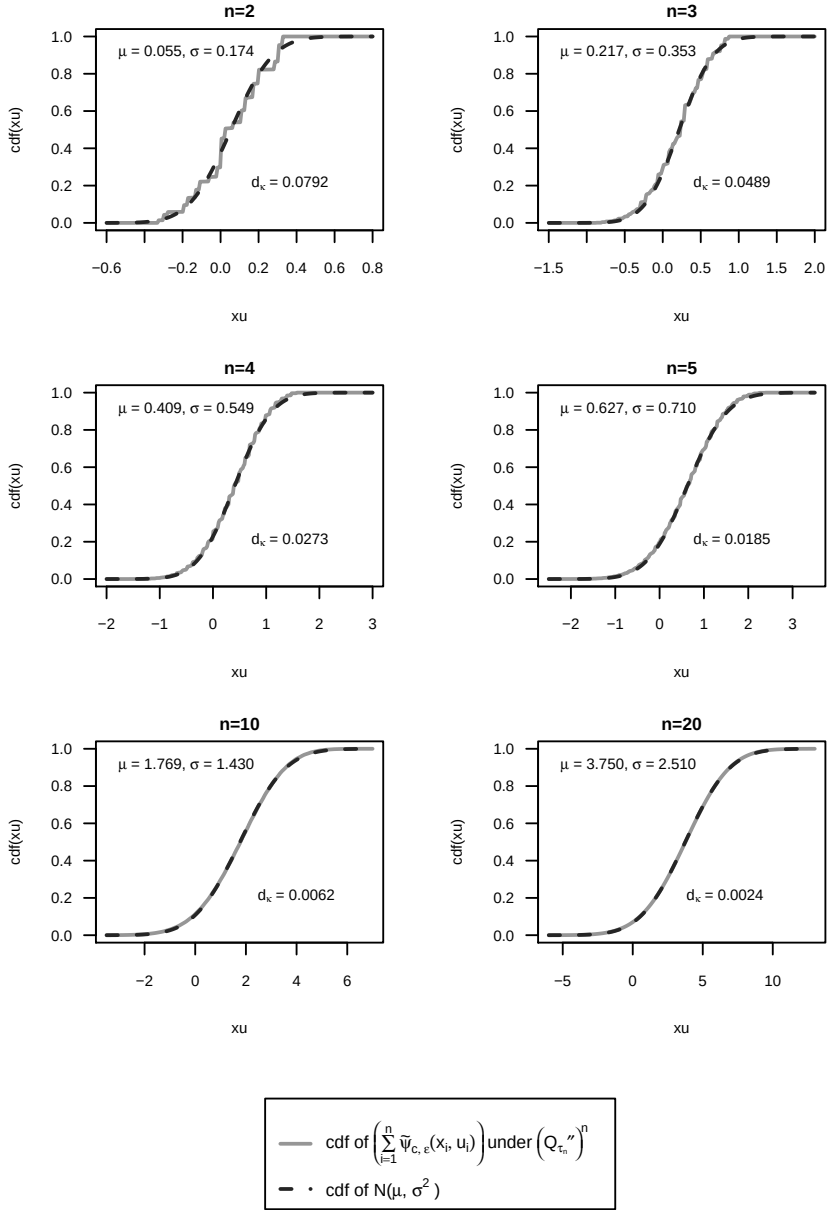

Figure 12.40: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\varepsilon(x) \equiv 0.5$ and $\tau = 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

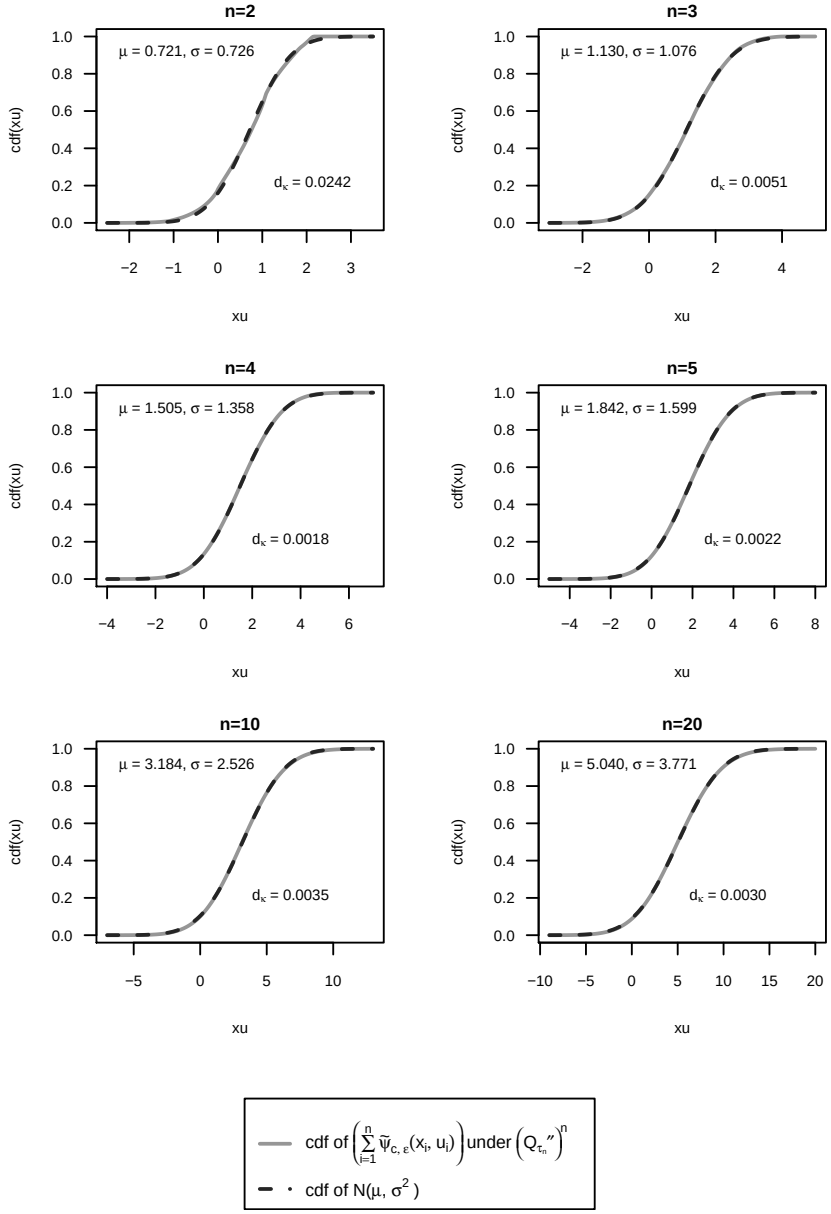

Figure 12.41: Finite-sample distributions for $K = \text{Unif}([-1,2])$, $\varepsilon(x) \equiv 0.1$ and $\tau = 1.645$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

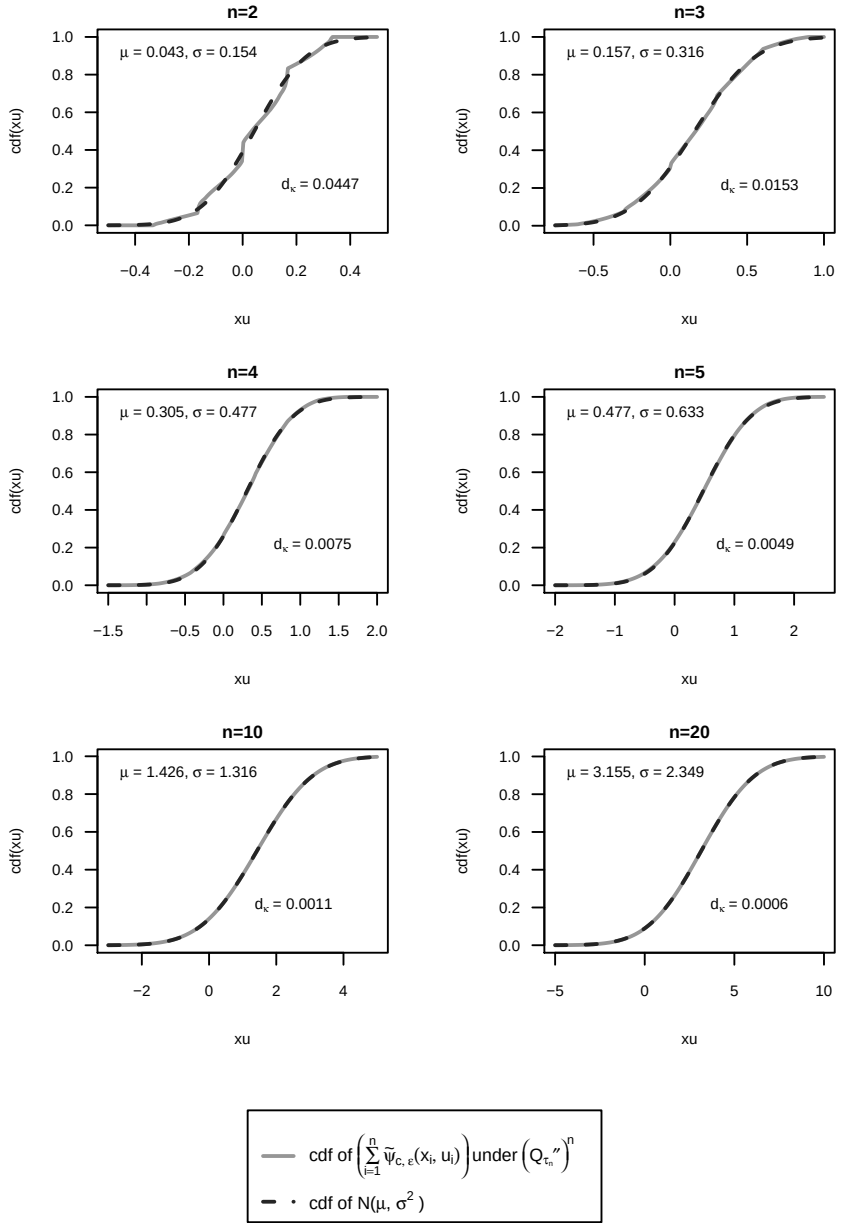

Figure 12.42: Finite-sample distributions for $K = \text{Unif}([-1,2])$, $\varepsilon(x) \equiv 0.5$ and $\tau = 2.576$ in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$).

12.3.4 Conditional Total Variation Neighborhoods

12.3.4.1 Hampel-Krasker Estimator

Like in case of conditional contamination neighborhoods (cf. Subsubsection 12.3.3.1) the Hampel-Krasker bounded influence estimator estimator arises as optimal in two ways. First, it is the finite-sample and asymptotic minimax estimator for unconditional total variation neighborhoods; confer Subsubsections 12.1.2.1 and 12.1.2.2. This result again contradicts the conjecture of Huber (1983) (cf. Sections 1, 7 and Rejoinder) that the Hampel-Krasker estimator is not concerned with errors-in-variables. Second, in case of conditional total variation neighborhoods it is the finite-sample minimax estimator for contamination curve $\delta_{n,\mathrm{HK}}(x)$ determined by

$$\big[1+\exp\big(-2\tau_n b\big)\big]\delta_{n,\mathrm{HK}}(x)=\exp\big(-2\tau_n b\big)\Phi\bigg(\tau_n|x|-\frac{b}{|x|}\bigg)-\Phi\bigg(-\tau_n|x|-\frac{b}{|x|}\bigg) \tag{12.3.19}$$

(cf. equation (12.2.27)), respectively the asymptotic minimax estimator for contamination curve $\delta_{\mathrm{HK}}(x)$ determined by

$$\delta_{\mathrm{HK}}(x)=\tau\bigg[|x|\varphi\bigg(\frac{b}{|x|}\bigg)-b\Phi\bigg(-\frac{b}{|x|}\bigg)\bigg] \tag{12.3.20}$$

(cf. equation (12.2.36)). Moreover, using Taylor expansions we obtain

$$\delta_{n,\mathrm{HK}}(x)\sim 0 \qquad \text{as } |x|\to 0 \tag{12.3.21}$$

$$\delta_{n,\mathrm{HK}}(x)\sim\big(1+\exp(2\tau_n b)\big)^{-1} \qquad \text{as } |x|\to\infty \tag{12.3.22}$$

and

$$\delta_{\mathrm{HK}}(x)\sim\frac{1}{\sqrt{2\pi}}\tau|x|\exp\bigg(-\frac{b^2}{2|x|^2}\bigg) \qquad \text{as } |x|\to 0 \tag{12.3.23}$$

$$\delta_{\mathrm{HK}}(x)\sim\frac{1}{\sqrt{2\pi}}\tau|x| \qquad \text{as } |x|\to\infty \tag{12.3.24}$$

As these approximations indicate and as we see in Figures 12.43 and 12.44, the contamination curves are very similar for small $|x|$. However, for large $|x|$ an enormous contamination is needed to obtain the Hampel-Krasker estimator in the asymptotic setup whereas in the finite-sample setup only a bounded amount of contamination is needed. In particular, the asymptotic behavior in the finite-sample setup is identical to the asymptotic behavior in the finite-sample setup and conditional contamination neighborhoods; confer equation (12.3.7). Since the results are very similar to the results given in Subsubsection 12.3.4.1, we again refer to the more detailed discussion given in Section 6 of Huber (1983) as well as to the various comments on pages 72–80 following Huber (1983).

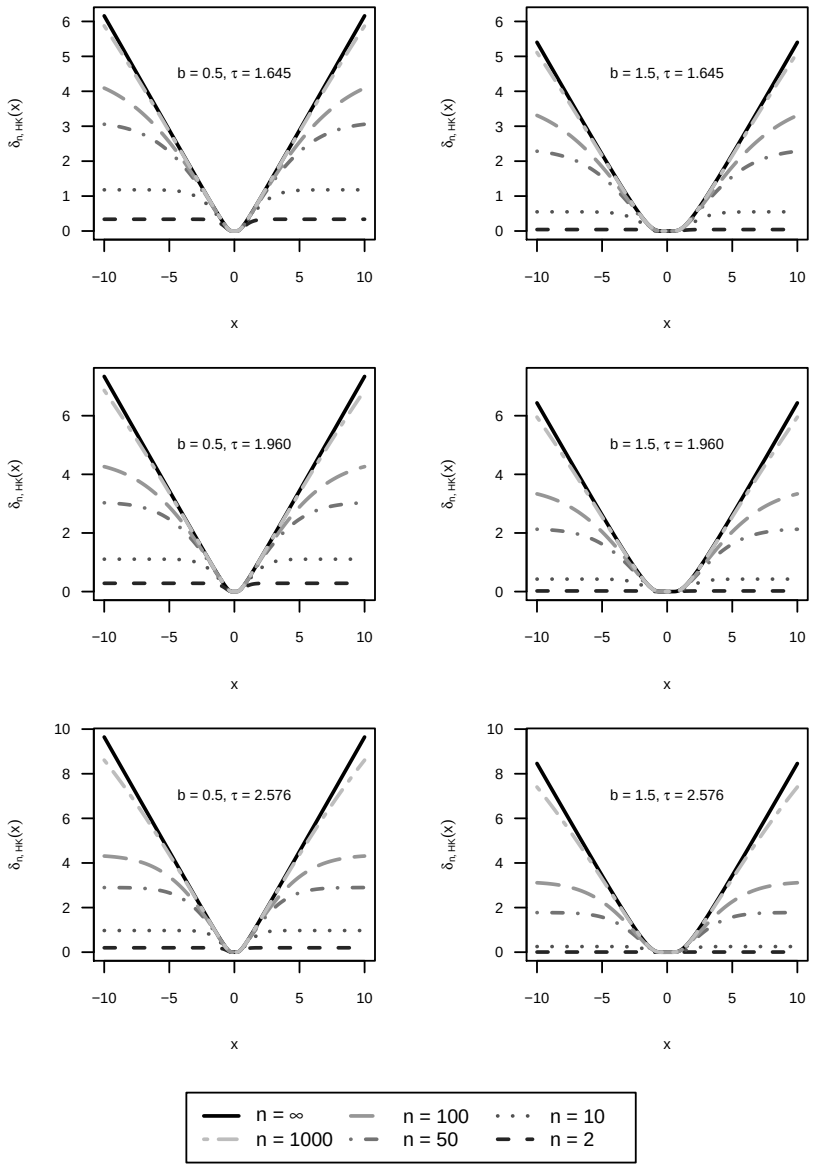

Figure 12.43: Contamination curve of the Hampel-Krasker estimator for $b(x) = b/|x| = 0.5/|x|, 1.5/|x|$ ($x \in [-10, 10]$) and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

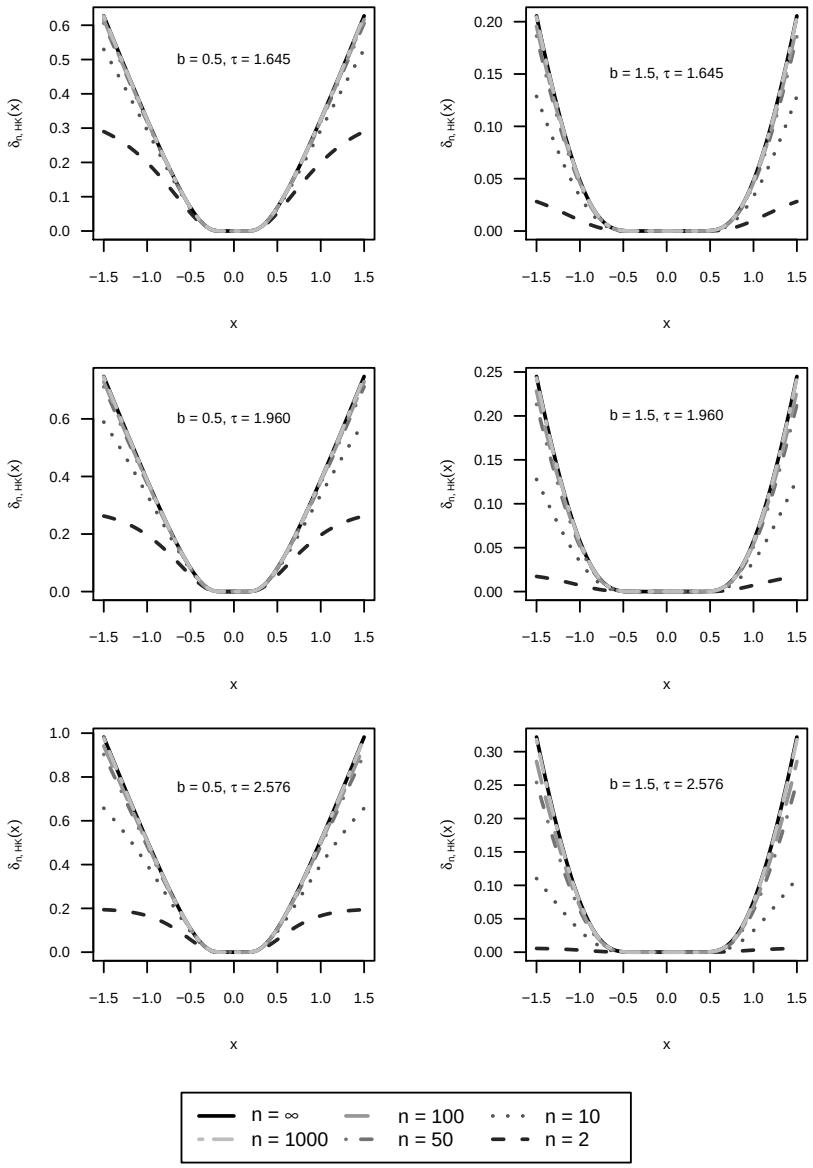

Figure 12.44: Contamination curve of the Hampel-Krasker estimator for $b(x) = b/|x| = 0.5/|x|, 1.5/|x|$ $(x \in [-1.5, 1.5])$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

12.3.4.2 Huber Estimator

Analogously to the case of conditional contamination neighborhoods (cf. Subsubsection 12.3.3.2) the Huber estimator is the finite-sample minimax estimator for contamination curve $\delta_{n,\mathrm{HU}}(x)$ determined by

$$\big[1+\exp\big(-2\tau_n|x|b\big)\big]\delta_{n,\mathrm{HU}}(x)=\exp\big(-2\tau_n|x|b\big)\Phi\big(\tau_n|x|-b\big)-\Phi\big(-\tau_n|x|-b\big) \tag{12.3.25}$$

(cf. equation (12.2.27)), respectively the asymptotic minimax estimator for contamination curve $\delta_{\mathrm{HU}}(x)$ determined by

$$\delta_{\mathrm{HU}}(x)=\tau|x|\big[\varphi\big(b\big)-b\Phi\big(-b\big)\big] \tag{12.3.26}$$

(cf. equation (12.2.36)). That is, we obtain the Huber estimator in the asymptotic setup if most of the contamination is put on large values of $|x|$. Furthermore, using Taylor expansions one obtains

$$\delta_{n,\mathrm{HU}}(x)\sim\tau_n|x|\big[\varphi\big(b\big)-b\Phi\big(-b\big)\big] \qquad \text{as } |x|\to 0 \tag{12.3.27}$$

$$\delta_{n,\mathrm{HU}}(x)\sim 0 \qquad \text{as } |x|\to\infty \tag{12.3.28}$$

As these approximations indicate the contamination curves are almost identical for small $|x|$; see also Figures 12.43 and 12.44. However, they behave very different for large $|x|$. In particular, we obtain the Huber estimator in the finite-sample setup if the amount of contamination is very small (tending to 0) for large $|x|$ whereas in the asymptotic setup it has to be proportional to $|x|$. Since the situation is very similar to Subsubsection 12.3.4.2, we refer to the more detailed discussion given in Section 6 of Huber (1983) as well as to the various comments on pages 72–80 following Huber (1983). Moreover, the results about contamination saddle points derived in Subsection 7.5.2 of Rieder (1994) in our setup apply to conditional total variation neighborhoods, too. For more details we refer to Remark 12.3.3.

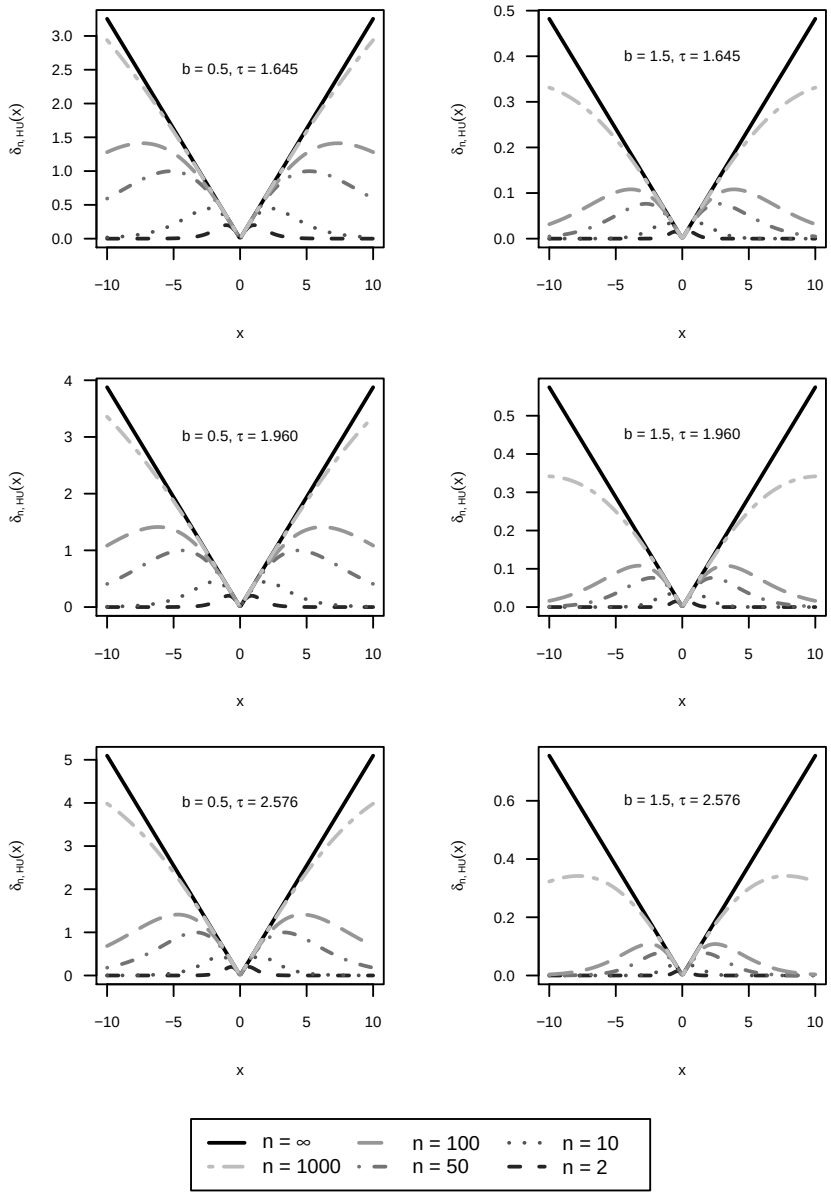

Figure 12.45: Contamination curve of the Huber estimator for $b(x) \equiv 0.5, 1.5$ ($x \in [-10, 10]$) and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

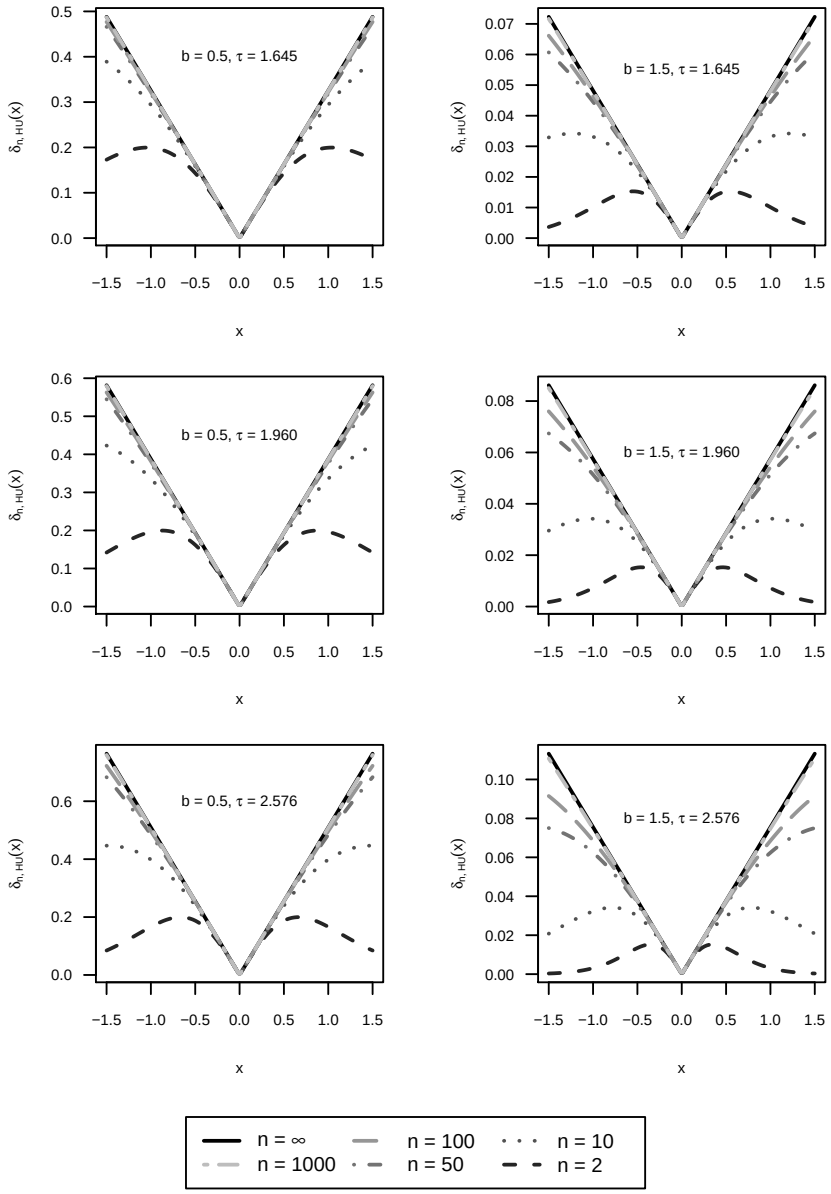

Figure 12.46: Contamination curve of the Huber estimator for $b(x) \equiv 0.5, 1.5$ ($x \in [-1.5, 1.5]$) and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

12.3.4.3 Optimal Clipping Function

We now consider constant contamination curves. In the finite-sample setup δ_n is characterized by a clipping function $b^{\mathrm{fi}}_{v,\delta}(x)$ which is determined by (12.2.26) and (12.2.27). In particular, one obtains

$$b^{\mathrm{fi}}_{v,\delta}(x) = 0 \qquad \text{if } |x| \leq \tau_n^{-1}\Phi^{-1}\big(0.5 + \delta_n\big) \tag{12.3.29}$$

$$b^{\mathrm{fi}}_{v,\delta}(x) \sim \frac{1}{2\tau_n|x|}\log\left(\frac{1-\delta_n}{\delta_n}\right) \qquad \text{as } |x| \to \infty \tag{12.3.30}$$

That is, the asymptotic behavior of $b^{\mathrm{fi}}_{v,\delta}$ is identical to the asymptotic behavior of $b^{\mathrm{fi}}_{c,\varepsilon}$; confer equation (12.3.16). In the asymptotic setup a constant contamination curve δ corresponds to a clipping function $b^{\mathrm{as}}_{v,\delta}(x)$ which is determined by (12.2.35) and (12.2.36). That is,

$$b^{\mathrm{as}}_{v,\delta}(x) = 0 \qquad \text{if } |x| \leq \frac{1}{\sqrt{2\pi}}\frac{\delta}{\tau} \tag{12.3.31}$$

$$\frac{\delta}{\tau|x|} = \varphi\big(b^{\mathrm{as}}_{v,\delta}(x)\big) - b^{\mathrm{as}}_{v,\delta}(x)\Phi\big(-b^{\mathrm{as}}_{v,\delta}(x)\big) \qquad \text{if } |x| > \frac{1}{\sqrt{2\pi}}\frac{\delta}{\tau} \tag{12.3.32}$$

Consequently, $b^{\mathrm{as}}_{v,\delta}$ increases with increasing $|x|$ and δ and may become arbitrarily large (cf. Figure 12.47); i.e., the asymptotic behavior $b^{\mathrm{as}}_{v,\delta}$ is very similar to the asymptotic behavior of $b^{\mathrm{as}}_{c,\varepsilon}$. Since the results are very similar to the results in case of conditional contamination neighborhoods (cf. Subsubsection 12.3.3.3), we refer to the more detailed discussion given in Section 6 of Huber (1983) as well as to the various comments on pages 72–80 following Huber (1983).

