Essays on International Trade and Migration

Dissertation
zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität Bayreuth

vorgelegt von
Steffen Sirries
geboren in
Erkelenz

Dekan:	Professor Dr. Knut Werner Lange
Erstberichterstatter:	Professor Dr. Mario Larch
Zweitberichterstatter:	Professor Dr. Hartmut Egger
Termin der mündlichen Prüfung:	16.03.2016

Der Grund warum der Neandertaler weiter gemacht hat, war ja nicht Geld oder Karriere, sondern andere Neandertaler.

—B. Stromberg

Acknowledgments

Writing a thesis, and being involved in the academic business, and enjoying life at the same time is a quite demanding task, if not impossible at a fully satisfying degree. I would like to thank all the people who *convoyed* me during this time. My family, my friends and the people I met during this time have always been and still are the most important motivational ingredient for me.

I specially would like to thank my supervisors. Foremost, I thank Mario Larch for being my "Doktorvater". I am grateful for the deep insights to research you gave me as well as to the academic business, and to the daily life at your chair. You were always demanding a great deal of me with respect to academics which definitely improved my work, my thoughts, and my thesis. Your academic rigor is inspiring and will hopefully accompany me my whole life. I thank Hartmut Egger for his supervision, for his ongoing effort in giving me input, and for his patience. Your enduring interest in economics and academic discussions is still impressive to me and it has been quite beneficial for my work and my attitude as a researcher. For the support of both of you it is true that only a tiny bit of what you actually taught me is documented in this thesis. As the one who initially gave the chance to start an academic career, as my very first academic supervisor, and as a friend I thank and thoughtfully miss the formidable Dieter Urban. I like to thank Martin Leschke for being part of my committee and I thank the DFG for funding under project 592405.

Furthermore, I thank all my colleagues and assistants at the RWTH Aachen University and at the University of Bayreuth and my coauthors for numerous discussions, happy lunches, and a good time at work and outside. Special thanks go to Raoul Ghosh, Simon Loretz, Sabine Ott, Cindy Kade, and Eva-Maria Rieger for your repeated support, personal interaction and friendship. I thank Pehr-Johan Norbäck and Christoph Moser for involving and supporting me in our projects. I am deeply indebted to Joschka Wanner and Benedikt Heid, without whom finishing my thesis would have been even more uncomfortable. With Beno I share much more than an office, a neighborhood and our interest in economics, econometrics, and trade. I am sure that I will live on the experience from this intense relationship for a long time and over every distance.

Actually, I do not know how to express my gratitude to my parents. I thank my parents for their deepest trust in me and that they never questioned what I do.

Also for this project, by far the most important relationship is my wife Kristin. I always missed you during long stays apart from you. You are the most patient partner one can have.

Thank you all!

Contents

List of Figures

List of Tables

1. Introduction

This volume was prepared as a dissertation in the broad field of the economics of international trade. One could easily classify this thesis altogether to belong to the broadly defined F code of the classification provided by the Journal of Economic Literature. However, the four chapters I present here ask four different research questions which could be seen as detached from each other within this field.

With Chapter 2, I provide a rigor first time empirical quantification of a theoretical model that predicts the multilateral pattern of international migration flows in the world. Specifically, I structurally estimate a micro-founded gravity equation for migration flows. The model allows me to conduct comparative static analyses which include general equilibrium changes in migration costs. With this framework, ex ante counterfactual analysis and the quantification of migration redirection effects are possible. For a sample of 33 European Union (EU) and OECD countries, I quantify effects on immigration from two scenarios. First, I provide direct and indirect immigration effects of Turkey becoming a member of the European Union. Second, I evaluate a deeper integration of the European Union single market from lowered language and correlated cultural barriers to migration. The results show that inference from consistent regression coefficients does not ensure a correct quantification of migration flows. Comparative static results differ quantitatively and qualitatively from predictions of consistently estimated coefficients. First, comparative static effects on immigration are substantially lower and second, immigration in third countries is affected negatively by bilaterally decreased migration frictions.

In Chapter 3 we ask how the welfare quantification of trade liberalization changes if we allow workers to be mobile within established frameworks. Precisely, so-called new quantitative trade models which are prominently used to evaluate welfare effects from trade liberalization so far assume labor to be immobile. This chapter therefore provides a first structurally estimable model of international trade with endogenous international migration choices of workers. We use the model for an ex ante comparative static welfare quantification of the Transatlantic Trade and Investment Partnership. We use dyadic trade and migration data for 36 OECD countries and find that quantitative welfare predictions change if workers are allowed to migrate. The results are informative about the complex welfare changes of international economic integration agreements with respect to the interaction of trade and migration frictions.

Chapter 4 contributes to the literature that tries to explain why the observation of reduced frictions with respect to international trade due to globalization does not show up if we infer elasticities of these frictions with established tools. More detailed, to solve this *distance puzzle*, we use a newly developed gravity equation estimator derived from a heterogeneous firm micro-structure. We use three different data sets and find that the distance coefficient increases over time when standard estimators are use, while a non-linear estimation of the newly developed estimation leads to a decline in the distance coefficient over time. We show that distance puzzle, thus, arises from a growing bias of standard estimates. We explain the latter by an increase of the importance of the bias from omitting the number of heterogeneous exporting firms relative to the bias from omitting zero trade flows. Furthermore, we show that simply including zero trade flows cannot solve the distance puzzle.

And Chapter 5 tries to clarify why domestic labor market effects of firm's internationalization strategies might differ across empirical investigations. This chapter precisely investigates the effects of offshoring and FDI on German establishment employment. We compare different modes and measures of offshoring and FDI, different estimation methods, different sets of control or selection variables, and two different micro-data sets in a unified methodical framework. We can confirm positive employment effects from general FDI, market seeking FDI, and even from cost saving FDI which we find in the literature, but find negative employment effects from international sourcing which includes domestic closures. We show that the results are sensitive to the mode of internationalization rather than to the estimation method, the choice of control or selection variables, or the employed data set. We argue that this can also explain diverse results in the literature. However, we document a robust negative employment effect of international sourcing whenever a domestic restructuring is causally aligned and can confirm this result also with a quasi natural control group which is unique to one of our data sets.

While it is true that all chapters could generally be seen as stand-alone contributions to more narrowly defined strands of literature within international trade, I want to highlight that the four chapters are additionally centered around and linked by two recently dominating topics within the international trade literature.

The first recurring topic is the gravity model of economic flow variables. Augmenting the pure empirical observation of gravity forces driving international goods flows with a general equilibrium economic theory, turned the gravity equation to a gravity model. By now this model is seen as the workhorse model to analyze international trade and specifically to quantify the welfare consequences of trade liberalization within so-called new quantitative trade models and elsewhere.

Clearly, Chapters 2, 3, and 4 source from recent developments in the economic gravity literature and can therefore be seen closely related to each other. All three chapters contain some version of