We additionally computed the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping function via (12.2.41); confer Figure 12.48. In contrast to conditional contamination neighborhoods (cf. Figure 12.32) $b^{\mathrm{as,c}}_{v,\delta}$ does not always lie below the finite-sample optimal clipping function. But, nevertheless it is negative for small and large values of $|x|$, respectively. Again, the approximation seems to work quite well for moderate values of $|x|$ and not too small sample sizes n. In particular, for $n = 10$ and $\tau = 2.576$ the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping function is already very close to the finite-sample optimal clipping function.

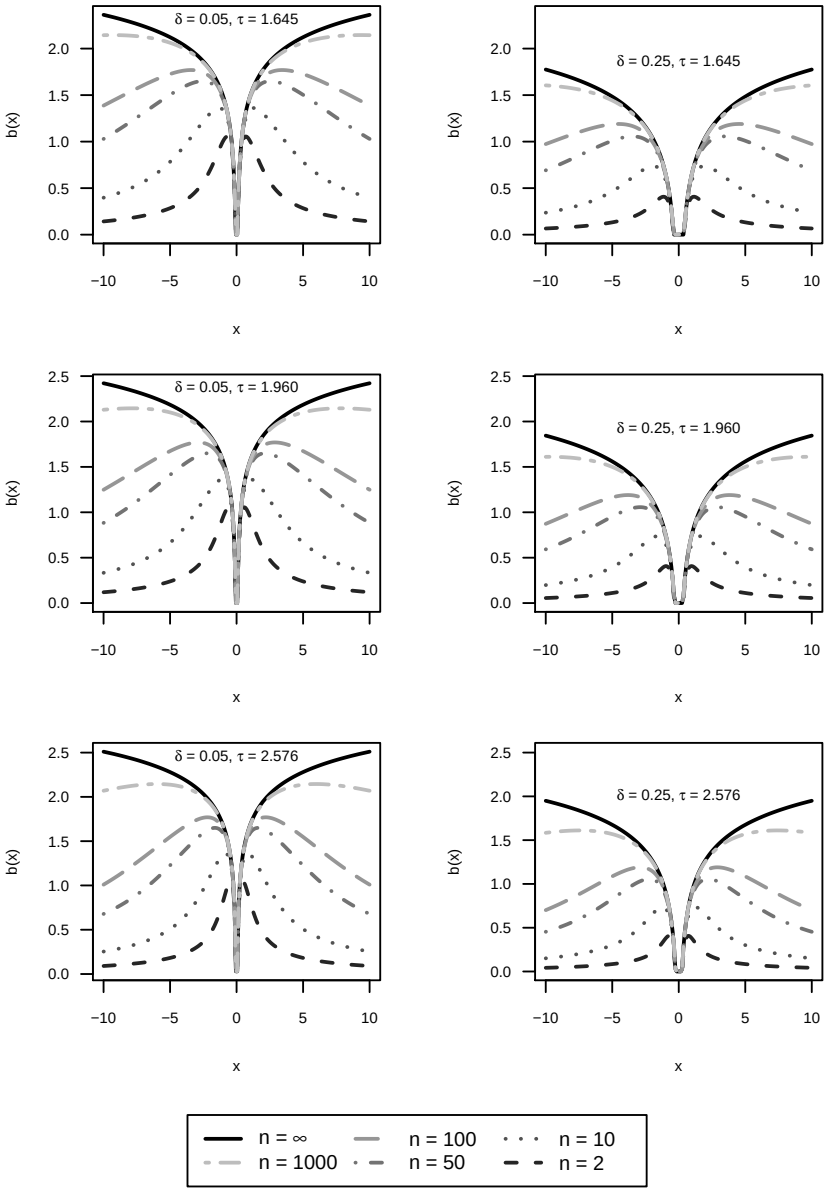

Figure 12.47: Finite-sample and asymptotic optimal clipping function for $\delta(x) \equiv 0.05, 0.25$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

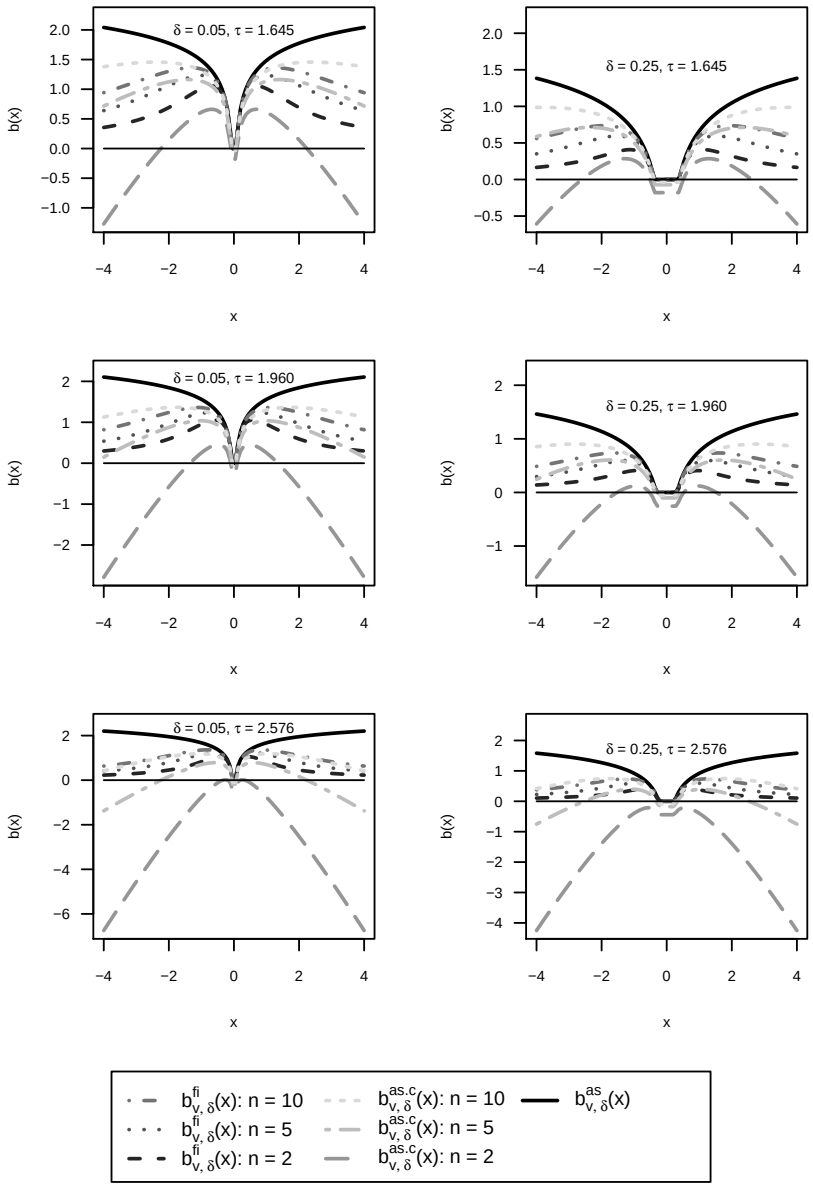

Figure 12.48: Finite-sample, $\mathrm{O}(n^{-1})$-corrected asymptotic optimal and asymptotic optimal clipping function for $\delta(x) \equiv 0.1, 0.5$ and $\tau = 1.645, 1.960, 2.576$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

12.3.4.4 Finite-Sample Risk

As in case of conditional contamination neighborhoods (c.f. Subsubsection 12.3.3.4), we consider constant contamination curves; i.e., $\delta(x) \equiv 0.05$ and $\delta(x) \equiv 0.25$, respectively. We choose $\delta(x) = \varepsilon(x)/2$ to obtain comparable results; confer equations (12.2.14) and (12.2.36). Given these contamination curves we determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}^{\text{fi}}_{v,\delta}$, the asymptotic minimax estimator $\tilde{S}^{\text{as}}_{v,\delta}$ and the estimator $S^{\text{as,c}}_{v,\delta}$ which is based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping function. Since clipping functions have to be non-negative, negative values of $b^{\text{as,c}}_{v,\delta}$ are set to 0.

As we saw in the previous subsubsection, there are clear differences between the optimal clipping functions, in particular, for large values of $|x|$. However, in case of the regressor distributions considered we have $|x| \leq 2$. Thus, not surprisingly, the differences concerning the corresponding finite-sample risks are again only small; see Figures 12.49 and 12.52. The only exceptions are the finite-sample risks of the estimator which is based on the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound for small sample sizes n.

Again, we take a closer look at the speed of convergence for the finite-sample risks. As Figures 12.50 and 12.53 indicate, the speed of it seems to be of order n^{-1} like in case of the optimal clipping functions; confer Lemma 12.2.2. As in all cases before, we consider $y = \text{Risk}^{\natural}_{\text{as}} - \text{Risk}^{\natural}_{\text{fi}}$ and apply the Box-Cox power transformation provided by the `MASS` package of Venables and Ripley (2002); i.e., we estimate lambda by means of maximum likelihood such that $y^{\text{lambda}} \approx n$. That is, lambda ≈ -1 indicates $y = \mathrm{O}(n^{-1})$. Indeed, most of the estimated values of lambda are very close to -1 which confirms our conjecture that we have a convergence of order n^{-1}; see Figures 12.35 and 12.38. Like in case of conditional contamination neighborhoods (c.f. Subsubsection 12.3.3.4), we conjecture that the relatively slow convergence in case $K = \text{Unif}([-1, 2])$ and $\delta(x) \equiv 0.25$ is caused by the fact that the finite-sample optimal clipping bound is equal to 0 for $|x| \leq \tau_n^{-1} \Phi^{-1}\big(0.5 + \delta_n(x)\big)$.

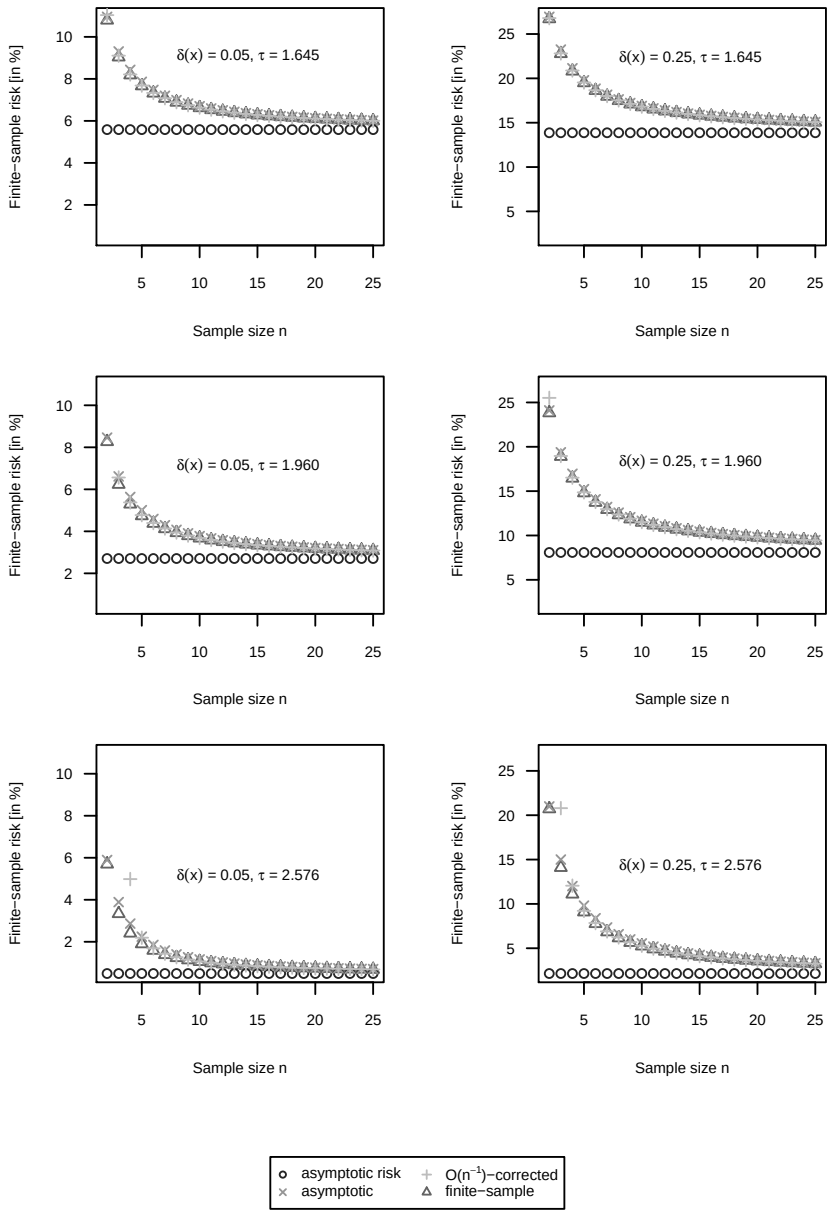

Figure 12.49: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $n \le 25$, $\delta(x) \equiv 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

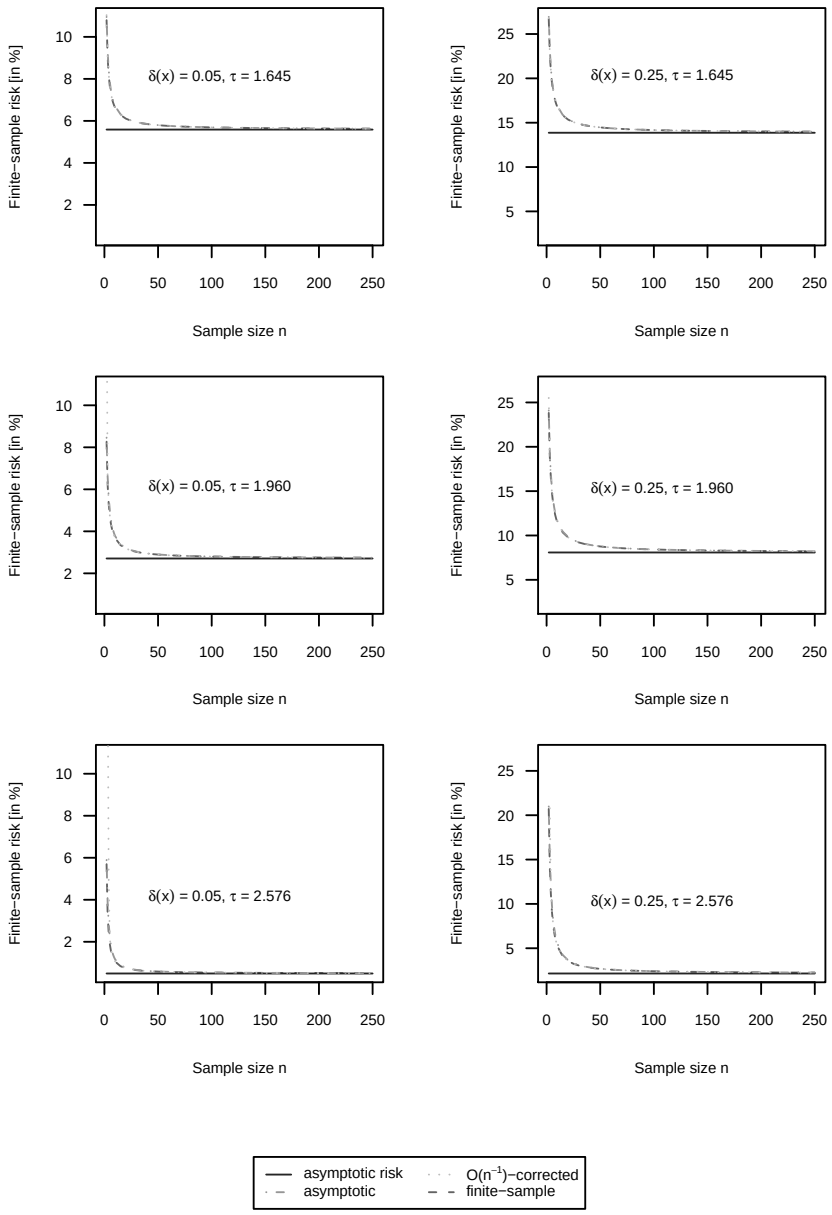

Figure 12.50: Finite-sample risk for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, increasing sample size n, $\delta(x) \equiv 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

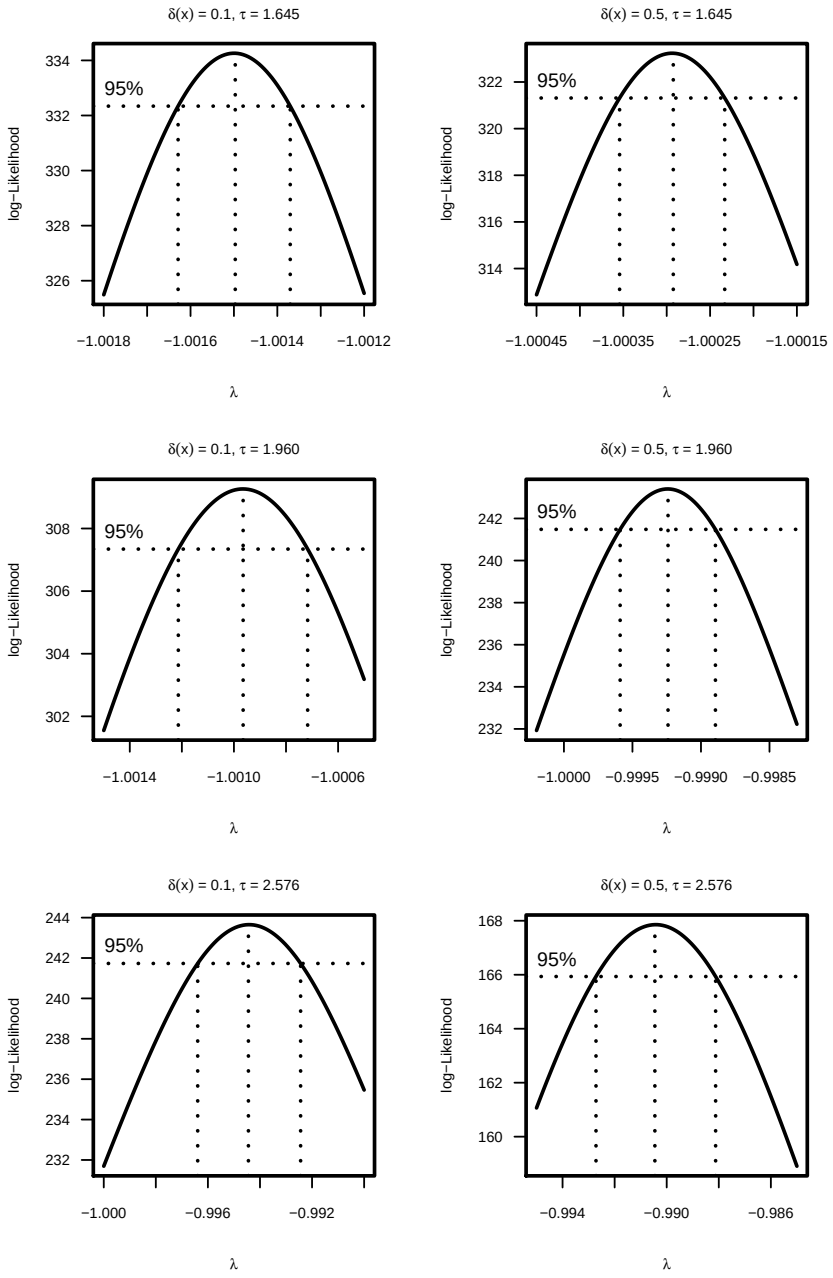

Figure 12.51: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$ and conditional total variation neighborhoods $(* = v,\, t = \delta)$.

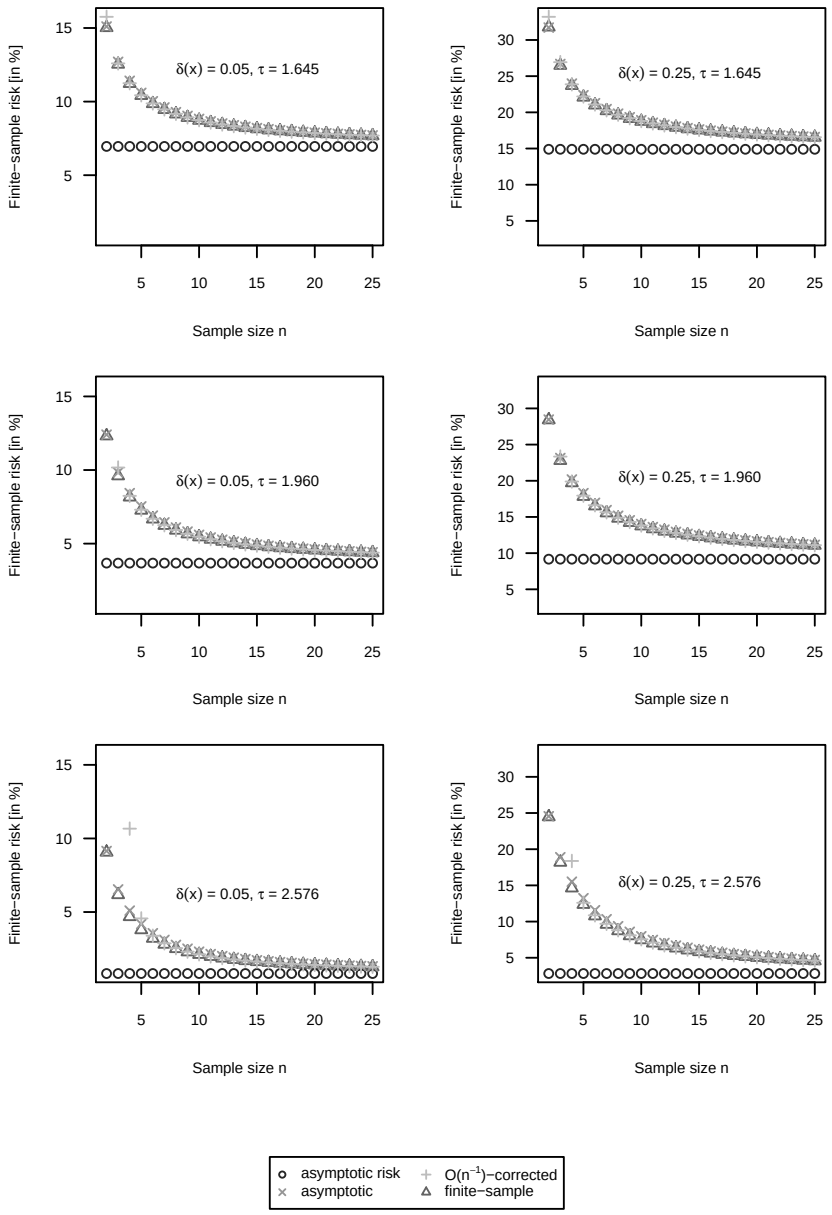

Figure 12.52: Finite-sample risk for $K = \text{Unif}([-1,2])$, $n \leq 25$, $\delta(x) \equiv 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional total variation neighborhoods $(* = v, t = \delta)$.

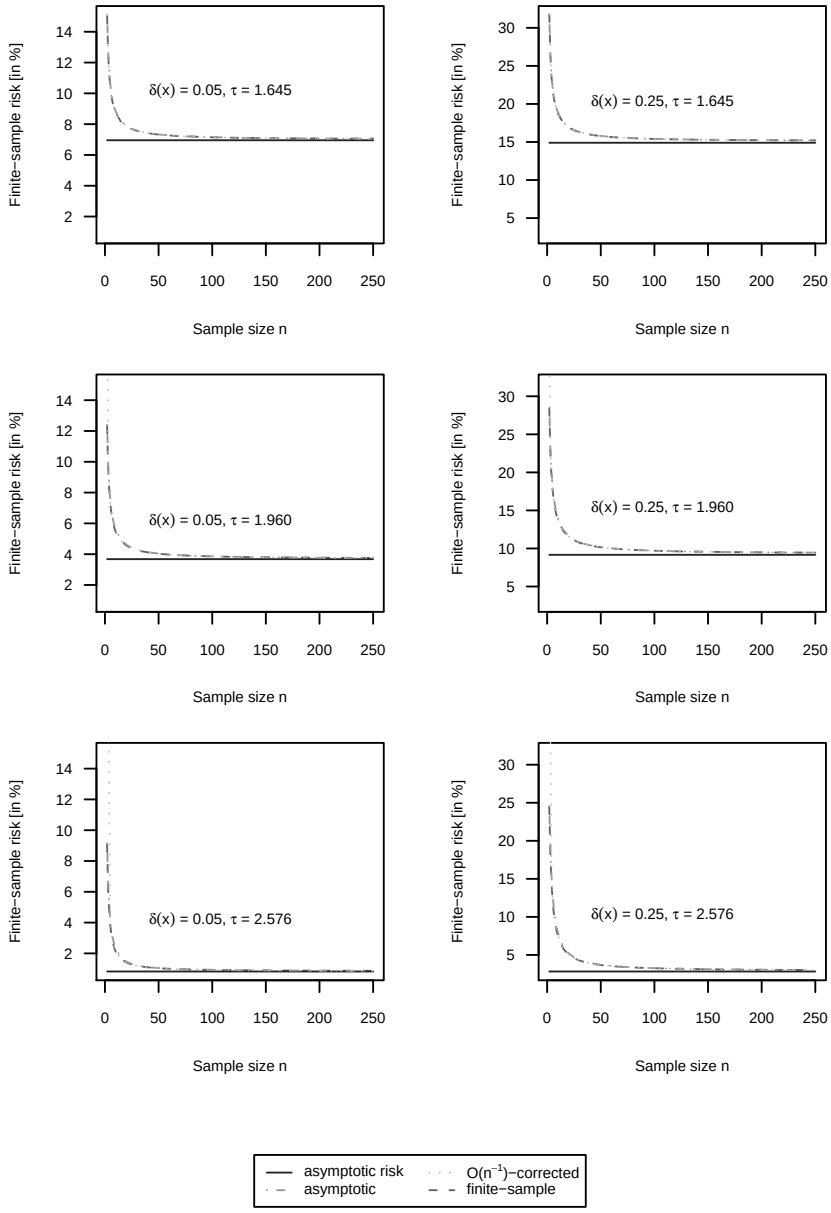

Figure 12.53: Finite-sample risk for $K = \text{Unif}([-1,2])$, increasing sample size n, $\delta(x) \equiv 0.05, 0.25$ and $\tau = 2.576, 1.960, 1.645$ in case of conditional total variation neighborhoods ($* = v$, $t = \delta$).

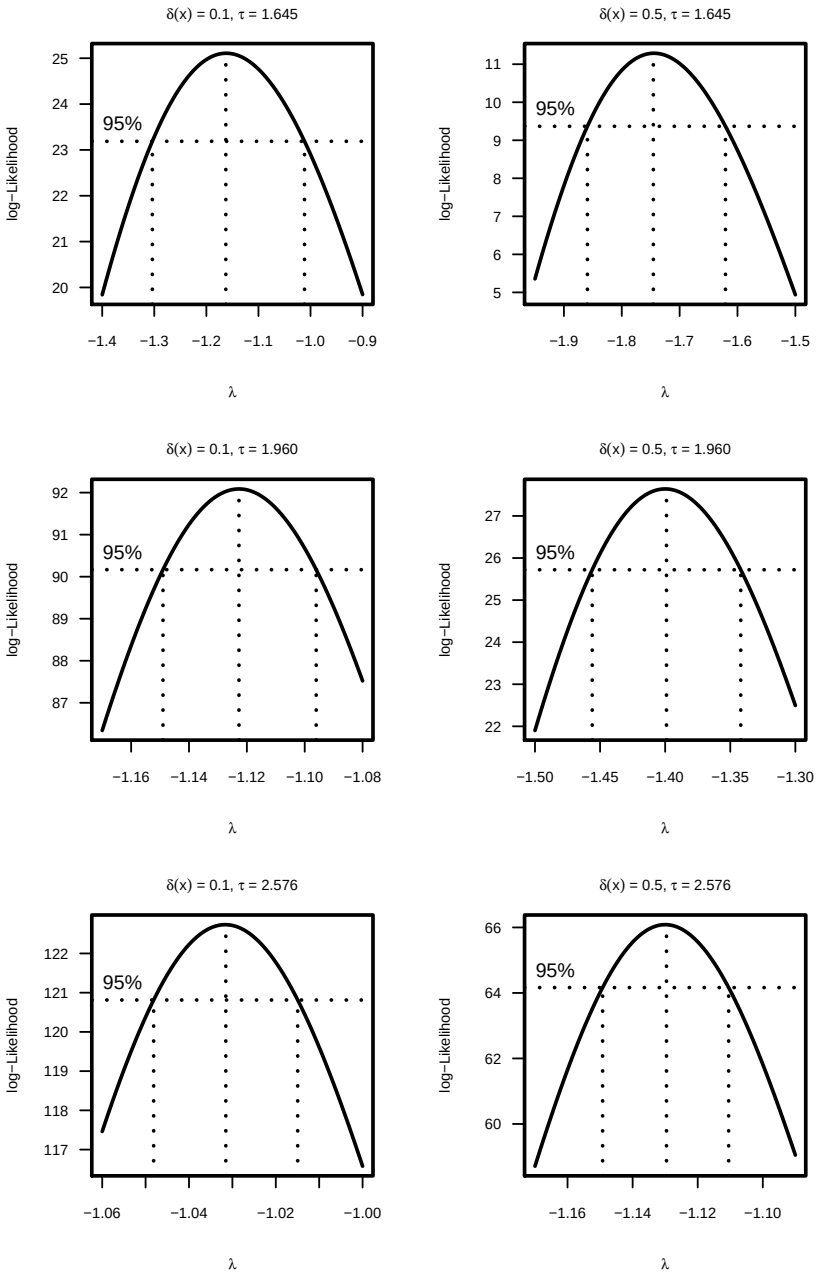

Figure 12.54: Results of the Box-Cox power transformation for the speed of convergence in case of the finite-sample risk of the finite-sample minimax estimator, $K = \text{Unif}([-1, 2])$ and conditional total variation neighborhoods ($* = v$, $t = \delta$).

12.3.4.5 Finite-Sample Distribution

As in all cases before, we use Algorithm A (cf. Subsubsection 12.2.3.2) to compute the cumulative distribution function of $\sum_{i=1}^n \tilde{\psi}_{v,\delta}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which is closest in Kolmogorov distance. By symmetry it suffices to consider only the cumulative distribution function of $\sum_{i=1}^n \tilde{\psi}_{v,\delta}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$. Moreover, we choose constant contamination curves and consider only the two "extreme" situations $\delta(x) \equiv 0.05$, $\tau = 1.645$ (see Figures 12.55 and 12.57) and $\delta(x) \equiv 0.25$, $\tau = 2.576$ (see Figures 12.56 and 12.58).

Remark 12.3.5 If $H''_{\tau_n}(dy \,|\, x)$ is absolutely continuous a.e. $H''_{\tau_n}(dx)$, then by Remark 12.2.4 (b) the distribution of $\tilde{S}^{\mathrm{fi}}_{v,\delta}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^n \tilde{\psi}_{v,\delta}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. ////

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^n \tilde{\psi}_{v,\delta}(x_i, u_i)$ under $\left(Q''_{\tau_n}\right)^n$ and of $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the `R` function `optim`; confer R Development Core Team (2005). As we see, in all cases about 5 observations are enough to get already quite close to a normal distribution (i.e., $d_\kappa < 0.02$ for $n \geq 5$) where the jumps included in the cumulative distribution functions decay exponentially in n; confer also Remark 11.3.3.
In Figures 12.57 and 12.58 the Kolmogorov distance is not monotone decreasing. Since the corresponding values are very small, we conjecture this is just a matter of numerical precision.
Somewhat surprisingly, the finite sample risk in case $\delta \equiv 0.25$ and $n = 2, 3, 4, 5$ is larger than in case of conditional contamination neighborhoods with $\varepsilon \equiv 0.5$ and $n = 2$; confer Figures 12.40, 12.42, 12.56 and 12.58. We conjecture this is caused by the fact that

$$b^{\mathrm{fi}}_{c,\varepsilon}(x) = 0 \qquad \text{for} \qquad |x| \leq 0.645, 0.561, 0.527, 0.502 \quad (n = 2, 3, 4, 5) \tag{12.3.33}$$

whereas

$$b^{\mathrm{fi}}_{v,\delta}(x) = 0 \qquad \text{for} \qquad |x| \leq 0.394, 0.390, 0.387, 0.386 \quad (n = 2, 3, 4, 5) \tag{12.3.34}$$

In particular, $b^{\mathrm{fi}}_{c,\varepsilon}(0.5) = 0$ for $n = 2, 3, 4, 5$ whereas $b^{\mathrm{fi}}_{v,\delta}(0.5) > 0$ for all n.

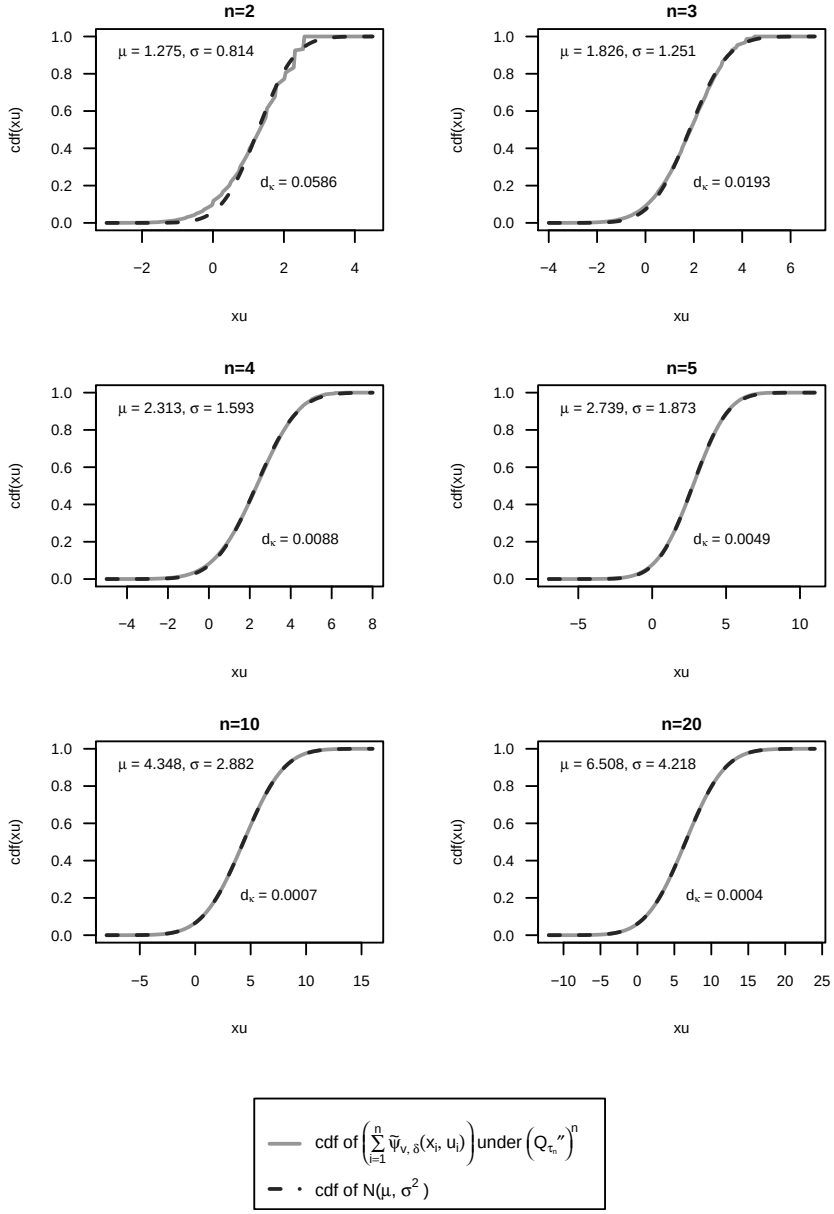

Figure 12.55: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\delta(x) \equiv 0.05$ and $\tau = 1.645$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

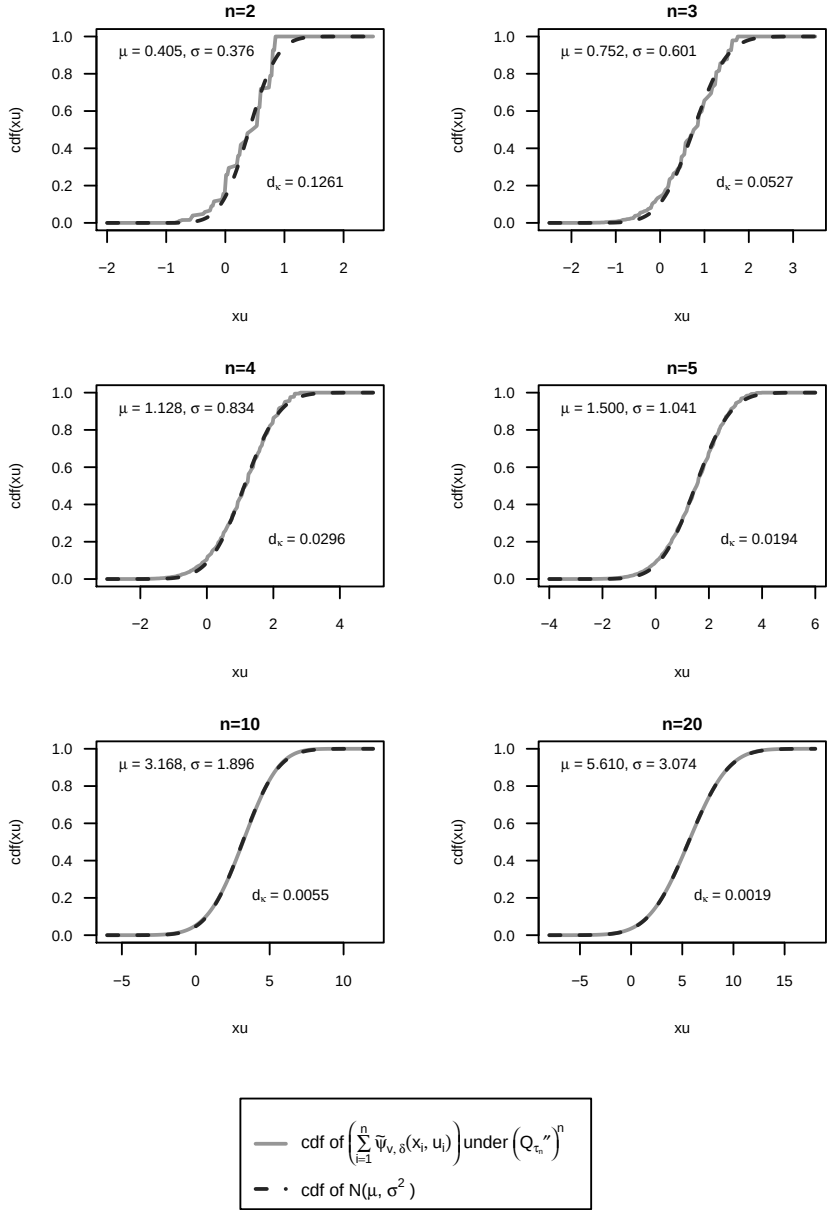

Figure 12.56: Finite-sample distributions for $K = \frac{1}{3}\big(\mathrm{I}_{\{0.5\}} + \mathrm{I}_{\{1.0\}} + \mathrm{I}_{\{1.5\}}\big)$, $\delta(x) \equiv 0.25$ and $\tau = 2.576$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

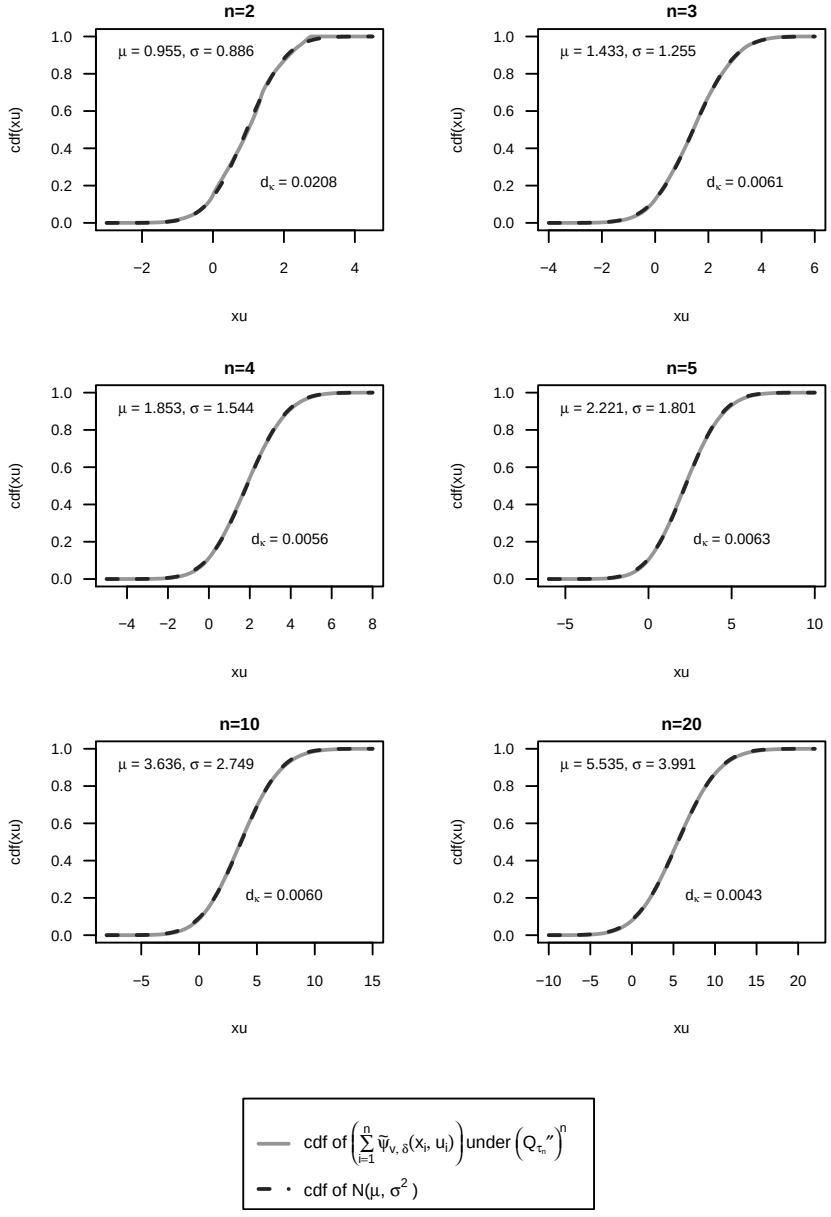

Figure 12.57: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\delta(x) \equiv 0.05$ and $\tau = 1.645$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

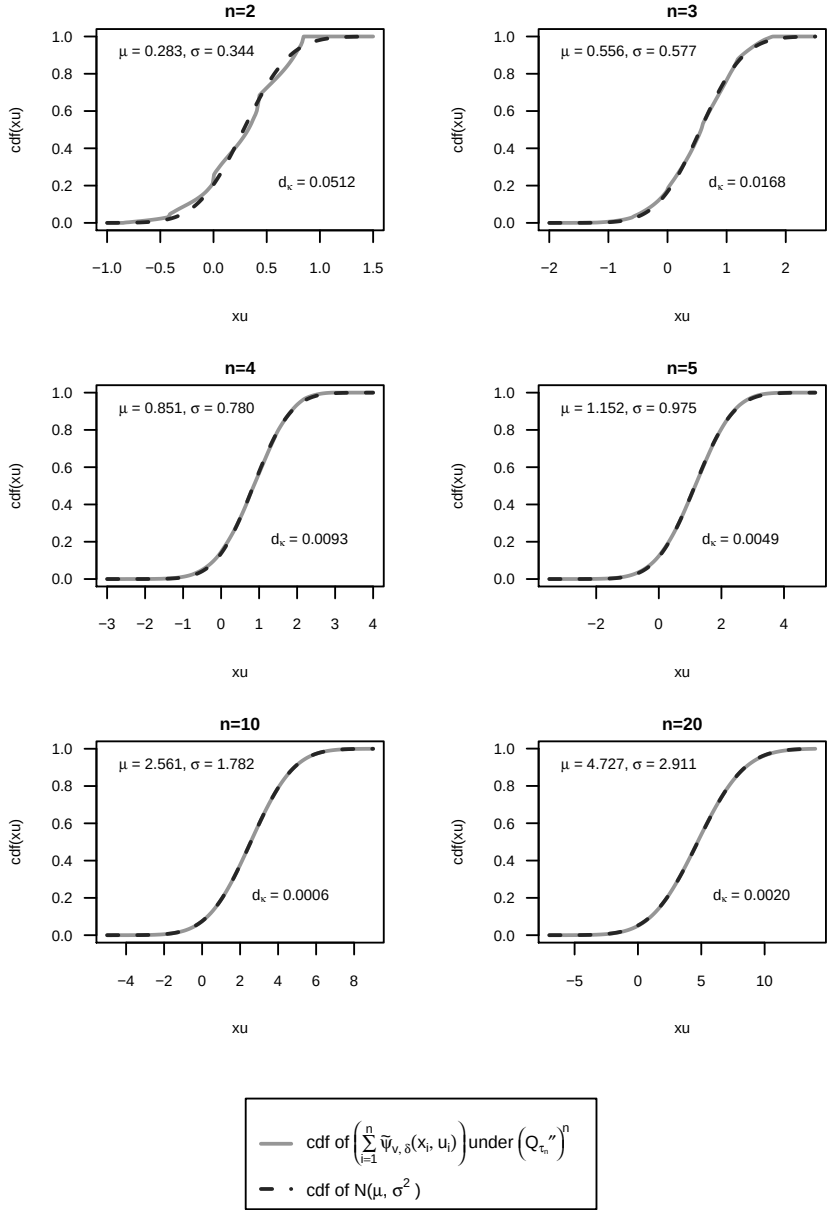

Figure 12.58: Finite-sample distributions for $K = \text{Unif}([-1, 2])$, $\delta(x) \equiv 0.25$ and $\tau = 2.576$ in case of conditional total variation neighborhoods $(* = v,\, t = \delta)$.

12.4 Implementation Using R

The computation of the optimal ICs with respect to the finite-sample as well as asymptotic under-/overshoot confidence risk can be done via our `R` package `ROptRegTS` which is part of our `R` bundle `RobASt` and is in detail described in Appendix D.4.

12.4.1 Unconditional Neighborhoods

First, we generate a normal linear regression family

```
> LM <- NormLinRegFamily(RegDistr = K)
```

where we choose some ideal univariate regressor distribution `K`; i.e., `K` is an object of class `UnivariateDistribution`. In addition to the `S4` class `InfRobRegTypeModel` we provide the `S4` class `FixRobRegTypeModel` which combines some regression-type family with a fixed neighborhood. Now, given some sample of size n, some width $\tau \in (0, \infty)$ and some neighborhood radius $r \in (0, \infty)$ we can define the following robust models

```
> Rob1 <-  InfRobRegTypeModel(center = LM,
+            neighbor = ContNeighborhood(radius = r))
> Rob2 <-  InfRobRegTypeModel(center = LM,
+            neighbor = TotalVarNeighborhood(radius = r))
```

respectively

```
> Rob3 <-  FixRobRegTypeModel(center = LM,
+            neighbor = ContNeighborhood(radius = r/sqrt(n)))
> Rob4 <-  FixRobRegTypeModel(center = LM,
+            neighbor = TotalVarNeighborhood(radius = r/sqrt(n)))
```

where we use $r/\sqrt{n}$ in case of the robust model with fixed unconditional neighborhoods to get comparable results; confer Remark 10.1.4 (b). To compute the asymptotic optimal IC, we can proceed as follows

```
> IC1 <- optIC(model = Rob1, risk = asUnOvShoot(width = τ))
> IC2 <- optIC(model = Rob2, risk = asUnOvShoot(width = τ))
```

Remark 12.4.1 **(a)** In case of the asymptotic under-/overshoot confidence risk, the algorithm works for any one-dimensional ideal error distribution with finite Fisher information of location.

(b) Since the solution in case of unconditional contamination neighborhoods is of total variation form (i.e., no centering constant but possibly asymmetric clipping), `IC1` is of class `TotalVarIC`. ////

The finite-sample optimal IC can be determined via

```
> IC3 <- optIC(model = Rob3,
+              risk = fiUnOvShoot(width = τ/sqrt(n)),
+              sampleSize = n)
> IC4 <- optIC(model = Rob4,
+              risk = fiUnOvShoot(width = τ/sqrt(n)),
+              sampleSize = n)
```

We use width $\tau/\sqrt{n}$ to get comparable results. In case of the finite-sample under-/overshoot confidence risk one can also specify whether Algorithm A or Algorithm B (cf. Subsubsection 12.1.3.2) should be used in the call to `optIC` by setting the additional parameter `Algo` to `"A"` or `"B"`, respectively. By default the faster Algorithm A is called. In case of the finite-sample under-/overshoot confidence risk it is also possible to distinguish between contamination "left" or "right" of some boundary region by setting the additional parameter `cont` to `"left"` or `"right"`, respectively. More details are given in Subsubsection 12.1.3.1.

Remark 12.4.2 **(a)** In contrast to the asymptotic risk, the solution in case of the finite-sample risk is only implemented for normal ideal error distribution.

(b) As in case of the asymptotic risk, the solution in case of unconditional contamination neighborhoods ($* = c$, $t = 0$) is of total variation form; i.e., `IC3` is of class `TotalVarIC`. ////

To compute the $\mathrm{O}(n^{-1/2})$-corrected optimal IC ($* = c$, $t = 0$), we start with `IC1` and modify the slots `clipUp`, `clipLo` and `stand` using equations (12.1.16) and (12.1.7). This can be done as follows

```
> IC5 <- IC1
> clipUp1 <- clipUp(IC1)/as.vector(stand(IC1))
> h1 <- E(K, function(x, b)pnorm(-b/abs(x)), b = clipUp1)
> clipUp5 <- r*(r + clipUp1*τ)/(sqrt(n)*2*τ*h1)
> clipUp5 <- max(0, clipUp1 - clipUp5)
> h2 <- E(K, A.fun, b = clipUp5)
> stand5 <- 1/(2*h2 - E(K, function(x)x^2)
> clipUp(IC5) <- stand5*clipUp5
> clipLo(IC5) <- -clipUp(IC5)
> stand(IC5) <- as.matrix(stand5)
```

where `A.fun` is defined as

```
> A.fun <- function(x, b){
+     if(x == 0) return(0)
+     return(x^2*pnorm(b/abs(x)))
+ }
```

Similarly, we can determine the $\mathrm{O}(n^{-1})$-corrected optimal IC ($* = v$, $t = 0$) via

```
> IC6 <- IC2
> clipUp2 <- clipUp(IC2)/as.vector(stand(IC2))
```

```
> h1 <- E(K, function(x, b)abs(x)^3*dnorm(b/abs(x)), b = clipUp1)
> h2 <- E(K, function(x, b)pnorm(-b/abs(x)), b = clipUp1)
> clipUp6 <- max(0, clipUp1 - τ*(2*clipUp2^2*r + τ*h1)/(6*n*h2))
> h3 <- E(K, A.fun, b = clipUp6)
> stand6 <- 1/(2*h3 - E(K, function(x)x^2)
> clipUp(IC6) <- stand6*clip6
> clipLo(IC6) <- -clipUp(IC6)
> stand(IC6) <- as.matrix(stand6)
```

by using equations (12.1.50) and (12.1.41).

Remark 12.4.3 After the installation of our `R` bundle `RobASt` one can find the `R` script `UnderOverShootRisk.R`, which contains some examples for the computation of the optimal ICs, in the directory "…/`RHome`/`library`/`ROptRegTS`/`scripts`/" where `RHome` stands for the local home directory of `R`. ////

12.4.2 Conditional Neighborhoods

As in case of unconditional neighborhoods we first generate a normal linear regression family

```
> LM <- NormLinRegFamily(RegDistr = K)
```

where we choose some ideal univariate regressor distribution `K`. Then, given some sample of size n, some width $\tau \in (0,\infty)$ and some radius curve $r\colon \mathbb{R} \to (0,\infty)$ we can instantiate the following robust models

```
> Rob1c <-  InfRobRegTypeModel(center = LM,
+             neighbor = CondContNeighborhood(radiusCurve = r))
> Rob2c <-  InfRobRegTypeModel(center = LM,
+             neighbor=CondTotalVarNeighborhood(radiusCurve = r))
```

respectively

```
> Rob3c <-  FixRobRegTypeModel(center = LM,
+           neighbor = CondContNeighborhood(radiusCurve = r_n))
> Rob4c <-  FixRobRegTypeModel(center = LM,
+           neighbor=CondTotalVarNeighborhood(radiusCurve = r_n))
```

where $r_n(x) = r(x)/\sqrt{n}$. We use r_n in case of the robust model with fixed conditional neighborhoods to get comparable results; confer Remark 10.1.4 (b). To determine the asymptotic optimal IC, we can proceed as follows

```
> IC1c <- optIC(model = Rob1c, risk = asUnOvShoot(width = τ))
> IC2c <- optIC(model = Rob2c, risk = asUnOvShoot(width = τ))
```

Remark 12.4.4 (a) In case of the asymptotic under-/overshoot confidence risk, the algorithm works for any one-dimensional ideal error distribution with finite Fisher information of location.

(b) Since the solution in case of conditional contamination neighborhoods ($* = c$, $t = \varepsilon$) is of total variation form (i.e., no centering function but possibly asymmetric clipping), `IC1c` is of class `CondTotalVarIC`. ////

The finite-sample optimal IC can be computed via

```
> IC3c <- optIC(model = Rob3c,
+               risk = fiUnOvShoot(width = τ/sqrt(n)),
+               sampleSize = n)
> IC4c <- optIC(model = Rob4c,
+               risk = fiUnOvShoot(width = τ/sqrt(n)),
+               sampleSize = n)
```

where we use width $\tau/\sqrt{n}$ to get comparable results.

Remark 12.4.5 **(a)** In contrast to the asymptotic risk, the solution in case of the finite-sample risk is only implemented for normal ideal error distribution.

(b) As in case of the asymptotic risk, the solution in case of conditional contamination neighborhoods is of total variation form; i.e., `IC3` is of class `CondTotalVarIC`.

(c) In case of the finite-sample under-/overshoot confidence risk and conditional neighborhoods we only implemented Algorithm A; confer Subsubsection 12.2.3.2. Again it is possible to distinguish between contamination left or right of some bound by setting the additional parameter `cont` to `"left"` or `"right"`, respectively. More details are given in Subsubsection 12.2.3.1. ////

To determine the $\mathrm{O}(n^{-1/2})$-corrected optimal IC ($* = c$, $t = \varepsilon$), we start with `IC1c` and modify the slots `clipUp`, `clipLo` and `stand` using equations (12.2.19) and (12.2.12). This can be done by using the following `R` code

```
> IC5c <- IC1c
> clipUp <- function(x){
> +  bf <- bfun; rf <- rfun
> +  b <- bf(x)/(A*abs(x))
> +  pmax(0, b - rf(x)*(rf(x) + b*τ)/(√n *2*τ*abs(x)*pnorm(-b)))
> +}
> body(clipUp) <- substitute( {bf <- bfun; rf <- rfun
> +      b <- bf(x)/(A*abs(x))
> +      pmax(0, b - rf(x)*(rf(x) + b*τ)/
> +        (√n *2*τ*abs(x)*pnorm(-b))) },
> +      list(bfun = clipUp(IC1c)@Map[[1]], rfun = r_n ,
> +        A = as.vector(stand(IC1c)), τ = τ))
> h2 <- E(K, A.fun, b = clipUp)
> stand5 <- 1/(2*h2 - E(K, function(x)x^2)
> clipUp5 <- function(x){ bf <- bfun; A*b(x)*abs(x) }
> body(clipUp5) <- substitute({ bf <- bfun; A*b(x)*abs(x) },
> +        list(bfun = clipUp, A = stand5))
> clipUp(IC5c) <- RealRandVariable(list(clipUp5), Domain=Reals())
> clipLo5 <- function(x){ bf <- bfun; -A*b(x)*abs(x) }
```

```
> body(clipLo5) <- substitute({ bf <- bfun; -A*b(x)*abs(x) },
> +        list(bfun = clipUp, A = stand5))
> clipLo(IC5c) <- RealRandVariable(list(clipLo5), Domain=Reals())
> stand(IC5c) <- as.matrix(stand5)
```

where `A.fun` is defined as

```
> A.fun <- function(x, b){ x^2*pnorm(b(x)) }
```

Similarly, we can determine the $\mathrm{O}(n^{-1})$-corrected optimal IC $(* = v)$ via

```
> IC6c <- IC2c
> clipUp <- function(x){
> +  bf <- bfun; rf <- rfun
> +  b <- bf(x)/(A*abs(x))
> +  pmax(0, b - τ*(2*b^2*rf(x) + τ*dnorm(b))/(6*n*pnorm(-b)))
> +}
> body(clipUp) <- substitute( {bf <- bfun; rf <- rfun
> +      b <- bf(x)/(A*abs(x))
> +      pmax(0, b - τ*(2*b^2*rf(x) + τ*dnorm(b))/
> +        (6*n*pnorm(-b))) },
> +      list(bfun = clipUp(IC2c)@Map[[1]], rfun = r_n ,
> +        A = as.vector(stand(IC2c)), τ = τ))
> h2 <- E(K, A.fun, b = clipUp)
> stand6 <- 1/(2*h2 - E(K, function(x)x^2)
> clipUp6 <- function(x){ bf <- bfun; A*b(x)*abs(x) }
> body(clipUp6) <- substitute({ bf <- bfun; A*b(x)*abs(x) },
> +        list(bfun = clipUp, A = stand6))
> clipUp(IC6c) <- RealRandVariable(list(clipUp6), Domain=Reals())
> clipLo6 <- function(x){ bf <- bfun; -A*b(x)*abs(x) }
> body(clipLo6) <- substitute({ bf <- bfun; -A*b(x)*abs(x) },
> +        list(bfun = clipUp, A = stand6))
> clipLo(IC6c) <- RealRandVariable(list(clipLo6), Domain=Reals())
> stand(IC6c) <- as.matrix(stand6)
```

by using equations (12.2.41) and (12.2.34).

Remark 12.4.6 After the installation of our `R` bundle `RobASt` one can find the `R` script `UnderOverShootRiskCond.R`, which contains some examples for the computation of the optimal ICs, under "`.../RHome/library/ROptRegTS/scripts/`" where `RHome` stands for the local home directory of `R`. ////

Outlook

During the work on this thesis many new questions arose, respectively "old" questions remained open. We will now give a subjective and incomplete task list (intending no order) which should be worked through in the near future:

Estimator Construction

(1) Find a principle for thinning out full infinitesimal neighborhoods. This should be statistically meaningful and should achieve certain mathematical advantages. The modification given on page xxxv excludes all samples which contain more than 50% contamination and enables the studies on higher order asymptotics in Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c) and Ruckdeschel (2005e). Similar modifications might provide a way to solve the construction problem for optimal ICs which are unbounded; confer Chapter 6 of Rieder (1994) and Section 7 of Rieder (2003).
(2) Search and implement appropriate initial estimators for the one-step construction in further models; confer Section 6.4 of Rieder (1994) and Section 2.3.

Lower Case

(3) Derive the general minimum variance solution to minimal bias bound; confer Chapters 5 and 7 of Rieder (1994). In the special case of linear regression with normal errors and finite design a first result was obtained by Müller (1987) and Kurotschka and Müller (1992). The general result is on the way; confer Ruckdeschel (2005c).
(4) Deduce gap conditions from multivariate discrete models. These conditions would imply that the MSE solution achieves the minimum bias already for finite radii; confer Subsection 2.1.2.

Convergence of Robust Models

(5) Generalize the concept of convergence of robust models, which so far has been solely based on the comparision of MSE solutions, and link it up with Le Cam's abstract convergence of experiments; confer Section 2.4 and Le Cam (1986).
(6) Analogously to Chapters 3 and 4, apply the notion to obtain approximative optimally robust estimators for multivariate models.

Numerical Issues

Convergence of Algorithms

(7) Prove the convergence of the fixed point algorithms used to compute optimally robust influence curves; confer Algorithm D.3.1.

Multivariate Distributions

(8) Extend our `R` package `distr` to multivariate distributions; confer Ruckdeschel et al. (2005). A first step has already been made in our `R` package `distrEx`; confer Appendix D.1.
(9) Broaden the scope of our convolution algorithm to multivariate, at least two- or three-dimensional distributions; confer Appendix C.

`R` Bundle `RobASt`

(10) Add further `S4` methods for the computation of optimally robust influence curves in case of smoothly parameterized ideal models to our `R` package `ROptEst`; confer Appendix D.3; for instance, for parametric models which are based on multivariate distributions.
(11) Implement `S4` methods for the computation of optimally robust influence curves in case of time series models. Several of these models have a certain regression structure (e.g., ARMA (p, q), ARCH (p)) and hence, fit well in the framework of our `R` package `ROptRegTS`; confer Drost et al. (1997), Section 9.3 and Appendix D.4.
(12) Enlarge our `R` package `ROptRegTS` (cf. Appendix D.4) to further regression-type models like generalized linear models; confer Rieder (1995).
(13) Add `R` packages for robust testing; confer for instance Huber (1965), Huber and Strassen (1973), Rieder (1977), Rieder (1978), Rieder (1981b) or Section 5.4 of Rieder (1994).
(14) Include diagnostic tools which enable the user to analyse robust estimation and testing results.

Appendix

Appendix A

Asymptotic Theory of Robustness for Regression Type Models

This chapter of the appendix is based on Rieder (1994) (particularly Chapter 7) and Rieder (2003). We give a short summary of results which hold for L_2 differentiable, structured, parametric models, like linear regression and several time series models. All models whose L_2 derivative has the following product form

$$\Lambda_\theta = H_\theta \Lambda_f \tag{A.0.1}$$

The linear regression model is in detail treated in Chapter 7 of Rieder (1994) and the derived results canonically extend to linear regression and scale; confer Subsubsection 7.2.1.2. Hence, we restrict our presentation to time series models and explain how the results of Chapter 7 in Rieder (1994) also apply to certain of these models. By identifying H_θ with the regressor X and Λ_f with the location scores Λ_f^{loc} of the error, the optimality results given in Theorems A.2.3, A.2.5 and A.2.7 are identical to the corresponding results in case of linear regression.

A.1 Introduction

We assume a (strictly) stationary and ergodic stochastic process $(X_t)_{t\in\mathbb{Z}}$ where $H_\theta = H_{\theta,j} = H_\theta(\bar{X}_{j-1})$ is a $\mathbb{B}^{\mathbb{Z}}$ measurable transformation of the extended past $\bar{X}_{j-1} = (X_{j-1}, X_{j-2}, \ldots, X_1, X_0, \ldots)$. Thus, H_θ is also (strictly) stationary and ergodic. For one-sided processes this is shown in Propositions 6.6 and 6.31 of Breiman (1968) and one can easily generalize this to two-sided processes; confer for instance Lemma A.3.2 of Ruckdeschel (2001) which was part of a lecture on time series models hold by H. Rieder in the winter term 2001/02. Moreover, $\Lambda_f = \Lambda_f(V_j)$ corresponds to the location $(\Lambda_f^{\mathrm{loc}}(V_j))$, respectively scale $(\Lambda_f^{\mathrm{sc}}(V_j))$ scores of the innovations at time j where we assume $V_j \overset{\text{i.i.d.}}{\sim} F$ and stochastically independent

of the extended past $\bar{X}_{j-1}$. By this independence, the product form (A.0.1) of the scores Λ_θ implies the following product form of the corresponding Fisher information

$$\mathcal{I}_\theta = \mathcal{I}_f \mathcal{K}_{\theta,F} \tag{A.1.1}$$

where $\mathcal{I}_f$ is the Fisher information of location, respectively scale of the ideal innovation distribution F. Furthermore, $\mathcal{K}_{\theta,F} = \mathrm{E}_\theta\, H_\theta H_\theta^\tau$ is the covariance matrix of the process laws $\mathcal{L}_\theta\{(X_t)_{t\in\mathbb{Z}}\}$.

Remark A.1.1 **(a)** Generally speaking, one has to replace the assumption independent and identically distributed in the context of linear regression by (strictly) stationary and ergodic in case of time series models. For the definition of L_2 differentiability in case of dependent variables we refer to Subsection 10.2.3 of Ruckdeschel (2001).

(b) L_2 differentiable time series models with L_2 derivative Λ_θ of product form (A.0.1) are for instance ARMA(p,q) (cf. Subsection 9.3.1) and ARCH(p) (cf. Subsection 9.3.2) models. More examples are given in Rieder (2003), pp 9. ////

Analogously to Definition 7.4.1 of Rieder (1994), we now define ICs for times series models; confer also Section 3 of Rieder (2003).

Definition A.1.2 *Suppose the time series model $\mathcal{P}$ is L_2 differentiable at θ, and assume some matrix $D \in \mathbb{R}^{p\times k}$ of full rank $p \le k$. Let $\alpha = 2, \infty$, respectively. Then, the set $\Psi_{2\bullet}^D(\theta)$ of all* square integrable, *and the subset $\Psi_{\infty\bullet}^D(\theta)$ of all* bounded, conditionally centered (partial) ICs *at P_θ, respectively, are*

$$\Psi_{\alpha\bullet}^D(\theta) = \left\{\psi_\theta \in L_2^k(P_\theta) \mid \mathrm{E}_\theta[\psi_\theta(\bar{x}_1)|\bar{x}_0] = 0,\ \mathrm{E}_\theta\, \psi_{\theta,1}\Lambda_{\theta,1}^\tau = D\right\} \tag{A.1.2}$$

where $\psi_{\theta,1} := \psi_\theta(\bar{x}_1)$ and $\Lambda_{\theta,1} := \Lambda_\theta(\bar{x}_1)$.

Remark A.1.3 The martingale difference sequence $\psi_{\theta,j} = \psi_\theta(\bar{X}_j)$ measures the influence of X_j given the past $\bar{X}_{j-1}$. Since $(X_t)_{t\in\mathbb{Z}}$ is (strictly) stationary and ergodic and $\psi_\theta \in \Psi_2^D(\theta)$ is $\mathbb{B}^\mathbb{Z}$ measurable, $\psi_{\theta,j}$ is also (strictly) stationary and ergodic. ////

Based on these ICs, we now define asymptotically linear estimators.

Definition A.1.4 *An asymptotic estimator for $D\theta \in \mathbb{R}^p$*

$$S = (S_n) \qquad S_n : (\Omega, \mathcal{A}_n) \to (\mathbb{R}^p, \mathbb{B}^p) \tag{A.1.3}$$

is called asymptotically linear at $P_{D\theta}$ *if there is a conditionally centered IC $\psi_\theta \in \Psi_{2\bullet}^D(\theta)$ such that*

$$\sqrt{n}\,(S_n - D\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \psi_\theta(\bar{x}_j) + o_{P_\theta^{(n)}}(n^0) \tag{A.1.4}$$

We call ψ_θ the IC, *of S at $P_{D\theta}$.*

Remark A.1.5 (a) $(\mathcal{A}_n)_{n\in\mathbb{N}}$ in the definition of S_n is an increasing filtration; i.e.,

$$\{\emptyset,\Omega\} =: \mathcal{A}_{-1} \subset \mathcal{A}_0 \subset \mathcal{A}_1 \subset \ldots \subset \mathcal{A} \tag{A.1.5}$$

where $(\Omega,\mathcal{A})$ is some measurable space.

(b) The joint distribution of the past $\bar{X}_n$ is given by

$$P_\theta^{(n)}(dx_n,\ldots,dx_1,d\bar{x}_0) = \prod_{j=1}^{n} P_\theta^{(n,j)}(dx_j\,|\,\bar{x}_{j-1})P_\theta^{(n,0)}(d\bar{x}_0) \tag{A.1.6}$$

with initial distribution $P_\theta^{(n,0)}(d\bar{x}_0)$ and transition probabilities $P_\theta^{(n,j)}(dx_j\,|\,\bar{x}_{j-1})$.

(c) If S is asymptotically linear at $P_{D\theta}$ with IC $\psi_\theta \in \Psi_{2\bullet}^D(\theta)$, then

$$\sqrt{n}\,(S_n - D\theta)(P_\theta^{(n)}) \xrightarrow{\mathrm{w}} \mathcal{N}\,(0,\mathrm{Cov}_\theta(\psi_{\theta,1})) \tag{A.1.7}$$

because of $\psi_\theta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta[\psi_{\theta,1}\,|\bar{x}_0] = 0$, and the central limit theorem for (strictly) stationary and ergodic martingale differences; confer McLeish (1974). The third condition $\mathrm{E}_\theta\,\psi_{\theta,1}\Lambda_{\theta,1}^\tau = D$, is equivalent to the locally uniform extension of this asymptotic normality; confer also Remark 1.1.6 (b).

(d) The Cramér-Rao bound reads

$$\mathrm{Cov}_\theta(\psi_{\theta,1}) \succeq D\mathcal{I}_\theta^{-1}D^\tau = \mathrm{Cov}_\theta(\psi_{h,\theta,1}) \tag{A.1.8}$$

and is achieved iff $\psi_{\theta,1} = \psi_{h,\theta,1} = D\mathcal{I}_\theta^{-1}\Lambda_{\theta,1}$; confer also Proposition 1.1.7. ////

Furthermore, in case of time series models, we consider neighborhoods $\mathcal{U}_{*,\varepsilon}(\theta,r)$ $(*=c,v)$ of transition probabilities which consist of all probability measures Q such that

$$Q^{(n,j)}(dx_j\,|\,\bar{x}_{j-1}) \in B_*\big(P_\theta^{(n,j)}(dx_j\,|\,\bar{x}_{j-1}), r\varepsilon^{(n,j)}(\bar{x}_{j-1})\big) \quad \text{a.e. } Q^{(n,0)}(d\bar{x}_0) \tag{A.1.9}$$

where

$$Q^{(n,0)}(d\bar{x}_0) \in B_*\big(P_\theta^{(n,0)}(d\bar{x}_0), r\varepsilon^{(n,0)}\big) \tag{A.1.10}$$

and $\varepsilon^{(n,j)}\colon (\mathbb{R}^{\mathbb{Z}},\mathbb{B}^{\mathbb{Z}}) \to (\bar{\mathbb{R}},\bar{\mathbb{B}})$ with $\varepsilon^{(n,0)}$ constant; confer Section 4 of Rieder (2003). For the definition of B_* $(*=c,v)$ see Section 1.2.

Remark A.1.6 In analogy to Subsection 7.2.2 of Rieder (1994), we consider average transition neighborhoods $\mathcal{U}_{*,\alpha}(\theta,r)$ of exponent $\alpha = 1,2$ whose contamination curves ε satisfy $\mathrm{E}\,\varepsilon \le 1$ $(\alpha=1)$, respectively $\mathrm{E}\,\varepsilon^2 \le 1$ $(\alpha=2)$. ////

The neighborhood submodels in case of time series models consist of the following simple perturbations $Q_n(q,r)$ of transition probabilities

$$Q_n^{(n,j)}(dx_j\,|\,\bar{x}_{j-1}) = \big(1 + r/\sqrt{n}\,q(\bar{x}_j)\big)P_\theta^{(n,j)}(dx_j\,|\,\bar{x}_{j-1}) \tag{A.1.11}$$

where $q \in L_\infty(P_\theta)$, $\mathrm{E}_\theta[q_1\,|\,\bar{x}_0] = 0$ and

$$q_1 = q(\bar{x}_1) \ge -\varepsilon(\bar{x}_0) \quad (*=c) \qquad\qquad \mathrm{E}_\theta[|q_1|\,|\,\bar{x}_0] \le 2\varepsilon(\bar{x}_0) \quad (*=v) \tag{A.1.12}$$

with radius curve $\varepsilon^{(n,j)} = \varepsilon$ $(j=1,\ldots,n)$. In addition, $Q_n^{(n,0)} = P_\theta^{(n,0)}$; confer Section 4 of Rieder (2003).

Remark A.1.7 (a) In case $\alpha = 1, 2$, the corresponding submodels include all submodels whose radius curves additionally fulfill $\mathrm{E}\,\varepsilon^\alpha \le 1$. This is in analogy to Subsection 7.2.4 of Rieder (1994).

(b) The specified submodels can be regarded as a generalized form of the IO (innovational outliers) model as they may depend on the past and they also cover the AO (additive outliers) and SO (substitutive outliers) model; confer Section 4 of Rieder (2003).

(c) Let S be an estimator that is asymptotically linear at $P_{D\theta}$ with IC $\psi_\theta \in \Psi_{2\bullet}^D(\theta)$. Then, for given simple perturbations Q_n,

$$\sqrt{n}\,(S_n - D\theta)\big(Q_n^{(n)}\big) \dashrightarrow \mathcal{N}_p\big(r\,\mathrm{E}_\theta\,\psi_{\theta,1}q_1,\,\mathrm{Cov}_\theta(\psi_{\theta,1})\big) \tag{A.1.13}$$

confer Section 4 of Rieder (2003). This corresponds to Proposition 1.2.5. ////

A.2 Optimally Robust Influence Curves

To simplify the notation we omit the index 1 and denote $\mathrm{E}_\theta[\psi_{\theta,1}\,|\,\bar{x}_0]$ by $\mathrm{E}_{\theta\bullet}\,\psi_\theta$. We also drop the fixed parameter θ and define $K = \mathcal{L}(H)$.
In this section we specify the solutions to the following Hampel type problems for $* = c$ and $t = \alpha = 1, 2$, respectively $* = v$ and $t = \alpha = 1$

$$E|\eta|^2 = \min\,! \qquad \eta \in \Psi_{2\bullet}^D,\ \omega_{*,t}(\eta) \le b \tag{A.2.1}$$

where the bias terms $\omega_{*,t}(\eta)$ read

$$\omega_{c,1} = \sup_P |\eta| \tag{A.2.2}$$

and

$$\omega_{c,2}^2 = \mathrm{E}\sup_{x_1} |\eta(x_1, \bar{x}_0)|^2 \tag{A.2.3}$$

$$\omega_{v,1} = \sup_K \sup_{|e|=1} (\sup_\bullet e^\tau\eta - \inf_\bullet e^\tau\eta) \tag{A.2.4}$$

confer Section 5 of Rieder (2003).

Remark A.2.1 Analogously to Lemma 7.3.4 of Rieder (1994), the terms $\omega_{*,t}$ ($* = c, v$ and $t = \alpha$) are positively homogeneous, subadditive, hence convex on $L_1^p(P)$, and weakly lower semicontinuous on $L_2^p(P)$. ////

Moreover, we give the solutions to the corresponding asymptotic MSE problems

$$\mathrm{maxMSE}\,(\eta, r) = E|\eta|^2 + r^2\omega_{c,t}(\eta)^2 = \min\,! \qquad \eta \in \Psi_{2\bullet}^D \tag{A.2.5}$$

where the radius $r \in (0, \infty)$ is fixed.

Remark A.2.2 As mentioned at the beginning of this chapter, the form of the solutions is identical to form of the solutions in case of linear regression. One just has to identify H with the regressor X and V with the error U. ////

First, we specify the solution in case $* = c$ and $t = \alpha = 1$.

Theorem A.2.3 (a) *Assume* $\operatorname{med}(\Lambda_f)$ *is unique. In case* $\omega_{c,1}^{\min} < b < \omega_{c,1}(\eta_h)$ *there exist a Borel measurable function* $z\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$, *and some matrix* $A \in \mathbb{R}^{p\times k}$, *such that*

$$\tilde\eta(H,V) = AH\big(\Lambda_f(V) - z(H)\big)w(H,V) \tag{A.2.6}$$

$$w(H,V) = \min\Big\{1, \frac{b}{|AH(\Lambda_f(V) - z(H))|}\Big\} \tag{A.2.7}$$

is the solution, and

$$\omega_{c,1}(\tilde\eta) = b \tag{A.2.8}$$

Conversely, if some $\tilde\eta \in \Psi_{2\bullet}^D$ *has form (A.2.6), for any function* $z\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$, *any matrix* $A \in \mathbb{R}^{p\times k}$, *and bound* $b \in (0,\infty)$, *then* $\tilde\eta$ *is the solution to problem (A.2.1).*

(b) *Given some radius* $r \in (0,\infty)$, *the unique solution* $\tilde\eta \in \Psi_{2\bullet}^D$ *to the MSE problem (A.2.5) is of sufficient and, if* $\operatorname{med}(\Lambda_f)$ *is unique, also of necessary form (A.2.6), with* $b \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *related via*

$$r^2 b = \mathrm{E}(|AH(\Lambda_f - z)| - b)_+ \tag{A.2.9}$$

(c) *With any* $m = \operatorname{med}(\Lambda_f)$, *we have*

$$\omega_{c,1}^{\min} = \max\Big\{\frac{\operatorname{tr} AD^\tau}{\mathrm{E}\,|AH|\,\mathrm{E}\,|\Lambda_f - m|}\;\Big|\; A \in \mathbb{R}^{p\times k}\setminus\{0\}\Big\} \tag{A.2.10}$$

There exist $A \in \mathbb{R}^{p\times k}\setminus\{0\}$ *and* $\bar\eta \in \Psi_{2\bullet}^D$ *achieving* $\omega_{c,1}^{\min}$, *respectively. And then necessarily*

$$\bar\eta(H,V) = b\frac{AH}{|AH|}\operatorname{sign}(\Lambda_f(V) - m) \qquad \text{on } \{AH(\Lambda_f(V) - m) \neq 0\} \tag{A.2.11}$$

Under condition

$$e \in \mathbb{R}^k,\ K(e^\tau H = 0) > 0 \quad\Longrightarrow\quad e = 0 \tag{A.2.12}$$

i.e., $K(E) = 0$ *for all* $(k-1)$*-dimensional linear subspaces* $E \subset \mathbb{R}^k$, *the solution is,*

$$\bar\eta(H,V) = b\frac{AH}{|AH|}\,\mathrm{I}(AH \neq 0)\big(\operatorname{sign}(\Lambda_f(V) - m) + \beta\,\mathrm{I}(\Lambda_f(V) = m)\big) \tag{A.2.13}$$

with $\beta \in [-1,1]$ *determined from* $\mathrm{E}\operatorname{sign}(\Lambda_f - m) + \beta P(\Lambda_f = m) = 0$.

Proof Analogously to Rieder (1994), Theorem 7.4.13. ////

Remark A.2.4 Due to the side condition $\mathrm{E}_\bullet\,\eta = 0$, the verification of the well-posedness of the optimization problem (A.2.1) gets complicated in case $* = c$ and $t = \alpha = 1$; confer Rieder (1994), p 276. Then, the additional assumption $\operatorname{med}(\Lambda_f)$ is unique is required (cf. Lemma C.2.4 of Rieder (1994)) except K has finite support. In addition, one can remove the constraint $\mathrm{E}_\bullet\,\eta = 0$, if F is symmetric; confer Subsection 9.2.1. ////

Second, we give the solution in case $* = c$ and $t = \alpha = 2$.

Theorem A.2.5 (a) *In case* $\omega_{c,2}^{\min} < b < \omega_{c,2}(\eta_h)$ *there exist constants* $z \in \mathbb{R}$ *and* $c, A \in (0,\infty)$ *such that the solution to problem (A.2.1) attains the following necessary and sufficient form*

$$\tilde{\eta}(H,V) = AD\mathcal{K}_F^{-1}H\big(\Lambda_f(V) - z\big)w(V) \qquad w(V) = \min\left\{1, \frac{c}{|\Lambda_f(V) - z|}\right\} \tag{A.2.14}$$

where

$$\mathcal{K}_F = \mathrm{E}\, HH^\tau \qquad z = \mathrm{E}\,\Lambda_f w[\mathrm{E}\, w]^{-1} \qquad A^{-1} = \mathrm{E}(\Lambda_f - z)^2 w \tag{A.2.15}$$

and

$$b = cA\sqrt{\operatorname{tr} D\mathcal{K}_F^{-1}D^\tau} \tag{A.2.16}$$

Conversely, if some $\tilde{\eta} \in \Psi_{2\bullet}^D$ *has form (A.2.14), for any* $z, A \in \mathbb{R}$ *and bound* $b \in (0,\infty)$, *then* $\tilde{\eta}$ *is the solution to problem (A.2.1).*

(b) *Given some radius* $r \in (0,\infty)$, *the unique solution* $\tilde{\eta} \in \Psi_{2\bullet}^D$ *to the MSE problem (A.2.5) is of necessary and sufficient form (A.2.14), with* $c \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *related via*

$$r^2 c = \mathrm{E}(|\Lambda_f - z| - c)_+ \tag{A.2.17}$$

(c) *With any* $m = \operatorname{med}(\Lambda_f)$, *the minimal bias is*

$$\omega_{c,2}^{\min} = \frac{\sqrt{\operatorname{tr} D\mathcal{K}_F^{-1}D^\tau}}{\mathrm{E}\,|\Lambda_f - m|} \tag{A.2.18}$$

and

$$\bar{\eta}(H,V) = \frac{D\mathcal{K}_F^{-1}H}{\mathrm{E}\,|\Lambda_f(V) - m|}\big(\operatorname{sign}(\Lambda_f(V) - m) + \beta\,\mathrm{I}(\Lambda_f(V) = m)\big) \tag{A.2.19}$$

with $\beta \in [-1,1]$ *determined from* $\mathrm{E}\operatorname{sign}(\Lambda_f - m) + \beta P(\Lambda_f = m) = 0$, *is the solution in case* $b = \omega_{c,2}^{\min}$.

Proof

(c) Analogously to Theorem 7.4.15 (c) in Rieder (1994).

(a) The first part is identical to the proof of Theorem 7.4.15 (a) in Rieder (1994) where the sufficient form

$$\tilde{\eta} = (\tilde{A}H\Lambda_f - a)\min\left\{1, \frac{\tilde{c}}{|\tilde{A}H\Lambda_f - a|}\right\} \tag{A.2.20}$$

$$\mathrm{E}_\bullet\left(|\tilde{A}H\Lambda_f - a| - \tilde{c}\right)_+ = \beta\tilde{c} \qquad \mathrm{E}\,\tilde{c}^2 = b^2 \tag{A.2.21}$$

of the solution $\tilde{\eta}$ is derived where $\tilde{A} \in \mathbb{R}^{k\times k}$, $a\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^k$ and $\tilde{c}\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. We then choose $\gamma \in (0,\infty)$ such that

$$b = \sqrt{\operatorname{tr} D\mathcal{K}_F^{-1}D^\tau}\,\gamma \tag{A.2.22}$$

As $b > \omega_{c,2}^{\min}$, we have $\gamma > \frac{1}{\mathrm{E}\,|\Lambda_f - m|}$. Moreover,

$$\omega_{c,2}(\eta_h) = \operatorname{tr} D\mathcal{K}_F^{-1} D^\tau \sup_\bullet |\mathcal{I}_f^{-1}\Lambda_f|^2 > b^2 = \operatorname{tr} D\mathcal{K}_F^{-1} D^\tau \gamma^2 \tag{A.2.23}$$

i.e., $\gamma < \sup_\bullet |\mathcal{I}_f^{-1}\Lambda_f|$. Now, consider the model $V \sim F$ with scores Λ_f and Fisher information $\mathcal{I}_f$. The corresponding Hampel type problem with bound γ is of form (1.3.7) and hence by Theorem 1.3.7 (a) there exist $z \in \mathbb{R}$, $A \in (0, \infty)$ such that the solution is of form

$$\tilde{\chi}(V) = A(\Lambda_f(V) - z) \min\left\{1, \frac{c}{|\Lambda_f(V) - z|}\right\} \qquad c = \gamma/A \tag{A.2.24}$$

Correspondingly to $\tilde{\chi}$, define

$$\tilde{\eta}(H, V) = D\mathcal{K}_F^{-1} H \tilde{\chi}(V) \tag{A.2.25}$$

Hence, $\tilde{\eta} \in \Psi_{2\bullet}$ and is of the sufficient form (A.2.20). In addition, we have

$$\mathrm{E}\,\gamma^2\,\mathrm{E}\,|D\mathcal{K}_F^{-1} H|^2 = \gamma^2 \operatorname{tr} D\mathcal{K}_F^{-1} D^\tau = b^2 \tag{A.2.26}$$

and

$$\mathrm{E}_\bullet\left(|AD\mathcal{K}_F^{-1} H(\Lambda_f - z)| - \gamma |D\mathcal{K}_F^{-1} H|\right)_+ = A|D\mathcal{K}_F^{-1} H|\,\mathrm{E}_\bullet\left(|\Lambda_f - z| - \tilde{c}\right)_+ \tag{A.2.27}$$

By identifying $\tilde{c}(H) = \gamma A |D\mathcal{K}_F^{-1} H|$ and defining $\beta = \gamma^{-1}\,\mathrm{E}\left(|\Lambda_f - z| - \gamma\right)_+$ we therefore obtain (A.2.21). Thus, $\tilde{\eta}$ is the solution.

The converse part is verified in the proof of Theorem 7.4.15 (a) in Rieder (1994).

(b) Follows by part (a) and Theorem 1.3.11 (b). ////

Remark A.2.6 The proof of part (a) of the previous theorem is based on arguments provided by H. Rieder. It simplifies the proof of Theorem 7.4.15 of Rieder (1994) as one can avoid an application of Lemma C.2.4 in Rieder (1994) and does not need $\operatorname{med}(\Lambda_f)$ unique to obtain the form (A.2.14) of the solution. ////

Finally, we state the solution in case $* = v$ and $t = \alpha = 1$.

Theorem A.2.7 (a) *Assume* $\operatorname{med}(\Lambda_f)$ *is unique. In case* $\omega_{v,1}^{\min} < b < \omega_{v,1}(\eta_h)$ *there exist a Borel measurable function* $c\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$, $c \le 0$, *and some matrix* $A \in \mathbb{R}^{p\times k}$, *such that*

$$\tilde{\eta}(H, V) = AH\Lambda_f(V) w(H, V) \tag{A.2.28}$$

$$w(H, V) = 1 \wedge \max\left\{\frac{c(H)}{|AH|\Lambda_f(V)}, \frac{c(H) + b}{|AH|\Lambda_f(V)}\right\} \tag{A.2.29}$$

is the solution, where

$$\mathrm{E}_\bullet\left(c - |AH|\Lambda_f\right)_+ = \mathrm{E}_\bullet\left(|AH|\Lambda_f - (c + b)\right)_+ \tag{A.2.30}$$

$$D = A\,\mathrm{E}\,HH^\tau\,\mathrm{E}_\bullet\,\Lambda_f^2 w \tag{A.2.31}$$

$$\omega_{v,1}(\tilde{\eta}) = b \tag{A.2.32}$$

Conversely, for any $A \in \mathbb{R}^{p\times k}$ *and* $b \in (0,\infty)$, *(A.2.30) defines some* $c\colon \mathbb{R}^{\mathbb{Z}} \to (-\infty, 0]$. *And if* $\tilde\eta$ *is of form (A.2.28)–(A.2.31), then* $\tilde\eta \in \Psi_{2\bullet}^{D}$ *and* $\tilde\eta$ *is the solution to problem (A.2.1).*

(b) *Given some radius* $r \in (0,\infty)$, *the unique solution* $\tilde\eta \in \Psi_{2\bullet}$ *to the MSE problem (A.2.5) is of sufficient and necessary form (A.2.28)–(A.2.31), with* $b \in (0,\infty)$ *and radius* $r \in (0,\infty)$ *related by*

$$r^2 b = \mathrm{E}_\bullet\left(|AH|\Lambda_f - (c+b)\right)_+ \tag{A.2.33}$$

(c) *It holds that*

$$\omega_{v,1}^{\min} = \max\left\{\frac{\operatorname{tr} AD^\tau}{\mathrm{E}\,|AH|\,\mathrm{E}(\Lambda_f)_+}\,\middle|\, A \in \mathbb{R}^{p\times k}\setminus\{0\}\right\} \tag{A.2.34}$$

Assume (A.2.12). Then, there exists some matrix $A \in \mathbb{R}^{p\times k}$ *such that the solution in case* $b = \omega_{v,1}^{\min}$ *is given by*

$$\tilde\eta(H,V) = b\frac{AH}{|AH|}\,\mathrm{I}(AH\neq 0)\left(\frac{P(\Lambda_f<0)}{P(\Lambda_f\neq 0)}\,\mathrm{I}(\Lambda_f>0) - \frac{P(\Lambda_f>0)}{P(\Lambda_f\neq 0)}\,\mathrm{I}(\Lambda_f<0)\right) \tag{A.2.35}$$

where

$$D = b\,\mathrm{E}\,\frac{AHH^\tau}{|AH|}\,\mathrm{I}(AH\neq 0)\,\mathrm{E}(\Lambda_f)_+ \tag{A.2.36}$$

Conversely, if $A \in \mathbb{R}^{p\times k}$ *is any matrix verifying (A.2.36) with* $b = \omega_{v,1}^{\min}$, *then* $\tilde\eta$ *defined by (A.2.35) is the solution.*

Proof Analogously to Rieder (1994), Theorem 7.4.17. ////

Remark A.2.8 If $\mathcal{L}(\Lambda_f)$ is symmetric, the solution in case of average conditional contamination neighborhoods ($* = c$, $t = \alpha = 1$) of radius r is automatically conditionally centered (i.e., $z(H_{\theta,1}) \equiv 0$ is a possible choice) and hence coincides with the solution in case of average conditional total variation neighborhoods ($* = v$, $t = \alpha = 1$) of radius $2r$. ////

Appendix B

The Kronecker Product and the Vec and Vech Operators

This appendix is based on Chapter 16 of Harville (1997) and provides some results about the Kronecker product of two matrices and the vec and vech operators needed in Chapter 7.

B.1 The Kronecker Product and the Vec Operator

We first define the Kronecker product for two matrices which is denoted by $\otimes$; confer Section 16.1 of Harville (1997).

Definition B.1.1 *Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ with $m,n,p,q \in \mathbb{N}$ and entries (a_{ij}) and (b_{ij}), respectively. Then, the* Kronecker product *is the $mp \times nq$ matrix*

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix} \tag{B.1.1}$$

The vec operator is defined as follows; confer Section 16.2 of Harville (1997).

Definition B.1.2 *Let $A \in \mathbb{R}^{m\times n}$ $(m,n \in \mathbb{N})$ with entries (a_{ij}). Then, the* vec operator *is the mn dimensional column vector*

$$\operatorname{vec}(A) = (a_{11},\dots,a_{m1},a_{12},\dots,a_{m2},\dots,a_{1n},\dots,a_{mn})^\tau \tag{B.1.2}$$

In Subsection 7.2.2 we make use of the following result.

Theorem B.1.3 *For any $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times q}$ $(m,n,p,q \in \mathbb{N})$, it holds*

$$\operatorname{vec}(ABC) = (C^\tau \otimes A)\operatorname{vec}(B) \tag{B.1.3}$$

PROOF Harville (1997), Theorem 16.2.1. ////

Remark B.1.4 As a consequence, we obtain for any $A, B \in \mathbb{R}^{m\times n}$ $(m, n \in \mathbb{N})$,

$$\operatorname{tr}(A^\tau B) = \big(\operatorname{vec}(A)\big)^\tau \operatorname{vec}(B) \tag{B.1.4}$$

confer Harville (1997), p 342. ////

B.2 The Vech Operator

We first give the definition of the vech operator; confer Section 16.4 of Harville (1997).

Definition B.2.1 *Let $A \in \mathbb{R}^{n\times n}$ $(n \in \mathbb{N})$ with entries (a_{ij}). Then, the* vech operator *is the $n(n+1)/2$ dimensional column vector*

$$\operatorname{vec}(A) = (a_{11}, \ldots, a_{m1}, a_{22}, \ldots, a_{m2}, \ldots, a_{nn})^\tau \tag{B.2.1}$$

Remark B.2.2 In case of a symmetric matrix A the matrix is determined by the $n(n+1)/2$ elements that are on or below (resp. above) the diagonal. Hence, such matrices contain redundant information. This information can be eliminated via the vech operator which is an application of the vec operator to the matrix A and afterwards deleting the supradiagonal elements of A. The definition is not restricted to symmetric matrices but, obviously, in case of asymmetric matrices some information about the elements of the matrix A is lost. ////

For symmetric matrices $A \in \mathbb{R}^{n\times n}$ $(n \in \mathbb{N})$ the column vector $\operatorname{vec}(A)$ can be reconstructed from the column vector $\operatorname{vech}(A)$ via the duplication matrix G_n which is defined as follows: (cf. Subsection 16.4.b of Harville (1997))

Definition B.2.3 *Let $A \in \mathbb{R}^{n\times n}$ $(n \in \mathbb{N})$ be symmetric. Then, the* duplication matrix *is the unique $(n^2 \times n(n+1)/2)$ dimensional matrix G_n, such that*

$$\operatorname{vec} A = G_n \operatorname{vech} A \tag{B.2.2}$$

Remark B.2.4 **(a)** The duplication matrix G_n can be described via its rows. For $i \geq j$ $(i = 1, \ldots, n;\ j = i, \ldots, n)$, the $[(j-1)n + i]$th and $[(i-1)n + j]$th rows of G_n are equal to the $[(j-1)(n - j/2) + i]$th row of $\mathbb{I}_{n(n+1)/2}$. Similarly, one can describe G_n via its columns; confer Harville (1997), p 352.

(b) Since every row of $\mathbb{I}_{n(n+1)/2}$ is a row of G_n, it contains $n(n+1)/2$ linearly independent rows. Hence, G_n has full column rank; i.e.,

$$\operatorname{rk}(G_n) = n(n+1)/2 \tag{B.2.3}$$

confer Harville (1997), p 352.

(c) For any $A, B \in \mathbb{R}^{n\times n}$, B symmetric we obtain by Theorem B.1.3 and symmetry of ABA^τ,

$$\operatorname{vech}(ABA^\tau) = H_n(A \otimes A)G_n \operatorname{vech} B \tag{B.2.4}$$

confer Harville (1997), p 357. ////

Appendix C

Convolution via Fast Fourier Transform

We first introduce the discrete Fourier transform (DFT) which may be calculated very fast using the fast Fourier transform (FFT) and state the convolution theorem for DFTs; confer Section C.1. In Section C.2, we establish the connection between the convolution of DFTs and the convolution of integer lattice distributions and propose a convolution algorithm that works for quite general probability distributions. Finally, we present the results of different checks which show that this algorithm yields astonishingly exact results; confer Section C.3.

C.1 Discrete Fourier Transform

This section is based on Lesson 8 of Gasquet and Witomski (1999). Let $m \in \mathbb{N}$ and let $(x_n)_{n\in\mathbb{Z}}$ be a sequence of complex numbers with period m; i.e., $x_{n+m} = x_n$ for all $n \in \mathbb{Z}$. Then, the discrete Fourier transformation (DFT) of order m is,

$$\mathrm{DFT}_m \colon \mathbb{C}^m \to \mathbb{C}^m,\ (x_0, x_1, \ldots, x_{m-1}) \mapsto (\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_{m-1}) \tag{C.1.1}$$

where

$$\hat{x}_n = \frac{1}{m} \sum_{j=0}^{m-1} x_j \omega_m^{jn} \qquad \omega_m = e^{-2\pi i/m},\ i = \sqrt{-1} \tag{C.1.2}$$

and we obtain the DFT $(\hat{x}_n)_{n\in\mathbb{Z}}$ of $(x_n)_{n\in\mathbb{Z}}$ by the periodic extension $\hat{x}_{n+m} = \hat{x}_n$ for all $n \in \mathbb{Z}$. The DFT_m is represented by the matrix Ω_m with entries ω_m^{jk} $(j, k = 0, 1, \ldots, m-1)$ and inverse $\Omega_m^{-1} = 1/m\,\overline{\Omega}_m$ ($\overline{\Omega}_m$ the conjugate DFT_m); i.e., DFT_m is linear and bijective.

Remark C.1.1 (a) Computing $\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_{m-1}$ directly from equation (C.1.2), requires $(m-1)^2$ complex multiplications and $m(m-1)$ complex additions. But, the algorithm developed by Cooley and Tukey (1965), which is known as fast Fourier transform (FFT), is of order $m \log m$. It works best for the case $m = 2^p$ $(p \in \mathbb{N})$;

confer Lesson 9 of Gasquet and Witomski (1999). If $m = 2^{10} = 1024$, the direct computation needs 1046529 multiplications and 1047552 additions whereas in case of the FFT only 4097 multiplications and 10240 additions are needed; confer Table 9.1 of Gasquet and Witomski (1999).

(b) If $(x_n)_{n\in\mathbb{Z}}$ is a sequence of real numbers, it is possible to reduce the cost of computation by half; confer Section 8.3 of Gasquet and Witomski (1999). ////

We now state the convolution theorem for DFTs which is the key property for the convolution algorithm given in Section C.2.

Theorem C.1.2 *Let $x = (x_n)_{n\in\mathbb{Z}}$ and $y = (y_n)_{n\in\mathbb{Z}}$ be two sequences of complex numbers with period m and let $\hat{x} = (\hat{x}_n)_{n\in\mathbb{Z}}$ and $\hat{y} = (\hat{y}_n)_{n\in\mathbb{Z}}$ be the corresponding DFTs. Then, the circular convolution product of x and y is defined as,*

$$x * y = \left(\sum_{j=0}^{m-1} x_j y_{n-j} \right)_{n\in\mathbb{Z}} \tag{C.1.3}$$

and it holds,

$$\hat{z} = m\,\hat{x}\,\hat{y} \qquad \text{with} \qquad z = x * y \tag{C.1.4}$$

where $\hat{x}\,\hat{y} = (\hat{x}_n \hat{y}_n)_{n\in\mathbb{Z}}$.

Proof For $n \in \mathbb{Z}$ we obtain,

$$\begin{aligned}\hat{z}_n &= \frac{1}{m} \sum_{j=0}^{m-1} z_j \omega_m^{jn} = \frac{1}{m} \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} x_k y_{j-k} \omega_m^{jn} = \frac{1}{m} \sum_{k=0}^{m-1} x_k \omega_m^{nk} \sum_{j=0}^{m-1} y_{j-k} \omega_m^{n(j-k)} \\ &= m\,\hat{x}_n\,\hat{y}_n\end{aligned}$$

by the periodicity of $(y_n)_{n\in\mathbb{Z}}$ and $\omega_m^{nm} = 1$ for all $n \in \mathbb{Z}$. ////

This Theorem implies the following result for N-fold convolution products.

Proposition C.1.3 *Let $x = (x_n)_{n\in\mathbb{Z}}$ be a sequence of complex numbers with period m and let $\hat{x} = (\hat{x}_n)_{n\in\mathbb{Z}}$ be the corresponding DFT. Then, it holds,*

$$\widehat{*_{i=1}^{N} x} = m^{N-1}\,\hat{x}^N \qquad N \in \mathbb{N} \tag{C.1.5}$$

Proof Immediately follows from Theorem C.1.2 by induction. ////

C.2 Convolution Algorithm

The convolution algorithm formulated in this section is based on the fact that the convolution of integer lattice distributions may be calculated very fast using FFT

and its inverse. Therefore, we first consider the following special integer lattice distributions

$$F_i(x) = \sum_{j=0}^{m-1} p_{i,j}\, \mathrm{I}_{[j,\infty)}(x) \qquad i = 1, 2 \tag{C.2.1}$$

with

$$p_{i,j} \geq 0 \qquad j = 0, 1, \ldots, m-1 \qquad \textstyle\sum_{j=0}^{m-1} p_{i,j} = 1 \tag{C.2.2}$$

where $x \in \mathbb{R}$ and $m = 2^q$ $(q \in \mathbb{N})$. We extend $p_{i,j}$ $(i = 1, 2,\ j = 0, \ldots, m-1)$ to two sequences $p_i = (p_{i,n})_{n\in\mathbb{Z}}$ of real numbers with period $2m$ via,

$$p_{i,j} = 0 \qquad i = 1, 2 \qquad j = m, \ldots, 2m-1 \qquad \text{(zero padding)} \tag{C.2.3}$$

and

$$p_{i,k+2m} = p_{i,k} \qquad \forall\, k \in \mathbb{Z} \tag{C.2.4}$$

Then, the convolution F of F_1 and F_2 is an integer lattice distribution given by

$$F(x) = (F_1 * F_2)(x) = \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} p_{1,j} p_{2,k}\, \mathrm{I}_{[k,\infty)}(x - j) \tag{C.2.5}$$

$$= \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} p_{1,j} p_{2,k}\, \mathrm{I}_{[k+j,\infty)}(x) \tag{C.2.6}$$

$$= \sum_{j=0}^{2m-1} \pi_j\, \mathrm{I}_{[j,\infty)}(x) \qquad \text{with} \qquad \pi_j := \sum_{k=0}^{2m-1} p_{1,k} p_{2,j-k} \tag{C.2.7}$$

where in particular $\pi_{2m-1} = 0$. Thus, in view of Theorem C.1.2, $\pi = (\pi_n)_{n\in\mathbb{Z}} = p_1 * p_2$ and we can compute π very fast using FFT and its inverse. This result was the starting point for Algorithm C.2.2.

Remark C.2.1 (a) A very similar algorithm was proposed by Bertram (1981) to numerically evaluate compound distributions in insurance mathematics where he assumes claim size distributions of lattice type. Numerical examples and comparisons to other methods can be found in Bühlmann (1984) and Feilmeier and Bertram (1987). A nice mathematical formulation of the corresponding algorithm (cf. Algorithm 1) is included in Grübel and Hermesmeier (1999). However, the main purpose of their article is the investigation of the aliasing error which occurs as a wrap-around effect due to summation modulo m (cf. equation (C.1.3)). In case of a claim size distribution of lattice type they obtain a simple general bound for the aliasing error and show that this error can be eliminated by a suitable change of measure (exponential tilting). But, this wrap-around error can also be made very small if we choose ε in step 1 of Algorithm C.2.2 small enough and q in step 2 large enough. Thus, in many cases this effect is negligible; confer also Remark C.2.3 (a). However, it is of interest if heavy-tailed probability distributions are involved.
Moreover, if one considers absolutely continuous probability distributions, a initial

discretization step is necessary; confer also step 2 of Algorithm C.2.2. The corresponding error is studied in Grübel and Hermesmeier (2000) and it is shown that this error, under certain smoothness conditions, can be reduced by an extrapolation technique (Richardson extrapolation).

(b) In Embrechts et al. (1993) the authors describe how one can use FFT to determine various quantities of interest in risk theory and insurance mathematics including the computation of the total claim size distribution, the mean and the variance of the claim size process and the probability of ruin. Moreover, using FFT it is also possible to find the stationary waiting time distribution for a given customer interarrival time distribution and a given service time distribution in the G/G/1 queueing model; confer Grübel (1991). ////

We state the algorithm for absolutely continuous distributions but with slight and obvious modifications this algorithm works for quite general distributions; confer Remark C.2.3.

Algorithm C.2.2
Assume two absolutely continuous distributions F_1, F_2 on $\mathbb{R}$.

Step 1: (Truncation)
If the support of F_i $(i = 1, 2)$ is unbounded or "too large", we define numbers $A_i, B_i \in \mathbb{R}$, for given $\varepsilon > 0$, such that,

$$F_i\big((-\infty, A_i)\big) = \frac{\varepsilon}{2} \qquad \text{and} \qquad F_i\big((B_i, \infty)\big) = \frac{\varepsilon}{2} \tag{C.2.8}$$

and put $A = \min\{A_1, A_2\}$ and $B = \max\{B_1, B_2\}$. If this is not the case, we define $A := \min\{F_1^{-1}(0), F_2^{-1}(0)\}$ and $B := \max\{F_1^{-1}(1), F_2^{-1}(1)\}$ where F_i^{-1} $(i = 1, 2)$ are the quantile functions of F_i.

Step 2: (Discretization on a real grid)
Given $m = 2^q$ $(q \in \mathbb{N})$ and F_i $(i = 1, 2)$, we define the lattice distributions

$$G_i(x) := \sum_{j=0}^{m-1} p_{i,j}\, \mathrm{I}_{[A+(j+0.5)h,\infty)}(x) \qquad h = \frac{B - A}{m} \tag{C.2.9}$$

where

$$p_{i,j} \;=\; F_i\big([A + jh, A + (j + 1)h]\big) \tag{C.2.10}$$

for $j = 0, 1, \ldots, m - 1$.

Step 3: (Transformation to an integer grid)
Based on G_i $(i = 1, 2)$, we define the integer lattice distributions

$$\tilde{G}_i(x) := \sum_{j=0}^{m-1} p_{i,j}\, \mathrm{I}_{[j,\infty)}(x) \qquad i = 1, 2 \tag{C.2.11}$$

and extend $p_{i,j}$ $(i = 1, 2,\ j = 0, \ldots, m-1)$ to two sequences $p_i = (p_{i,n})_{n\in\mathbb{Z}}$ of real numbers with period $2m$ via,

$$p_{i,j} = 0 \qquad i = 1, 2 \qquad j = m, \ldots, 2m-1 \qquad \text{(zero padding)} \tag{C.2.12}$$

and

$$p_{i,k+2m} = p_{i,k} \qquad \forall\, k \in \mathbb{Z} \tag{C.2.13}$$

Step 4: (Convolution by FFT on integer grid)
We calculate $\tilde{G} = \tilde{G}_1 * \tilde{G}_2$ by FFT and its inverse as given in (C.2.7); i.e.,

$$\tilde{G}(x) = \sum_{j=0}^{2m-1} \pi_j \, \mathrm{I}_{[j,\infty)}(x) \qquad \pi_j := \sum_{k=0}^{2m-1} p_{1,k} p_{2,j-k} \tag{C.2.14}$$

where in particular $\pi_{2m-1} = 0$.

Step 5: (Back-transformation to real grid)
Given $\tilde{G}$, we obtain $G = G_1 * G_2$ by,

$$G(x) = \sum_{j=0}^{2m-2} \pi_j \, \mathrm{I}_{[2A+(j+1.5)h,\infty)}(x) \tag{C.2.15}$$

That is, we additionally use a continuity correction of $h/2$, which clearly improves the results; confer Remark C.3.6 (a).

Step 6: (Smoothing)
Next, we use interpolation of the values of G on

$$\{2A,\ 2A + 1.5h,\ 2A + 2.5h,\ \ldots,\ 2B - 0.5h,\ 2B\} \tag{C.2.16}$$

by linear functions to get a continuous approximation $F^\natural$ of $F = F_1 * F_2$. We obtain a continuous approximation $f^\natural$ of the density f of F by multiplying $\{0, \pi_0, \pi_1, \ldots, \pi_{2m-2}, 0\}$ by h and interpolating these values on the grid $\{2A, 2A + h, \ldots, 2B - h, 2B\}$ (no continuity correction) using linear functions.

Step 7: (Standardization)
To make sure that the approximation $F^\natural$ is indeed a probability distribution, we standardize $F^\natural$ and $f^\natural$ by $F^\natural\big([2A, 2B]\big)$ and $\int f^\natural(x)\,dx$, respectively, where $\int f^\natural(x)\,dx$ may be calculated numerically exact by using the fact that $f^\natural$ is a piecewise linear function.

Remark C.2.3 **(a)** In view of Remark C.2.1 the results of this algorithm can even be improved under certain additional conditions (e.g., smoothness of the densities) and in certain special cases (e.g., in case of heavy-tailed distributions).

(b) In step 1, a support is considered as "too large" if a uniform grid with a reasonable step-length produces too many grid points. In the same sense, the loss of mass included in step 1 of Algorithm C.2.2 is, to some extent, controllable and in many cases negligible.

(c) Obviously, Algorithm C.2.2 may easily be applied to lattice distributions F_1, F_2 on $\mathbb{R}$ that are defined on the same grid. In this case the algorithm can essentially be reduced to steps 1-5 and 7. Moreover, the results are numerically exact if the lattice distributions have finite support; confer Section C.3. In this case the algorithm consists only of steps 2-5.

(d) This algorithm is implemented in our `R` package `distr` (cf. Ruckdeschel et al. (2005)) which was developed in joint work with P. Ruckdeschel, T. Stabla and F. Camphausen and treats one-dimensional distributions by means of `S4` classes (cf. Chambers (1998)). To determine a continuous approximation of the quantile function we interpolate $\{2A, 2A + 1.5h, \ldots, 2B - 0.5h, 2B\}$ on the values of the standardized $F^{\natural}$ by means of linear functions (numerical inversion). There are also some instructive examples, like the computation of an approximation to the stationary regressor distribution of an AR(1) process, together with the corresponding `R` sources included in Ruckdeschel et al. (2005). A description of this algorithm in the object orientated context of `S4` classes can be found in Kohl et al. (2005).

(e) In the framework of this thesis we modify Algorithm C.2.2 in view of Proposition C.1.3 to compute an approximation of the exact finite-sample distribution of M estimates; confer Part V. This needs some care since the underlying distributions contain points which carry mass; confer Subsection 11.3.1 and Subsubsections 12.1.3.1 and 12.2.3.1. But, we can handle this and obtain very accurate results; confer Subsection 11.3.2 and Subsubsections 12.1.3.2 and 12.2.3.2. In Subsection 11.3.3, respectively more detailed in Ruckdeschel and Kohl (2005) we compare the results of this algorithm with other approximations of the exact finite-sample distribution of M estimates, like the saddle point approximation of Daniels (1954) and Hampel (1974a). In the latter study we also take into account the results of higher order asymptotics contained in Ruckdeschel (2004a), Ruckdeschel (2004b), Ruckdeschel (2004c) and Ruckdeschel (2005e). ////

C.3 Checks

In the framework of this thesis Algorithm C.2.2 is used for the computation of an approximation of the exact finite-sample distribution of M estimates; confer Part V. Therefore, we now present checks for n-fold convolution products. We determine the precision of this convolution algorithm in terms of the total variation distance of the densities,

$$d_v(P,Q) = \tfrac{1}{2}\int |p-q|\,d\mu = \sup_{B\in\mathbb{B}} \big|P(B) - Q(B)\big| \tag{C.3.1}$$

where $P, Q \in \mathcal{M}_1(\mathbb{B})$ with $dP = p\,d\mu$, $dQ = q\,d\mu$ for some σ-finite measure μ on $(\mathbb{R}, \mathbb{B})$ and the Kolmogorov distance of the cumulative distribution functions,

$$d_\kappa(P,Q) = \sup_{t\in\mathbb{R}} \big|P\big((-\infty,t]\big) - Q\big((-\infty,t]\big)\big| \tag{C.3.2}$$

Obviously, $d_\kappa \le d_v$ as the supremum in case of the total variation distance is taken over more sets. In the sequel $d_v^{\natural}$ and $d_\kappa^{\natural}$ stand for the numerical approximations of d_v and d_κ.

The first example treats binomial distributions and shows that the convolution algorithm is numerically exact for integer lattice distributions with finite support. In particular, this implies that the computation of the FFT and its inverse is numerically exact.

Example C.3.1 Assume $F = \mathrm{Binom}\,(k,p)$ with $k \in \mathbb{N}$ and $p \in (0,1)$. Then, the n-fold convolution product is $F^{*n} = \mathrm{Binom}\,(nk,p)$ $(n \in \mathbb{N})$. Let f_n and $f^\natural$ be the probability functions of F^{*n} and $F^\natural$, respectively. Then, we may determine $d_v^\natural$ and $d_\kappa^\natural$ numerically exact by,

$$d_v^\natural(F, F^\natural) = \tfrac{1}{2} \sum_{j=0}^{nk} |f_n(j) - f^\natural(j)| \tag{C.3.3}$$

and

$$d_\kappa^\natural(F, F^\natural) = \max_{j \in \{0,\ldots,nk\}} \left| F^{*n}([0,j]) - F^\natural([0,j]) \right| \tag{C.3.4}$$

We obtain the results contained in Table C.1 which show that Algorithm C.2.2 is numerically exact in case of binomial distributions, where the values of k and p are chosen arbitrarily. The computation time for the convolution at $n = 1000$, $k = 50$ and $p = 0.4$ on an Athlon with 2.5 GHz and 512 MB RAM is about 0.2 seconds. ////

n	k	p	$d_v^\natural$	$d_\kappa^\natural$
2	10	0.5	1.9e−16	1.1e−16
3	5	0.3	3.1e−16	2.2e−16
5	20	0.7	5.9e−16	3.9e−16
10	30	0.8	1.4e−15	1.1e−15
50	25	0.1	4.4e−15	3.4e−15
100	15	0.2	8.5e−15	8.3e−15
1000	50	0.4	7.0e−14	6.6e−14

Table C.1: The precision of the convolution of binomial distributions via FFT; confer Example C.3.1.

In case of the Poisson distribution the results of the convolution algorithm turn out to be very accurate, too.

Example C.3.2 We consider $F = \mathrm{Pois}\,(\lambda)$ with $\lambda \in (0,\infty)$. Then, it holds, $F^{*n} = \mathrm{Pois}\,(n\lambda)$ $(n \in \mathbb{N})$. Since the support of F is $\mathbb{N}_0$, we use $A = 0$ and $B = F^{-1}(1 - 1\mathrm{e}{-15})$ in step 1 of Algorithm C.2.2 and determine $d_v^\natural$ and $d_\kappa^\natural$

numerically exact by,

$$d_v^\natural(F, F^\natural) = \tfrac{1}{2} \sum_{j=0}^{M} |f_n(j) - f^\natural(j)| \tag{C.3.5}$$

and

$$d_\kappa^\natural(F, F^\natural) = \max_{j \in \{0,\ldots,M\}} \left| F^{*n}([0,j]) - F^\natural([0,j]) \right| \tag{C.3.6}$$

where M is the $1 - 3\text{e}{-16}$ quantile of F^{*n}. For further details on the computation of $d_v^\natural$ and $d_\kappa^\natural$ we refer to Remark C.3.3. We obtain the results contained in Table C.2 which demonstrate the high precision of the convolution algorithm in case of Poisson distributions where the parameter λ is chosen arbitrarily. The computation time for the convolution at $n = 1000$ and $\lambda = 50$ on an Athlon with 2.5 GHz and 512 MB RAM is about 0.6 seconds. ////

n	λ	$d_v^\natural$	$d_\kappa^\natural$
2	0.1	2.9e−16	3.3e−16
3	1.5	2.6e−16	3.3e−16
5	10.0	2.8e−15	2.9e−15
10	7.5	4.9e−15	4.8e−15
50	25.0	2.0e−14	1.6e−14
100	15.0	1.9e−14	1.4e−14
1000	50.0	3.4e−13	3.3e−13

Table C.2: The precision of the convolution of Poisson distributions via FFT; confer Example C.3.2.

Remark C.3.3 **(a)** The differences between $d_\kappa^\natural$ and $d_v^\natural$ are below the computational accuracy in case of $n = 2, 3, 5, 10$ which is about $2\text{e}{-16}$. That is, the values of $d_\kappa^\natural$ and $d_v^\natural$ are numerically not distinguishable.

(b) By checking the output of the convolution algorithm, we found that the cumulative distribution function of the Poisson distribution `ppois` included in `R` (cf. R Development Core Team (2005)) becomes less exact for bigger λ; confer also the `R-help` mail archives of January 2004 and March 2004 which can be found under

`http://maths.newcastle.edu.au/~rking/R/help/04/01/0513.html`

and

`http://maths.newcastle.edu.au/~rking/R/help/04/03/0036.html`

respectively. Thus, we directly sum over the numerically exact values of the probability function `dpois` in order to obtain the numerically exact Kolmogorov distances given in Table C.2. ////

In the next two examples we consider the convolution of absolutely continuous distributions. We determine the total variation distance $d_v^\natural(F, F^\natural)$ by numerical integration using the `R` function `integrate`; confer R Development Core Team (2005). To compute an approximation of the Kolmogorov distance, we evaluate $d_\kappa^\natural(F, F^\natural)$ on a random grid which we obtain by sampling pseudo-random numbers of Unif ([Min, Max]) where Min and Max are the $\varepsilon/10$ and $1 - \varepsilon/10$ quantile of F^{*n}. We first present the results for normal distributions.

Example C.3.4 Assume $F = \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$. Then, it holds, $F^{*n} = \mathcal{N}(n\mu, n\sigma^2)$ $(n \in \mathbb{N})$. Starting with $\mathcal{N}(0, 1)$ and A and B as defined in step 1 of Algorithm C.2.2 we obtain $\tilde{A} = \sigma A + \mu$ and $\tilde{B} = \sigma B + \mu$ in case of $\mathcal{N}(\mu, \sigma^2)$. That is, the grid transforms the same way as the normal distributions do. Thus, we expect the precision of the results to be independent of μ and σ. This is indeed confirmed by our numerical calculations; confer Table C.3. We therefore may consider $\mu = 0$ and $\sigma = 1$ for the study of the accuracy of the convolution algorithm subject to $n \in \mathbb{N}$, $\varepsilon > 0$ (step 1) and $q \in \mathbb{N}$ (step 2). The results included in Table C.4 show that the precision is almost independent of n. It mainly depends on q where the maximum accuracy, we can reach, is of order ε. ////

n	ε	q	μ	σ	$d_v^\natural$	$d_\kappa^\natural$
2	1e−08	12	-10.0	100.0	3.2e−07	1.4e−07
			-2.0	5.0	3.2e−07	1.4e−07
			0.0	1.0	3.2e−07	1.4e−07
			1.0	50.0	3.2e−07	1.4e−07
			100.0	1000.0	3.2e−07	1.4e−07

Table C.3: The precision of the convolution of normal distributions via FFT is independent of the parameters μ and σ; confer Example C.3.4.

Our last example treats the convolution of exponential distributions which leads to gamma distributions.

Example C.3.5 We consider $F = \text{Exp}(\lambda) = \text{Gamma}(1, \lambda)$ with $\lambda \in (0, \infty)$. Then, it holds, $F^{*n} = \text{Gamma}(n, \lambda)$ $(n \in \mathbb{N})$. Analogously to the normal case (cf. Example C.3.4), the grid transforms the same as the exponential distributions do; i.e., $\tilde{A} = 1/\lambda\, A$ and $\tilde{B} = 1/\lambda\, B$. Thus, we expect the precision of the results to be independent of λ. This is again confirmed by our numerical computations; confer Table C.5. Next, we study the dependence of the accuracy of Algorithm C.2.2 on $n \in \mathbb{N}$, $\varepsilon > 0$ and $q \in \mathbb{N}$ where we may choose $\lambda = 1.0$ without restriction. As in Example C.3.4 the precision is almost independent of n. It mainly depends on q where the maximum accuracy, we can reach, is of order ε; confer Table C.6. ////

n	ε	q	$d_v^\natural$	$d_\kappa^\natural$
2	1e−06	8	5.6e−05	2.6e−05
		9	1.4e−05	5.6e−06
		10	4.5e−06	1.7e−06
		11	2.5e−06	1.8e−06
		12	2.1e−06	1.9e−06
	1e−08	10	4.8e−06	2.4e−06
		12	3.2e−07	1.4e−07
		13	8.8e−08	2.5e−08
		14	3.5e−08	1.9e−08
		15	2.3e−08	1.9e−08
	1e−10	12	4.0e−07	1.9e−07
		14	2.5e−08	1.2e−08
		16	1.7e−09	6.3e−10
		17	5.4e−10	1.9e−10
		18	2.9e−10	2.0e−10
5	1e−08	12	1.9e−07	7.5e−08
		13	6.0e−08	2.8e−08
		14	5.0e−08	3.5e−08
10	1e−08	12	1.4e−07	4.3e−08
		13	1.0e−07	3.7e−08
		14	1.0e−07	4.8e−08
50	1e−08	12	5.0e−07	8.2e−09
		13	5.0e−07	5.3e−08
		14	5.0e−07	6.7e−08

Table C.4: The precision of the convolution of normal distributions via FFT; confer Example C.3.4.

n	ε	q	λ	$d_v^\natural$	$d_\kappa^\natural$
2	1e−08	12	0.01	1.3e−06	2.5e−06
			0.5	1.3e−06	2.5e−06
			1.0	1.3e−06	2.5e−06
			5.0	1.3e−06	2.5e−06
			10.0	1.3e−06	2.5e−06

Table C.5: The Precision of the convolution of exponential distributions via FFT is independent of the parameter λ; confer Example C.3.5.

Remark C.3.6 (a) Without continuity correction (step 6) the Kolmogorov distances would clearly increase. For $n = 2$, $\varepsilon =$ 1e−08 and $q = 12$ we obtain in case of normal distributions $d_\kappa^\natural =$ 3.9e−04 instead of 1.4e−07 and in case of

exponential distributions $d_\kappa^\natural = 8.3\text{e}{-}04$ instead of $2.5\text{e}{-}06$.

(b) In case of normal and exponential distributions the computation time for $n = 2$, $\varepsilon = 1\text{e}{-}10$ and $q = 18$ on an Athlon with 2.5 GHz and 512 MB RAM is about 30 seconds, for $n = 50$, $\varepsilon = 1\text{e}{-}08$ and $q = 14$, it is about 14 seconds and for $n = 2$, $\varepsilon = 1\text{e}{-}08$ and $q = 12$ it is about 0.4 seconds.

(c) The Example C.3.5 reveals one minor flaw of Algorithm C.2.2. The support of $\text{Gamma}\,(n, \lambda)$ is $[0, \infty)$ whereas the convolution algorithm is only very exact in $[2A + (n/2 + 0.5)h, \ldots, 2B - (n/2 + 0.5)h]$. That is, for small n $(n \leq 5)$ the Kolmogorov distance is $F\big([0, 2A + (n/2 + 0.5)h)\big) - F^\natural\big([0, 2A + (n/2 + 0.5)h)\big)$. However, for bigger n this inexactness disappears as there is less and less mass in $[0,\ 2A + (n/2 + 0.5)h)$. Moreover, since $(n/2 + 0.5)h$ is very small, this also causes the numerical inaccuracy of $d_v^\natural$ for small n and leads to $d_\kappa^\natural > d_v^\natural$. ////

n	ε	q	$d_v^\natural$	$d_\kappa^\natural$
2	1e−06	8	1.9e−04	3.6e−04
		9	4.7e−05	9.0e−05
		10	1.3e−05	2.2e−05
		11	3.9e−06	5.6e−06
		12	1.8e−06	2.0e−06
	1e−08	10	2.1e−05	4.0e−05
		12	1.3e−06	2.5e−06
		13	3.3e−07	6.3e−07
		14	9.0e−08	1.6e−07
		15	3.2e−08	3.9e−08
	1e−10	12	2.1e−06	3.9e−06
		14	1.3e−07	2.5e−07
		16	7.9e−09	1.5e−08
		17	2.1e−09	3.9e−09
		18	6.0e−10	9.6e−10
5	1e−08	12	1.7e−06	1.7e−06
		13	4.3e−07	4.2e−07
		14	1.3e−07	9.7e−08
10	1e−08	12	2.3e−06	2.2e−06
		13	6.1e−07	5.3e−07
		14	1.9e−07	1.1e−07
50	1e−08	12	4.6e−06	4.6e−06
		13	1.2e−06	1.0e−06
		14	4.0e−07	3.2e−07

Table C.6: The Precision of the convolution of exponential distributions via FFT; confer Example C.3.5.

Appendix D

Optimally Robust Estimation via S4 Classes and Methods

Our R bundle RobASt (version 0.3-9) consists of about 18.000 lines of R code and the converted R help files build a pdf-manual of 291 pages. Hence, we can not give a full description of bundle RobASt in this appendix. We rather explain the main designing principles and give an overview about the implemented S4 classes and methods. For more details about the full functionality of our R packages we refer to the corresponding source code and help pages, respectively. A short description of the R packages RobLox and RobRex which are also part of bundle RobASt can be found in Section 8.9 and Subsection 7.6.2, respectively.
Since there are many new S4 methods, the loading time for the two main packages ROptEst and ROptRegTS is quite long. In Table D.1 the loading times on an AMD Athlon with 2.5 GHz and 512 MB RAM using R 2.0.1 on Windows 2000 are given where the additionally required packages were already loaded.

Package	**Loading Time in sec.**
distrEx	1.66
RandVar	5.92
ROptEst	10.14
RobLox	0.06
ROptRegTS	18.41
RobRex	0.06

Table D.1: Loading times for the R packages included in R bundle RobASt.

D.1 R Package distrEx

This package extends the functionality of our R package distr (cf. Ruckdeschel et al. (2005)) which is on CRAN since April 2004.

D.1.1 S4 Classes

The first extensions are S4 classes for multivariate distributions as well as univariate conditional distributions; confer Figure D.1. In case of discrete multivariate distributions (DiscreteMVDistribution) and discrete conditional distributions (DiscreteCondDistribution) we add the slot support as in case of class DiscreteDistribution of package distr. In the univariate cases support is of class numeric whereas in the multivariate case support is of class matrix. In case of univariate conditional distributions (UnivariateCondDistribution) we additionally introduce the slot cond which includes an object of the new S4 class Condition. So far, Condition has the subclass EuclCondition which enables us to define univariate conditional distributions where the condition is some Euclidean space.
In addition, we define S4 classes for vectors of distributions which are DistrList for general distributions and UnivarDistrList for univariate distributions. It is also possible to coerce objects of class Distribution and UnivariateDistribution to objects of class DistrList and UnivarDistrList, respectively.

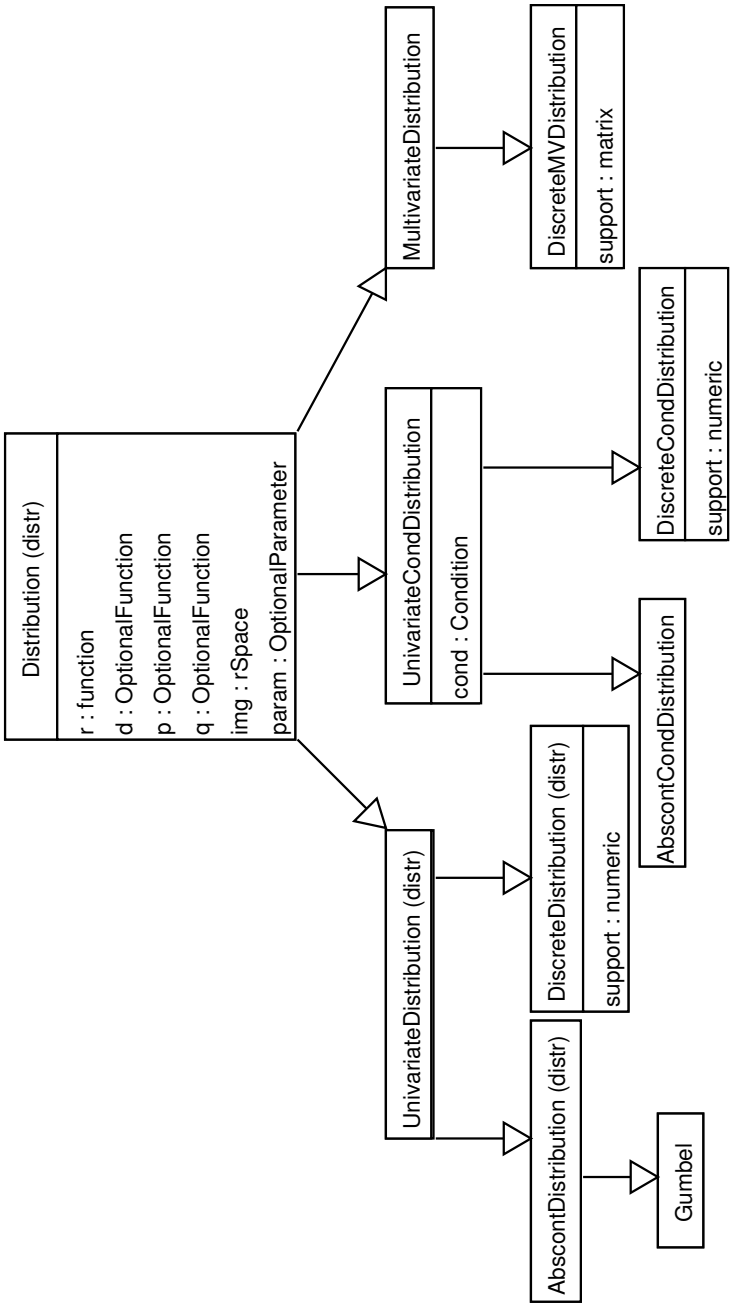

Figure D.1: Class `Distribution` and subclasses (subclasses of `AbscontDistribution` and `DiscreteDistribution` contained in package `distr` are omitted.)

D.1.2 Functions and Methods

Following Section 7.3 of Chambers (1998) we introduce generating functions which in most cases have the same names as the classes. These functions help the user to create objects of the corresponding classes. In package `distrEx` there are the generating functions given in Table D.2.

Generating Function	Description
`DiscreteDistribution`	Generates an object of class `DiscreteDistribution` (via some support vector and corresponding probabilities)
`DiscreteMVDistribution`	Generates an object of class `DiscreteMVDistribution` (via some support matrix and corresponding probabilities)
`DistrList`	Generates an object of class `DistrList`
`EuclCondition`	Generates an object of class `EuclCondition`
`EuclideanSpace`	Generates an object of class `EuclideanSpace` (part of package `distr`)
`Gumbel`	Generates an object of class `Gumbel` (Gumbel distribution)
`LMCondDistribution`	Generates an object of class `AbscontCondDistribution` (conditional distribution of a linear regression model given the regressor)
`LMParameter`	Generates an object of class `LMParameter` (parameter of a linear regression model)
`Naturals`	Generates an object of class `Naturals` (part of package `distr`)
`Reals`	Generates an object of class `Reals` (part of package `distr`)
`UnivarDistrList`	Generates an object of class `UnivarDistrList`

Table D.2: Generating functions of package `distrEx`.

Moreover, following Gentleman (2003) one should not use the `@`-Operator to inspect, respectively modify slots as this breaks the data abstraction and relies on the implementation. To avoid this, one should rather implement accessor and replacement functions for the slots of `S4` classes. We follow this advice and define accessor functions for all slots of newly introduced classes. We also implement replacement functions not for all but for those slots a user should modify. Consequently, we strongly discourage the use of the `@`-operator to inspect or even modify slots.

In addition, we implement methods for the generic function `show` for the new `S4` classes `DistrList`, `EuclCondition`, `LMParameter`, `MultivariateDistribution` and `UnivariateCondDistribution` as well as methods for the generic function `plot`

for DistrList, MultivariateDistribution and UnivariateCondDistribution. In case of MultivariateDistribution and UnivariateCondDistribution the methods for the generic function plot are defined but not yet implemented.

Besides the accessor and replacement functions mentioned above, we introduce further new generic functions and corresponding S4 methods. In Table D.3 we give a short description of these new generic functions.

Generic Function	Description
ContaminationSize	computation of convex contamination (pseudo-)distance of two probability distributions P and Q
ConvexContamination	generation of convex contaminations
E	computation of (conditional) expectations
HellingerDist	computation of the Hellinger distance of two distributions
KolmogorovDist	computation of the Kolmogorov distance of two distributions
liesInSupport	testing the support of a distribution
m1df	computation of clipped first moments
m2df	computation of clipped second moments
TotalVarDist	computation of the total variation distance of two distributions

Table D.3: New generic functions of package distrEx (without accessor and replacement functions).

The generic functions E, m1df and m2df are essential for the computation of optimal ICs as implemented in packages ROptEst and ROptRegTS. Given some object D of class Distribution, and some function f the expectation of D, respectively of f under D can easily be computed via

```
> E(object = D)
```

respectively

```
> E(object = D, fun = f)
```

Furthermore, given some object CD of class UnivariateCondDistribution we can compute the conditional distribution of CD given some condition C, respectively of f under CD via

```
> E(object = CD, cond = C)
```

respectively

```
> E(object = CD, fun = f, cond = C)
> E(object = CD, fun = f, cond = C, withCond = TRUE)
```

If function `f` also depends on the condition `C`, this can be taken into account by setting the parameter `withCond` to `TRUE`. Finally, given some bound `b` we can obtain the clipped first, respectively second moment of a univariate distribution `D` by just calling

```
> m1df(object = D, upper = b)
```

respectively

```
> m2df(object = D, upper = b)
```

By means of these functions together with the distribution classes provided by packages `distr` and `distrEx` we are able to implement one algorithm which works for a whole class of distributions.
For more details about the full functionality of package `distrEx` we refer to the source code and the corresponding help pages, respectively.

D.1.3 Odds and Ends

The next steps which should be done for package `distrEx` are:

- further functionals for distributions like variance, median, MAD, IQR, skewness or kurtosis
- further distances between distributions like Cramér von Mises, Lévy or Prokhorov distance
- further generating functions
- methods for the generic function `plot` for classes `MultivariateDistribution` and `UnivariateCondDistribution`
- `S4` classes for multivariate distributions like multivariate normal or multivariate T distribution; confer package `mvtnorm` of Genz et al. (2004)
- increase computation speed for expectations

So far, the `R` function `apply`, respectively `sapply` (cf. R Development Core Team (2005)) is called in methods for the generic function `E`. To avoid this, a further parameter `useApply` shall be introduced which allows to control the use of `apply`, respectively `sapply`.

D.2 R Package RandVar

In this package we implement random variables by means of S4 classes and methods.

D.2.1 S4 Classes

The S4 class RandVariable (cf. Figure D.2) has the slots Map, Domain and Range where Map contains a list of functions which are measurable maps from Domain to Range. The elements contained in the list Map must be functions in one argument named x. We do not allow further parameters for these functions as this would lead to inconsistent objects. Strictly speaking, an object of class RandVariable would represent not only one random variable but a whole set of random variables depending on these parameters.

The slots Domain and Range are filled with an object of class OptionalrSpace; i.e., they contain NULL or an object of class rSpace (see package distr). In case of EuclRandVariable and RealRandVariable the slot Range is filled with an object of class Euclideanspace and Reals, respectively. The class EuclRandMatrix additionally has the slot Dim which is a vector of integers and contains the dimensions of the Euclidean random matrix.

Using these S4 classes there are two possibilities to implement a $\mathbb{R}^k$ valued random variable. First, we could define a EuclRandVariable whose slot Map contains a list with one function which maps to $\mathbb{R}^k$; i.e., the slot Range is a k-dimensional Euclidean space. Second, we could define a EuclRandVariable whose slot Map contains a list with n functions (projections) which map to $\mathbb{R}^m$ where $k = m*n$. Now, the slot Range is an m-dimensional Euclidean space. Since it is sometimes convenient to regard a $\mathbb{R}^k$ valued random variable as measurable map consisting of $\mathbb{R}^{k_i}$ valued maps where $\sum k_i = k$, we introduce a further class called EuclRandVarList. With this class we can now define $\mathbb{R}^k$ valued random variables as a list of $\mathbb{R}^{k_i}$ valued random variables with compatible domains. More precisely, the elements of a EuclRandVarList may even have very different ranges (not necessarily Euclidean spaces) they only need to have compatible domains which is checked via the generic function compatibleDomains.

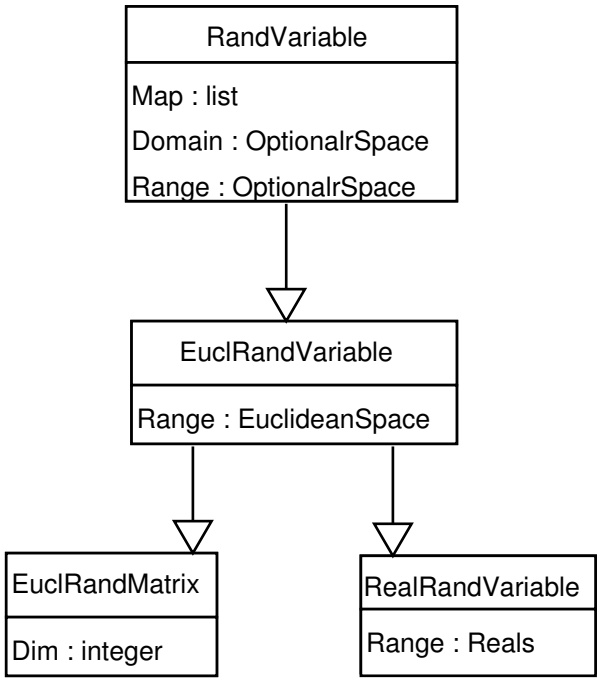

Figure D.2: Class `RandVariable` and subclasses.

D.2.2 Functions and Methods

As in case of package `distrEx` (cf. Subsection D.1.2), we follow the advices of Section 7.3 of Chambers (1998) and Gentleman (2003). That is, we introduce generating functions as well as accessor and replacement functions. A short description of the implemented generating functions is given in Table D.4.

Generating Function	**Description**
`EuclRandMatrix`	Generates an object of class `EuclRandMatrix`
`EuclRandVariable`	Generates an object of class `EuclRandVariable`
`EuclRandVarList`	Generates an object of class `EuclRandVarList`
`RandVariable`	Generates an object of class `RandVariable`
`RealRandVariable`	Generates an object of class `RealRandVariable`

Table D.4: Generating functions of package `RandVar`.

While there are accessor functions for all slots of the newly defined `S4` classes, replacement functions are only implemented for those slots a user should modify.

Our next goal was that one can use these classes of random variables like ordinary numeric vectors or matrices. Hence, we overloaded the S4 group generic functions Arith and Math as well as matrix multiplication %*%. For the matrix multiplication of EuclRandVarLists we additionally introduced the operator %m%. Now, if we have random variables X and Y, a numerical vector v and a numerical matrix M (with compatible dimensions) we can for instance generate

```
> exp(X-v)
> X %*% Y
```

or

```
> M %*% X
```

We also implemented S4 methods for the generic function E of package distrEx. That is, given some distribution D, respectively some conditional distribution CD and some random variable X we can compute the (conditional) expectation of X under D, respectively CD simply by

```
> E(object = D, fun = X)
> E(object = D, fun = X, cond = C)
```

where C is some given condition.

In addition, we define methods for the generic function show for the classes RandVariable, EuclRandMatrix and EuclRandVarList. There are also methods for the generic functions dimension (see package distr), length, ncol, nrow, t and [(cf. package base of R Development Core Team (2005)). For more details we refer to the corresponding help pages.

Finally, we introduce several new generic functions. A brief description of these functions is given in Table D.5.

Generic Function	**Description**
%m%	matrix multiplication for EuclRandVarLists
compatibleDomains	test if the domains of two random variables are compatible
evalRandVar	evaluation of random variables
imageDistr	image distribution of some distribution under some random variable
numberOfMaps	number of functions contained in the slots Map of the members of a EuclRandVarList

Table D.5: New generic functions of package RandVar (without accessor and replacement functions).

For more details about the full functionality of package RandVar we refer to the source code and the corresponding help pages, respectively.

D.2.3 Odds and Ends

The main issue is to reduce the computation time for methods using objects of class `RandVariable` and its subclasses as these classes play an important role in the computation of optimally robust estimators; confer Sections D.3 and D.4. In particular, we are looking for ways to increase the computation speed of `evalRandVar` and `E`.

D.3 `R` Package `ROptEst`

The aim of package `ROptEst` is to provide methods for the computation of optimally robust estimators for various models and risks.

D.3.1 `S4` Classes

Ideal Model

We define the `S4` class `ProbFamily` which stands for some family of probability measures and derive subclasses for parametric, respectively L_2 differentiable parametric families; confer Figure D.4. Special properties and additional (meta-)information can be noted in the slot `props` of these classes.

Semi-Symbolic Calculus

Since invariance properties like symmetry in many cases lead to simpler algorithms, we define the `S4` class `Symmetry` and corresponding subclasses; confer Figure D.3. With these classes we introduce what we call *semi-symbolic calculus*. This can be explained as follows:
Assume we know about some invariance property (e.g. symmetry) of our model which cannot be decided (exactly) on basis of numeric evaluations. However, as a logical statement one can use this property and even deduce further invariances.
For example, with the help of the `S4` class `Symmetry` and its subclasses, which can also be put together to lists as `FunSymmList`s or `DistrSymmList`s, we can obtain simpler/faster algorithms (cf. Algorithm D.3.1) or simplify the computation of high dimensional integrals.
Second, in classical (linear) hierarchical inheritance relations it is not always clear in which order one should inherit certain properties (e.g. symmetry, boundedness, ...). Should one for instance derive bounded functions as subclasses of symmetric functions or the other way round? We bypass this problem by using symbolic flags like `Symmetry` as slots of larger `S4` classes instead of defining new classes. With this approach we are able to take these properties into account and to interfere into the dispatching mechanism.

Robust Model

To obtain a fixed (contamination of fixed radius), respectively infinitesimal (radius shrinking at rate of $\sqrt{n}$) robust model (cf. Figure D.6) an object of class `ParamFamily`, respectively `L2ParamFamily` is equipped with an unconditional neighborhood (cf. Figure D.5).

Risk

To determine some optimal (robust) estimator, one has to decide which risk should be used to compare possible estimators. For this purpose we introduce the S4 class RiskType and corresponding subclasses (see Figure D.7) where we distinguish between finite-sample (fiRisk) and asymptotic (asRisk) risks.

Influence Curve

As we are mainly interested in optimally robust estimators which in addition are asymptotically linear with some IC, we finally define the S4 class InfluenceCurve and its subclasses; see Figure D.8. Since objects of class IC are used for various computations we tried to make sure that such objects are not too big. Hence, not the whole L_2 differentiable family is saved in one slot of class IC but only the call to generate this family.

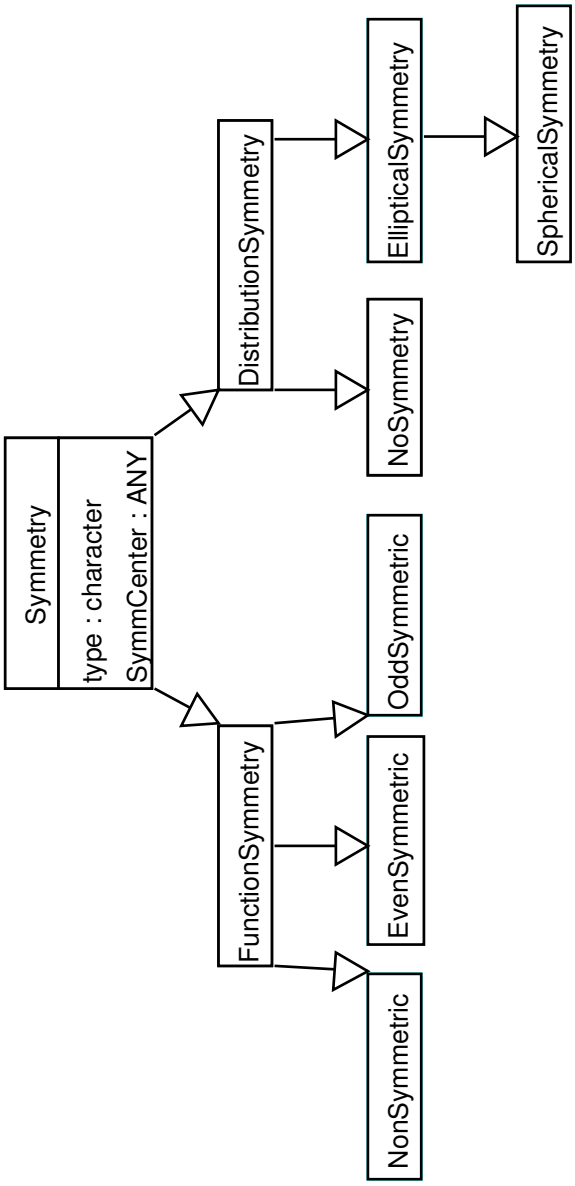

Figure D.3: Class `Symmetry` and subclasses.

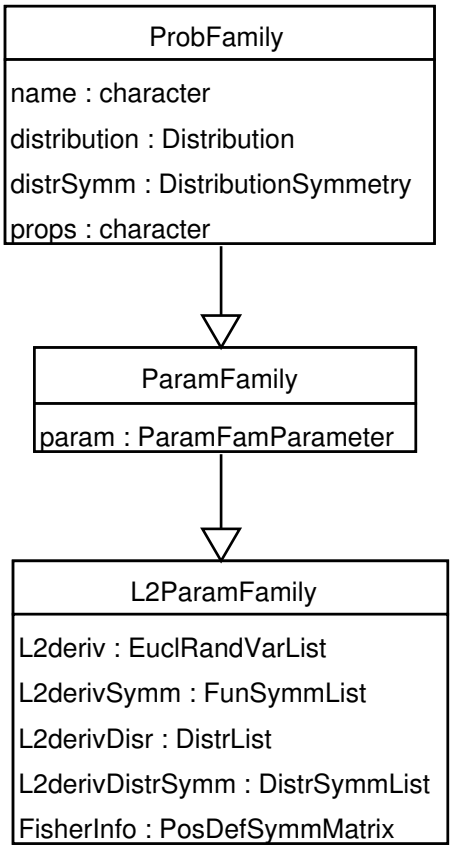

Figure D.4: Class ProbFamily and subclasses.

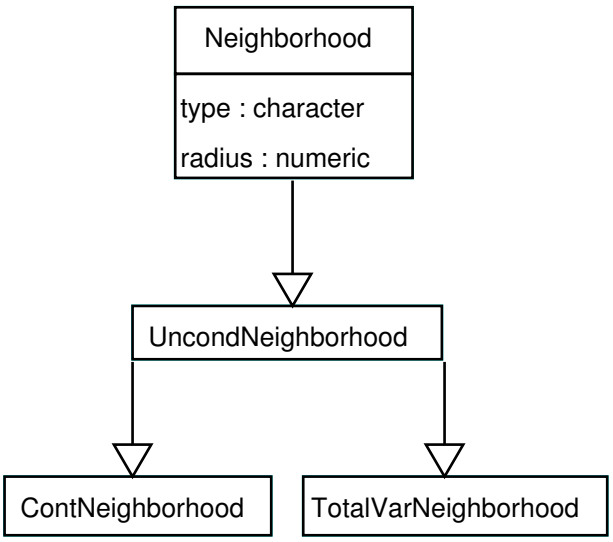

Figure D.5: Class `Neighborhood` and subclasses.

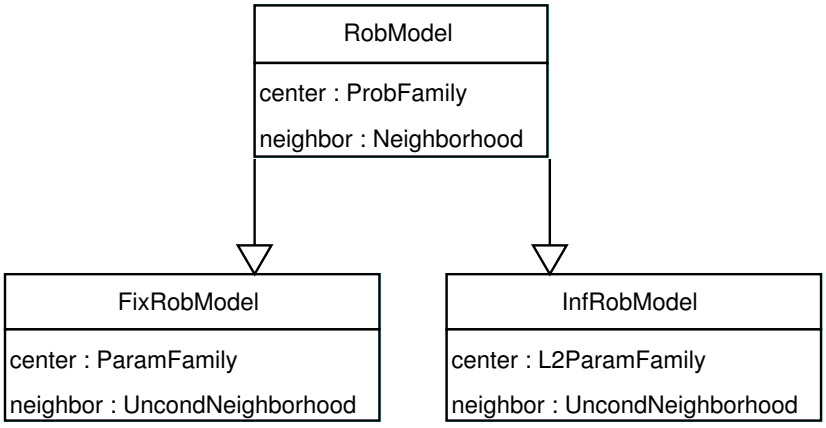

Figure D.6: Class `RobModel` and subclasses.

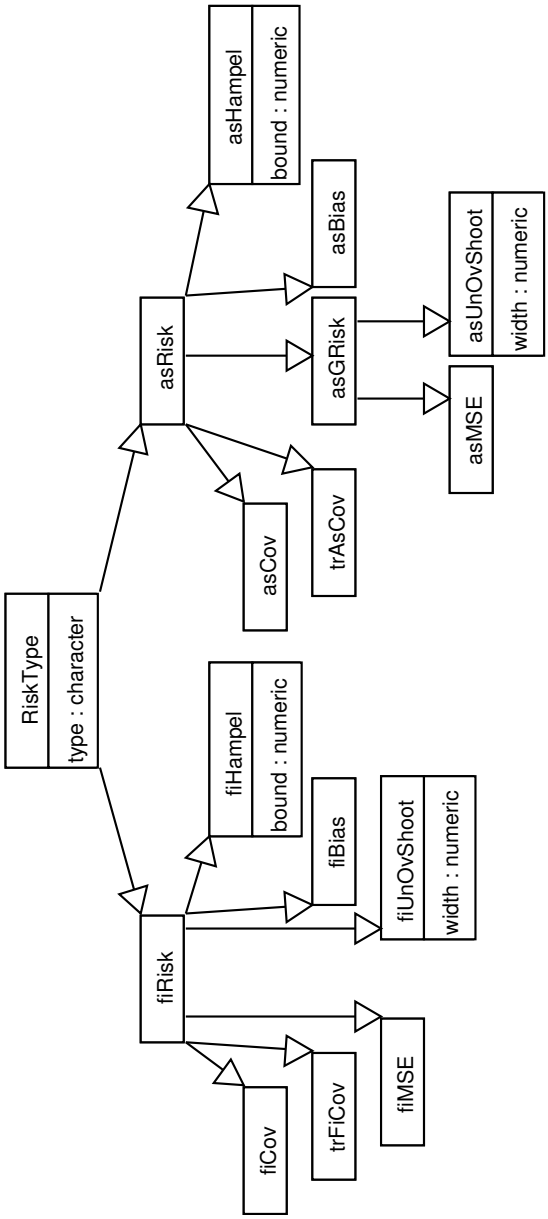

Figure D.7: Class `RiskType` and subclasses.

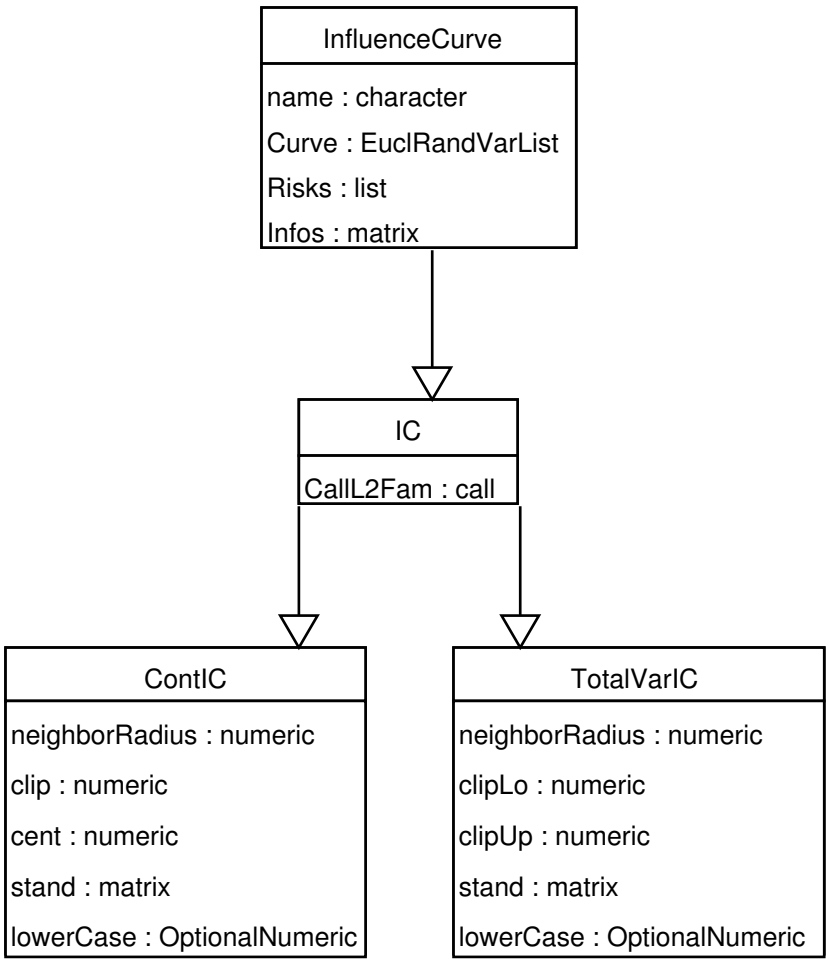

Figure D.8: Class `InfluenceCurve` and subclasses.

D.3.2 Functions and Methods

As in case of distrEx and RandVar we follow the advices of Section 7.3 of Chambers (1998) and Gentleman (2003). That is, we introduce generating functions as well as accessor and replacement functions. A short description of the implemented generating functions is given in Tables D.6 and D.7. While there are accessor functions for all slots of the newly defined S4 classes, replacement functions are only implemented for those slots a user should modify.

Generating Function	Description
asBias	Generates an object of class asBias
asCov	Generates an object of class asCov
asHampel	Generates an object of class asHampel
asMSE	Generates an object of class asMSE
asUnOvShoot	Generates an object of class asUnOvShoot
BinomFamily	Generates an object of class L2ParamFamily which represents a Binomial family
ContIC	Generates an object of class ContIC
ContNeighborhood	Generates an object of class ContNeighborhood
DistrSymmList	Generates an object of class DistrSymmList
EllipticalSymmetry	Generates an object of class EllipticalSymmetry
EvenSymmetric	Generates an object of class EvenSymmetric
ExpScaleFamily	Generates an object of class L2ParamFamily which represents an exponential scale family
fiBias	Generates an object of class fiBias
fiCov	Generates an object of class fiCov
fiHampel	Generates an object of class fiHampel
fiMSE	Generates an object of class fiMSE
fiUnOvShoot	Generates an object of class fiUnOvShoot
FixRobModel	Generates an object of class FixRobModel
FunSymmList	Generates an object of class FunSymmList
GammaFamily	Generates an object of class L2ParamFamily which represents a Gamma family
GumbelLocationFamily	Generates an object of class L2ParamFamily which represents a Gumbel location family
IC	Generates an object of class IC
InfluenceCurve	Generates an object of class InfluenceCurve
InfRobModel	Generates an object of class InfRobModel
L2ParamFamily	Generates an object of class L2ParamFamily
LnormScaleFamily	Generates an object of class L2ParamFamily which represents a lognormal scale family

Table D.6: Generating functions of package ROptEst (part 1).

Generating Function	Description
NonSymmetric	Generates an object of class NonSymmetric
NormLocationFamily	Generates an object of class L2ParamFamily which represents a normal location family
NormLocationScaleFamily	Generates an object of class L2ParamFamily which represents a normal location and scale family
NormScaleFamily	Generates an object of class L2ParamFamily which represents a normal scale family
NoSymmetry	Generates an object of class NoSymmetry
OddSymmetric	Generates an object of class OddSymmetric
ParamFamily	Generates an object of class ParamFamily
ParamFamParameter	Generates an object of class ParamFamily
PoisFamily	Generates an object of class L2ParamFamily which represents a Poisson family
PosDefSymmMatrix	Generates an object of class PosDefSymmMatrix
SphericalSymmetry	Generates an object of class SphericalSymmetry
TotalVarIC	Generates an object of class TotalVarIC
TotalVarNeighborhood	Generates an object of class TotalVarNeighborhood
trAsCov	Generates an object of class trAsCov
trFiCov	Generates an object of class trFiCov

Table D.7: Generating functions of package ROptEst (part 2).

If one wants to estimate the parameter of an L_2 differentiable parametric family which does not have its own generating function, one can use the generating function L2ParamFamily. To demonstrate the use of L2ParamFamily let us now consider the Poisson family censored at 0; i.e., we observe $X \sim \mathrm{Pois}(\lambda)$ only if $X \neq 0$. Given the parameter lambda we first define the corresponding distribution

```
> supp <- support(Pois(lambda))
> supp <- supp[supp > 0]
> p0 <- exp(-lambda)
> PC <-DiscreteDistribution(supp = supp,
+                  prob = dpois(supp, lambda = lambda)/(1-p0))
```

Second, we derive and define the L_2 derivative of this family,

```
> fun <- function(x){ x/lambda - 1/(1-exp(-lambda)) }
> L2PC <- RealRandVariable(list(fun), Domain = Reals())
```

Now, we can instantiate the corresponding object of class L2ParamFamily via

```
> PCFam <- L2ParamFamily(name = "censored Poisson family",
+                        distribution = PC, main = lambda,
+                        L2deriv = EuclRandVarList(L2PC))
```

Of course, one can also implement a new generating function for this family. In particular, if the considered L_2 differentiable parametric family is more complicated we recommend to do so.

To derive some optimal IC one can call the S4 generic function optIC with the given model and the chosen risk. So far, there are methods for the signatures given in Table D.8 where asRisk up to now may be replaced by asCov, asHampel, asMSE or asBias. In any case optIC returns an object of class InfluenceCurve; confer Figure D.8.

model	risk
L2ParamFamily	asCov
InfRobModel	asRisk
InfRobModel	asUnOvShoot
FixRobModel	fiUnOvShoot

Table D.8: Methods for generic function optIC in package ROptEst.

For the computation of optimally robust ICs the generic functions given in Table D.9 are used. These functions are rarely called directly.

Generic Function	**Description**
generateIC	generation of influence curve
getAsRisk	computation of asymptotic risk
getFiRisk	computation of finite-sample risk
getFixClip	computation of optimal clipping bound
getFixRobIC	computation of optimally robust IC
getInfCent	computation of optimal centering constant, resp. optimal lower clipping bound
getInfClip	computation of optimal clipping bound
getInfGamma	computation of optimal clipping bound
getInfRobIC	computation of optimally robust IC
getInfStand	computation of standardizing matrix

Table D.9: Generic functions for the computation of optimal (robust) ICs in package ROptEst.

To demonstrate how these function work together, we now state the algorithm which is used to compute the optimally robust IC for an object of class InfRobModel and a convex asymptotic risk (asGRisk). These asymptotic risks correspond to the

class of risks covered by Ruckdeschel and Rieder (2004). We state the algorithm for a k-dimensional ($k > 1$) L_2 differentiable parametric family and unconditional contamination neighborhoods.

Algorithm D.3.1
The methods for `optIC` may be regarded as a user interface. Most of the computations are done in the corresponding method for `getInfRobIC`. In the considered setup the algorithm is mainly based on the slots `distribution` (here: `Distr`) and `L2deriv` of the considered L_2 differentiable parametric family.

Step 1: (Radius Check) If the specified radius is equal or smaller than the bound `.Machine$double.eps^0.5`, it is regarded as 0 and the (classical) optimal IC in sense of the Cramér-Rao bound is returned.

Step 2: (Symmetry Check) It is checked whether there are certain symmetries which simplify the computation of the optimal centering constant z or of the standardizing matrix A, respectively.

Step 3: (Clipping) The optimal clipping bound b is computed via

```
> b <- try(uniroot(getInfClip, lower=.Machine$double.eps^0.75,
+                  upper = upper, tol = tol, L2deriv = L2deriv,
+                  risk = risk, Distr = Distr,
+                  neighbor = neighbor, stand = A, cent = z,
+                  trafo = trafo)$root, silent = TRUE)
```

where `lower`, `upper` and `tol` are parameters of `uniroot`. The method for function `getInfClip` calls `getInfGamma` via

```
> return(neighbor@radius^2*clip +
+        getInfGamma(L2deriv = L2deriv, risk = risk,
+                neighbor = neighbor, Distr = Distr,
+                stand = stand, cent = cent, clip = clip))
```

If `uniroot` fails, this might by caused by the fact that the considered radius is larger than the maximum radius which is admitted for the given risk. Hence, in this case the minimum asymptotic bias solution is returned.

Step 4: (Centering) The optimal centering constant z is computed via

```
> z <- getInfCent(L2deriv = L2deriv, neighbor = neighbor,
+                 Distr = Distr, z.comp = z.comp, stand = A,
+                 cent = z, clip = b)
```

where `z.comp` tells `getInfCent` which components of z have to be computed; i.e., are not equal to 0. For instance, if the distribution of the L_2 derivative is spherical symmetric a possible choice for the optimal centering constant is 0; i.e., in this case nothing has to be computed.

Step 5: (Standardization) The standardizing matrix A is computed via

```
> A <- getInfStand(L2deriv = L2deriv, neighbor = neighbor,
+                  Distr = Distr, A.comp = A.comp, stand = A,
+                  clip = b, cent = z, trafo = trafo)
```

Now, `A.comp` tells `getInfStand` which components of A have to be computed. For instance, if the distribution of the L_2 derivative is elliptical symmetric this leads to a standardizing matrix of diagonal form. Moreover, if it is even spherical symmetric we get $A = \alpha\mathbb{I}$ with $\alpha \in \mathbb{R}$.

The Steps 3–5 are repeated until convergence.

Step 6: (Risk) The asymptotic risk is computed via

```
> Risk <- getAsRisk(risk = risk, L2deriv = L2deriv,
+                   neighbor = neighbor, clip = b, cent = a,
+                   stand = A, trafo = trafo)
```

The computed results are returned to `optIC`.

Step 7: (Generation of IC) The optimally robust IC is generated and returned via

```
> return(generateIC(model@neighbor, model@center, res))
```

where `res` is a list containing the results computed by `getInfRobIC`.

Remark D.3.2 (a) The previous algorithm works as proposed in Remark 5.5.2 of Rieder (1994). That is, the centering constant z and the standardizing matrix A are determined iteratively.

(b) Algorithm D.3.1 works for all (!) k-dimensional ($k > 1$) L_2 differentiable parametric families which are based on a univariate distribution. This great generality is made possible by the use of distribution and random variable objects from packages `distr`, `distrEx` and `RandVar` in combination with the generic function `E` of package `distrEx`.

(c) In case of an one-dimensional parametric family the algorithm directly uses the distribution of the L_2 derivative and makes use of the functions `m1df` and `m2df` of package `distrEx`. In addition, only the Steps 3 and 4 have to be repeated until convergence as the standardizing matrix (here: $A \in \mathbb{R}$) can be computed afterwards; confer Chapters 3–5. Moreover, we use the `R` function `uniroot` (cf. R Development Core Team (2005)) to compute the centering constant z numerically exact in every iteration step.

(d) In case of unconditional total variation neighborhoods the algorithm is very similar. The only difference is the computation of the optimal lower clipping bound in Step 4 instead of the optimal centering constant. As it turns out, it is numerically more stable to use

$$r^2 b = \mathrm{E}\left(A\Lambda - (c+b)\right)_+ \tag{D.3.1}$$

instead of (1.3.46) in Step 3 of Algorithm D.3.1.

(e) If we consider the under-/overshoot confidence risk treated in Part V of this thesis, the algorithm again is analogous to Algorithm D.3.1 with the modifications mentioned in parts (c) and (d) of this remark. ////

After the computation of some optimally robust IC the corresponding optimally robust estimator has to be constructed. In case of a location model we provide the generic function `locMEstimator` which determines the corresponding optimally robust estimator by solving an M equation. More generally, the optimally robust estimator can be constructed by means of one-step construction. For this purpose we introduce the generic function `oneStepEstimator`. In many parametric models the Kolmogorov(–Smirnov) minimum distance estimator may serve as initial estimator which is implemented via the generic function `ksEstimator`.

Besides the already mentioned generic functions and methods, we define methods for the generic functions `plot` for classes `ParamFamily`, `L2ParamFamily` and `IC`. Moreover, we define methods for the generic function `show` for classes `Symmetry`, `ParamFamParameter`, `ParamFamily`, `Neighborhood`, `FixRobModel`, `InfRobModel`, `RiskType`, `asUnOvShoot`, `asHampel`, `fiUnOvShoot`, `fiHampel`, `InfluenceCurve`, `IC`, `ContIC` and `TotalVarIC`. Finally, we introduce several new generic functions. A brief description of these functions is given in Table D.10.

Generic Function	**Description**
`addInfo`	add information to slot `Infos` of some IC
`addProp`	add property to slot `props` of a family of probability measures
`addRisk`	add risk to slot `Risks` of some IC
`checkIC`	check centering and Fisher consistency of some IC
`checkL2deriv`	check L_2 derivative of an L_2-differentiable family of probability measures
`evalIC`	evaluation of ICs
`getIneffDiff`	computation of inefficiency differences
`getRiskIC`	computation of a risk for an IC
`infoPlot`	plot absolute and relative information of ICs
`ksEstimator`	computation of the Kolmogorov(–Smirnov) minimum distance estimator
`leastFavorableRadius`	computation of least favorable radii
`locMEstimator`	computation of location M estimators
`lowerCaseRadius`	computation of lower case radii
`oneStepEstimator`	computation of one-step estimators
`optRisk`	computation of optimal (i.e., minimal) risks
`radiusMinimaxIC`	computation of radius–minimax ICs

Table D.10: Further new generic functions in package `ROptEst` (without accessor and replacement functions).

For more details about the full functionality of package `ROptEst` we refer to Sections 3.5, 4.5, 5.5, 6.5, 8.9 and 11.5 as well as to the source code and the corresponding help pages.

D.3.3 Odds and Ends

There are many things that remain to be done. The most important are:

- lower case for dimension $k > 1$
- methods for further asymptotic and finite-sample risks
- generating functions for multivariate L_2 differentiable parametric families
- methods for L_2 differentiable parametric models which are based on multivariate distributions
- take care of further invariance properties (e.g., group models)
- use package `Matrix` (cf. Bates and Maechler (2005))

D.4 R Package ROptRegTS

This package is designed to provide optimally robust estimators for regression and time series models.

D.4.1 S4 Classes

Starting form the S4 class `ParamFamily` of package `ROptEst` we define the S4 classes `RegTypeFamily` and `L2RegTypeFamily` (cf. Figure D.9) which shall represent (L_2 differentiable) regression-type families as given in Appendix A. Again, we introduce several slots where symmetry properties can be saved. We also extend the classes `Neighborhood`, `RobModel` and `IC` of package `ROptEst` as given in Figures D.10–D.12.

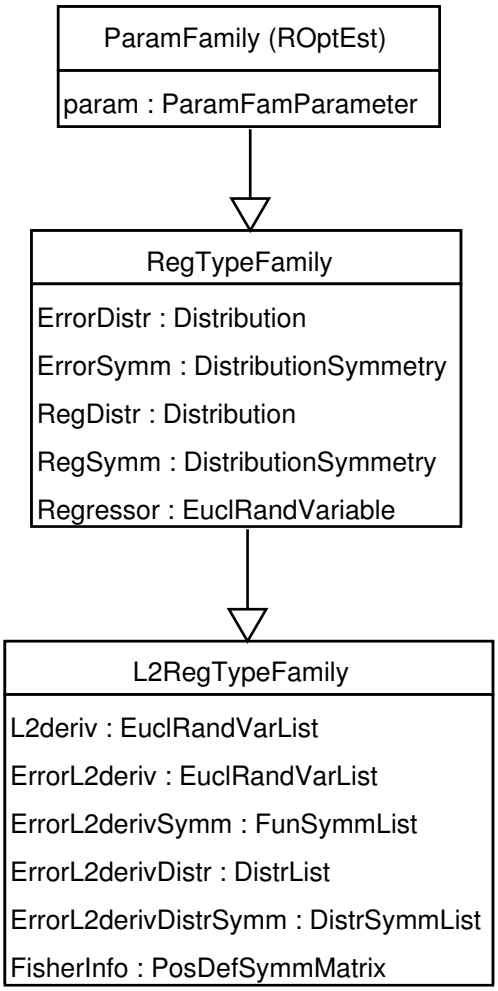

Figure D.9: Class `RegTypeFamily` and subclass `L2RegTypeFamily`.

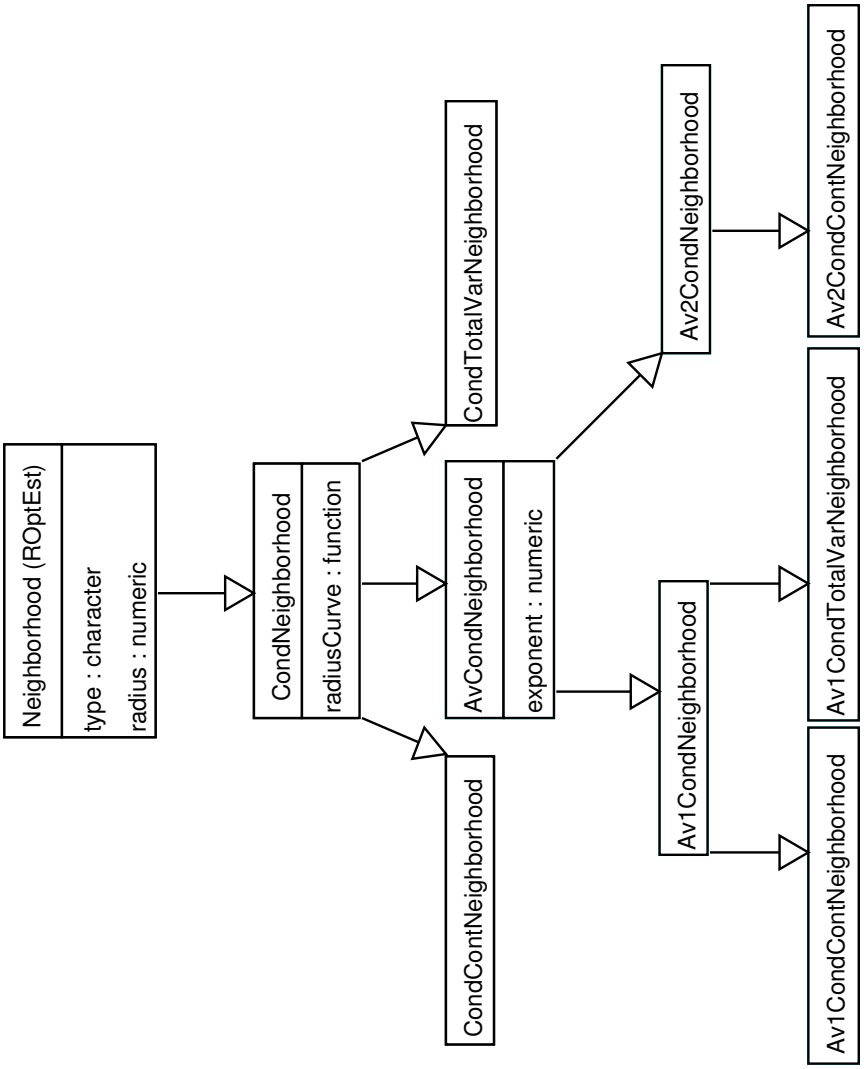

Figure D.10: Class `CondNeighborhood` and subclasses.

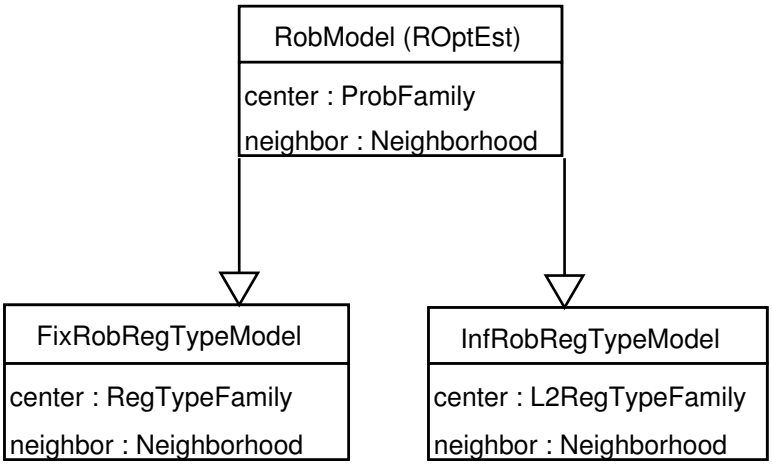

Figure D.11: Classes `FixRobRegTypeModel` and `InfRobRegTypeModel`.

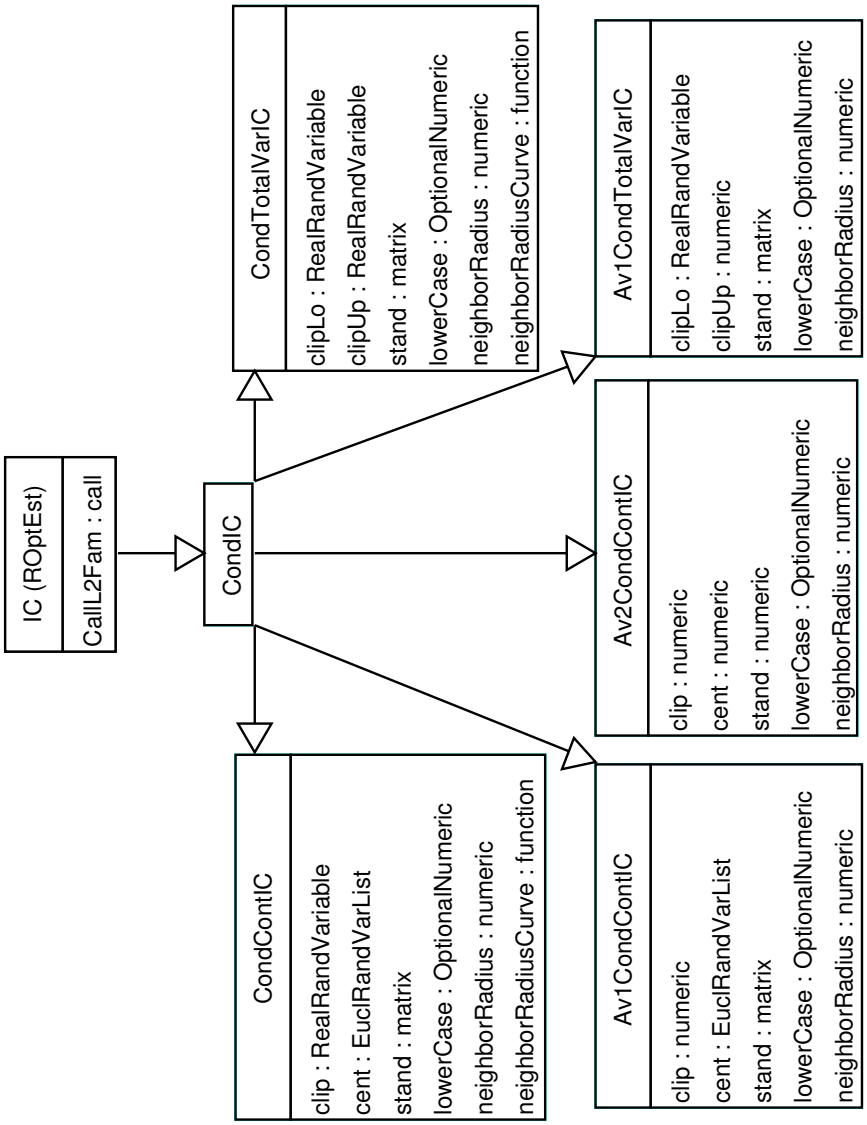

Figure D.12: Class `CondIC` and subclasses.

D.4.2 Functions and Methods

We again follow the advices of Section 7.3 of Chambers (1998) and Gentleman (2003); i.e., we introduce generating functions as well as accessor and replacement functions for the newly defined classes. A brief description of the implemented generating functions is given in Tables D.11 and D.12. While there are accessor functions for all slots of the newly defined `S4` classes, replacement functions are again only implemented for those slots a user should modify.

Generating Function	**Description**
`Av1CondContIC`	Generates an object of class `Av1CondContIC`
`Av1CondContNeighborhood`	Generates an object of class `Av1CondContNeighborhood`
`Av1CondTotalVarIC`	Generates an object of class `Av1CondTotalVarIC`
`Av1CondTotalVarNeighborhood`	Generates an object of class `Av1CondTotalVarNeighborhood`
`Av2CondContIC`	Generates an object of class `Av2CondContIC`
`Av2CondContNeighborhood`	Generates an object of class `Av2CondContNeighborhood`
`CondContIC`	Generates an object of class `CondContIC`
`CondContNeighborhood`	Generates an object of class `CondContNeighborhood`
`CondIC`	Generates an object of class `CondIC`
`CondTotalVarIC`	Generates an object of class `CondTotalVarIC`
`CondTotalVarNeighborhood`	Generates an object of class `CondTotalVarNeighborhood`
`FixRobRegTypeModel`	Generates an object of class `FixRobRegTypeModel`
`InfRobRegTypeModel`	Generates an object of class `InfRobRegTypeModel`
`L2RegTypeFamily`	Generates an object of class `L2RegTypeFamily`
`NormLinRegFamily`	Generates an object of class `L2RegTypeFamily` which represents a linear regression family with normal errors and random regressor

Table D.11: Generating functions of package `ROptRegTS` (part 1).

Generating Function	Description
NormLinRegInterceptFamily	Generates an object of class L2RegTypeFamily which represents a linear regression family with normal errors, random regressor and unknown intercept
NormLinRegScaleFamily	Generates an object of class L2RegTypeFamily which represents a linear regression family with normal errors, random regressor and unknown error scale
RegTypeFamily	Generates an object of class RegTypeFamily

Table D.12: Generating functions of package ROptRegTS (part 2).

To compute some optimal IC one can call the S4 generic function optIC with the given regression-type model and the chosen risk. So far, there are methods for the signatures given in Table D.13 where asRisk up to now may be replaced by asCov, asMSE or asBias.

model	risk
L2RegTypeFamily	asCov
InfRobRegTypeModel	asRisk
InfRobRegTypeModel	asUnOvShoot
FixRobRegTypeModel	fiUnOvShoot

Table D.13: Methods for generic function optIC in package ROptst.

Analogously to package ROptEst (cf. Subsection D.3.2) the generic functions given in Table D.14 are used to compute optimally robust ICs. These functions are rarely called directly.

Remark D.4.1 The algorithm used to compute optimally robust ICs consists of the same steps as Algorithm D.3.1. But, depending on the conditional neighborhood, one now has to determine an optimal centering or clipping function, respectively. For more details we refer to the source code, respectively the help pages of the corresponding methods. ////

After the computation of some optimally robust IC the corresponding optimally robust estimator has to be constructed. One possibility is to use one-step constructions which can be computed via the generic function oneStepEstimator. However, up to now, we have not implemented appropriate initial estimators. In case of linear regression, robust estimators like M estimators or the least trimmed square estimator could serve as starting estimators. These estimators can for in-

Generic Function	Description
`generateIC`	generation of influence curve
`getAsRiskRegTS`	computation of asymptotic risk
`getFiRiskRegTS`	computation of finite-sample risk
`getFixClipRegTS`	computation of optimal clipping bound/function
`getFixRobRegTypeIC`	computation of optimally robust IC
`getInfCentRegTS`	computation of optimal centering constant/function, resp. optimal lower clipping bound/function
`getInfClipRegTS`	computation of optimal clipping bound/function
`getInfGammaRegTS`	computation of optimal clipping bound/function
`getInfRobRegTypeIC`	computation of optimally robust IC
`getInfStandRegTS`	computation of standardizing matrix

Table D.14: Generic functions for the computation of optimal (robust) ICs in package `ROptRegTS`.

stance be computed via functions `rlm` and `ltsreg` of package `MASS`; confer Venables and Ripley (2002).
Besides the already mentioned generic functions and methods, we define methods for the generic functions `show` for classes `RegTypeFamily`, `CondNeighborhood`, `AvCondNeighborhood`, `FixRobRegTypeModel`, `InfRobRegTypeModel`, `CondContIC`, `Av1CondContIC`, `Av2CondContIC`, `CondTotalVarIC` and `Av1CondTotalVarIC`.
For more details about the full functionality of package `ROptRegTS` we refer to Subsections 7.6.1, 9.2.3 and Section 12.4 as well as to the source code and the corresponding help pages.

D.4.3 Odds and Ends

As in case of package `ROptEst` there are many open issues. The most important are:

- introduce generic function `getRiskRegTypeIC` as analogue to `getRiskIC` of package `ROptEst`
- appropriate initial estimators for one-step construction
- generating functions for time series models
- methods for time series models
- methods for further asymptotic and finite-sample risks
- take care of further invariance properties (e.g., group models)
- use package `Matrix` (cf. Bates and Maechler (2005))

Bibliography

Abramowitz M. and Stegun I.A. (Eds.) (1984): *Handbook of mathematical functions with formulas, graphs, and mathematical tables.*. Selected Government Publications. National Bureau of Standards., Washington, D.C. Reprint of the 1972 ed.

Alm S.E. (2003): Monotonicity of the difference between median and mean of gamma distributions and of a related Ramanujan sequence. *Bernoulli*, **9**(2): 351–371.

Andrews D.F., Bickel P.J., Hampel F.R., Huber P.J., Rogers W.H. and Tukey J.W. (1972): *Robust estimates of location. Survey and advances*. Princeton University Press, Princeton, N. J.

Artin E. (1964): *The gamma function*. Athena Series. Selected Topics in Mathematics. New York-Chicago-San Francisco-Toront-London:Holt, Rinehart and Winston. VII, 39 p.

Bates D. and Maechler M. (2005): *Matrix: A Matrix package for R*. R package version 0.96-3.

Bauer H. (1990): *Maß- und Integrationstheorie. (Measure and integration theory)*. Walter de Gruyter.

Beaton A.E. and Tukey J.W. (1974): The fitting of power series, meaning polynomials, illustrated on band- spectroscopic data. Discussions. *Technometrics*, **16**: 147–185.

Bednarski T. and Müller C.H. (2001): Optimal bounded influence regression and scale M-estimators in the context of experimental design. *Statistics*, **35**(4): 349–369.

Bertram J. (1981): Numerische Berechnung von Gesamtschadenverteilungen. *Bl., Dtsch. Ges. Versicherungsmath.*, **15**: 175–194.

Bickel P. (1981): Quelques aspects de la statistique robuste. In: *Ecole d'ete de probabilites de Saint-Flour IX-1979, Lect. Notes Math. 876, 2-72* .

—— (1984): Robust regression based on infinitesimal neighbourhoods. *Ann. Stat.*, **12**: 1349–1368.

Bickel P.J. (1982): On adaptive estimation. *Ann. Statist.*, **10**(3): 647–671.

Bickel P.J., Klaassen C.A., Ritov Y. and Wellner J.A. (1998): *Efficient and adaptive estimation for semiparametric models.*. Springer.

Billingsley P. (1968): *Convergence of probability measures*. John Wiley and Sons, Inc.

Breiman L. (1968): *Probability*. Addison-Wesley Series in Statistics. Addison-Wesley.

Brockwell P.J. and Davis R.A. (1991): *Time series: theory and methods*. Springer Series in Statistics. Springer, 2. edition.

Bühlmann H. (1984): Numerical evaluation of the compound Poisson distribution: Recursion or fast Fourier transform? *Scand. Actuar. J.*, **1984**: 116–126.

Cadigan N. and Chen J. (2001): Properties of robust M-estimators for Poisson and negative binomial data. *J. Stat. Comput. Simulation*, **70**(3): 273–288.

Chambers J. (1998): *Programming with data: a guide to the S language*. New York-Berlin-Heidelberg: Springer.

Choi K. (1994): On the medians of gamma distributions and an equation of Ramanujan. *Proc. Am. Math. Soc.*, **121**(1): 245–251.

Chung K.L. (2000): *A course in probability theory. 3rd ed.*. Orlando, FL: Academic Press. xvi, 419 p. $ 59.95 .

Cooley J. and Tukey J.W. (1965): An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, **19**: 297–301.

Daniels H.E. (1954): Saddlepoint approximations in statistics. *Ann. Math. Statistics*, **25**: 631–650.

Drost F.C., Klaassen C.A. and Werker B.J.M. (1997): Adaptive estimation in time-series models. *Ann. Stat.*, **25**(2): 786–817.

Embrechts P., Grübel R. and Pitts S. (1993): Some applications of the fast Fourier transform algorithm in insurance mathematics. *Stat. Neerl.*, **47**(1): 59–75.

Engle R.F. (1982): Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**: 987–1007.

Feilmeier M. and Bertram J. (1987): *Anwendung numerischer Methoden in der Risikotheorie. (Application of numerical methods in risk theory)*. Schriftenreihe Angewandte Versicherungsmathematik, Heft 16. Deutsche Gesellschaft fr Versicherungsmathematik. Karlsruhe: Verlag Versicherungswirtschaft e.V.

Feller W. (1968): *An introduction to probability theory and its applications. I*. New York-London-Sydney: John Wiley and Sons, Inc. XVIII, 509 p. , 3. edition.

—— (1971): *An introduction to probability theory and its applications. Vol II. 2nd ed.*. Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons, Inc. XXIV, 669 p. , 2. edition.

Field C. and Ronchetti E. (1990): *Small sample asymptotics.*. IMS Lecture Notes - Monograph Series. 13. Hayward, CA: Institute of Mathematical Statistics. 151 p. .

Fraiman R., Yohai V.J. and Zamar R.H. (2001): Optimal robust M-estimates of location. *Ann. Stat.*, **29**(1): 194–223.

Gasquet C. and Witomski P. (1999): *Fourier analysis and applications. Filtering, numerical computation, wavelets. Transl. from the French by R. Ryan*. Texts in Applied Mathematics. 30. New York, NY: Springer. xviii, 442 p.

Gather U. and Schultze V. (1999): Robust estimation of scale of an exponential distribution. *Stat. Neerl.*, **53**(3): 327–341.

Gentleman R. (2003): *Object Orientated Programming. Slides of a Short Course held in Auckland*. http://www.stat.auckland.ac.nz/S-Workshop/Gentleman/Methods.pdf.

Genz A., Bretz F. and port by Torsten Hothorn R. (2004): *mvtnorm: Multivariate Normal and T Distribution*. R package version 0.7-1.

Gilbert P. (2004): *setRNG: Set (normal) random number generator and seed*. R package version 2004.4-1.
URL: *http://www.bank-banque-canada.ca/pgilbert*

Grübel R. (1991): Algorithm AS 265: G/G/1 via fast Fourier transform. *Applied Statistics*, **40**(2): 355–365.

Grübel R. and Hermesmeier R. (1999): Computation of compound distributions. I. Aliasing errors and exponential tilting. *Astin Bull.*, **29**(2): 197–214.

—— (2000): Computation of compound distributions. II. Discretization errors and Richardson extrapolation. *Astin Bull.*, **30**(2): 309–331.

Hajek J. (1972): Local asymptotic minimax and admissibility in estimation. In: *Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970, 1, 175-194*.

Hall P. (1992): *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag.

Hampel F.R. (1968): *Contributions to the theory of robust estimation*. Dissertation, University of California, Berkely, CA.

—— (1974a): Some small sample asymptotics. In: *Proc. Prague Symp. Asympt. Stat., Vol. II, Prague 1973, 109-126* .

—— (1974b): The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.*, **69**: 383–393.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986): *Robust statistics. The approach based on influence functions*. Wiley Series in Probability and Mathematical Statistics. Wiley.

Harville D.A. (1997): *Matrix algebra from a statistician's perspective.*. New York, NY: Springer. xvii, 630 p. DM 108.00; S 788.40; sFr. 98.50.

Hoeffding W. (1963): Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, **58**: 13–30.

Huber P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**: 73–101.

—— (1965): A robust version of the probability ratio test. *Ann. Math. Stat.*, **36**: 1753–1758.

—— (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **10**: 269–278.

—— (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.

—— (1983): Minimax aspects of bounded-influence regression. *J. Am. Stat. Assoc.*, **78**: 66–80.

—— (1997): *Robust statistical procedures. 2 ed.*. CBMS-NSF Regional Conference Series in Applied Mathematics. 68. Philadelphia, PA: SIAM. ix, 67 p.

Huber P.J. and Strassen V. (1973): Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. of Statist.*, **1**: 251–263.

Ibragimov I. (1967): the Chebyshev-Cramér asymptotic expansions. *Theor. Probab. Appl.*, **12**: 454–469.

Johnson N.L., Kotz S. and Balakrishnan N. (1994): *Continuous univariate distributions. Vol. 1.2nd ed.*. Chichester: Wiley,. xix, 756 p.

—— (1995): *Continuous univariate distributions. Vol. 2.2nd ed.*. New York, NY: Wiley. xix, 719 p.

Kohl M., Ruckdeschel P. and Stabla T. (2005): General Purpose Convolution Algorithm for Distributions in S4 Classes by means of FFT.
http://www.uni-bayreuth.de/departments/math/org/mathe7/KOHL/pubs/comp.pdf.

Kreiss J.P. (1987): On adaptive estimation in stationary ARMA processes. *Ann. Stat.*, **15**: 112–133.

Kurotschka V. and Müller C. (1992): Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Stat.*, **20**(1): 331–350.

Le Cam L. (1986): *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer.

Le Cam L. and Lo Yang G. (1990): *Asymptotics in statistics: some basic concepts*. Springer Series in Statistics. Springer.

Lehmann E. and Casella G. (1998): *Theory of point estimation. 2nd ed.*. Springer Texts in Statistics. New York, NY: Springer. xxvi, 589 p. .

Marazzi A. (1993): *Algorithms, routines, and S functions for robust statistics. The FORTRAN library ROBETH with an interface to S-PLUS. With the collaboration of Johann Joss and Alex Randriamiharisoa*. Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks / Cole Advanced Books & Software, Pacific Grove, CA.

Marazzi A., Paccaud F., Ruffieux C. and Beguin C. (1998): Fitting the distributions of length of stay by parametric models. *Medical Care*, **36**(6): 915–927.

Marazzi A. and Ruffieux C. (1996): Implementing M-estimators of the gamma distribution. In: *Rieder, Helmut (ed.), Robust statistics, data analysis, and computer intensive methods. In honor of Peter Huber's 60th birthday. New York, NY: Springer-Verlag. Lect. Notes Stat., Springer-Verlag. 109, 277-297* .

McLeish D.L. (1974): Dependent central limit theorems and invariance principles. *Ann. of Probab.*, **2**: 620–628.

Müller C. (1987): *Optimale Versuchspläne für robuste Schätzfunktionen in linearen Modellen. (Optimal experimental designs for robust estimation functions in linear models)*. Dissertation, Berlin (FRG): Freie Univ. Berlin, Fachbereich Mathematik, Diss. xii, 124 S. .

Nelson D.B. (1990): Stationarity and persistence in the GARCH$(1,1)$ model. *Econometric Theory*, **6**(3): 318–334.

Parthasarathy K.R. (1967): *Probability measures on metric spaces*. Probability and Mathematical Statistics. A Series of Monographs and Textbooks. Academic Press.

Pfanzagl J. (1988): *Elementare Wahrscheinlichkeitsrechnung. (Elementary probability theory)*. De Gruyter Lehrbuch. Berlin etc.: Walter de Gruyter. xii, 313 S. DM 68.00/hbk; DM 38.00/pbk .

R Development Core Team (2005): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Reeds J.A. (1985): Asymptotic number of roots of Cauchy distribution likelihood equations. *Ann. Stat.*, **13**: 775–784.

Rieder H. (1977): Least favorable pairs for special capacities. *Ann. Stat.*, **5**: 909–921.

—— (1978): A robust asymptotic testing model. *Ann. Stat.*, **6**: 1080–1094.

—— (1980): Estimates derived from robust tests. *Ann. Stat.*, **8**: 106–115.

—— (1981a): On local asymptotic minimaxity and admissibility in robust estimation. *Ann. Stat.*, **9**: 266–277.

—— (1981b): Robustness of one- and two-sample rank tests against gross errors. *Ann. Stat.*, **9**: 245–265.

—— (1989): A finite-sample minimax regression estimator. *Statistics*, **20**(2): 211–221.

—— (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.

—— (1995): Robustness in structured models. In: *Rinne, Horst (ed.) et al., Grundlagen der Statistik und ihre Anwendungen. Festschrift für Kurt Weichselberger. Berlin: Physica-Verlag. 172-187* .

—— (1996): Estimation of mortalities. *Bl., Dtsch. Ges. Versicherungsmath.*, **22**(4): 787–816.

—— (2003): Robust estimation for time series models based on infinitesimal neighborhoods. Available under
http://www.uni-bayreuth.de/departments/math/org/mathe7/RIEDER/pubs/zzk.pdf.

Rieder H., Kohl M. and Ruckdeschel P. (2001): The Costs of not Knowing the Radius. Submitted. Appeared as discussion paper Nr. 81. SFB 373 (Quantification and Simulation of Economic Processes), Humboldt University, Berlin; also available under
http://www.uni-bayreuth.de/departments/math/org/mathe7/RIEDER/pubs/RR.pdf.

Rieder H. and Ruckdeschel P. (2001): Short Proofs on L_r–Differentiability. *Stat. Decis.*, **19**: 419–425.

Rousseeuw P. and Yohai V. (1984): Robust regression by means of S-estimators. In: *Robust and nonlinear time series analysis (Heidelberg, 1983)*, Vol. 26 of *Lecture Notes in Statist.*, p. 256–272. Springer, New York.

Rousseeuw P.J. and Croux C. (1993): Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.*, **88**(424): 1273–1283.

Ruckdeschel P. (2001): *Ansätze zur Robustifizierung des Kalman–Filters*. Dissertation, Universität Bayreuth, Bayreuth.

—— (2004a): Consequences of Higher Order Asymptotics of M-Estimators.
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/cmest.pdf.

—— (2004b): Higher Order Asymptotics for the MSE of M-Estimators on Shrinking Neighborhoods.
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/mest2.pdf.

—— (2004c): Higher Order Asymptotics for the MSE of the Median on Shrinking Neighborhoods.
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/medmse.pdf.

—— (2005a): A Motivation for $1/\sqrt{n}$ -Shrinking-Neighborhoods?
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/whysqn.pdf.

—— (2005b): Ergänzungen zu Lagrange Multiplikatoren. technical report.

—— (2005c): Ergänzungen zur Lower-Case-Lösung. technical report.

—— (2005d): Fisher Information in Transformation Models. unpublished manuscript.

—— (2005e): Higher Order Asymptotics for the MSE of One-Step-Estimators on Shrinking Neighborhoods. unpublished manuscript.

Ruckdeschel P. and Kohl M. (2005): Computation of the Finite Sample Risk of M-Estimators on Neighborhoods.
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/howto.pdf.

Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2005): S4 Classes for Distributions.
http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR.

Ruckdeschel P. and Rieder H. (2004): Optimal influence curves for general loss functions. *Stat. Decis.*, **22**: 201–223. Also available under
http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/GOptIC.pdf.

Ruckstuhl A. and Welsh A. (2001): Robust fitting of the binomial model. *Ann. Stat.*, **29**(4): 1117–1136.

Salibian-Barrera M. (2000): *Contributions to the theory of robust inference*. Dissertation, University of British Columbia, Vancouver. Available under
http://mathstat.math.carleton.ca:16080/~matias/thesis.pdf.

Salibian-Barrera M. and Zamar R. (2004): Uniform asymptotics for robust location estimates when the scale is unknown. *Ann. Statist.*, **32**(4): 1434–1447.

Serfling R. (2004): Multivariate symmetry and asymmetry. In: *Encyclopedia of Statistical Sciences. (N. Balakrishnan, C. B. Read and B. Vidakovic, eds.)*. Wiley, New York, 2. edition. In press. Preprint:
http://www.utdallas.edu/~serfling/papers/ess5011.pdf.

Shen L.Z. (1995): On optimal B-robust influence functions in semiparametric models. *Ann. Stat.*, **23**(3): 968–989.

Simpson D.G., Carroll R.J. and Ruppert D. (1987): M-estimation for discrete data: Asymptotic distribution theory and implications. *Ann. Stat.*, **15**: 657–669.

Staab M. (1984): *Robust parameter estimation for ARMA models*. Dissertation, Universität Bayreuth, Bayreuth.

Stabla T. (2005): *Robuste Adaptive Schätzung*. Diplomarbeit, Universität Bayreuth, Bayreuth.

Stein C. (1956): Efficient nonparametric testing and estimation. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, p. 187–195. University of California Press, Berkeley and Los Angeles.

Stephenson A. (2004): *evd: Functions for extreme value distributions*. R package version 2.1-6.
URL: *http://www.maths.lancs.ac.uk/˜stephena/*

Stigler S.M. (1973): The asymptotic distribution of the trimmed mean. *Ann. Stat.*, **1**: 472–477.

—— (1977): Do robust estimators work with real data? *Ann. Stat.*, **5**: 1055–1098.

Swensen A.R. (1980): *Asymptotic inference for a class of stochastic processes*. Dissertation, University of California, Berkely, CA.

Tukey J. (1960): A survey of sampling from contaminated distribution. In: *Contrib. Probab. Stat., Essays in Honor of H. Hotelling, 448-485* .

van der Vaart A. (1998): *Asymptotic statistics.*, Vol. 3 of *Cambridge Series on Statistical and Probabilistic Mathematics*. Cambridge Univ. Press., Cambridge.

Venables W. and Ripley B. (2002): *Modern Applied Statistics with S-Plus*. Statistics and Computing. Springer, 4. edition.

Victoria-Feser M.P. and Ronchetti E. (1994): Robust methods for personal-income distribution models. *Can. J. Stat.*, **22**(2): 247–258.

Witting H. (1985): *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.

Wuertz D., many others and see the SOURCE file (2005): *fBasics: Financial Software Collection - fBasics*. R package version 201.10060.
URL: *http://www.rmetrics.org*

Yohai V.J. (1987): High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**(2): 642–656.

Author Index

Subject Index

R Index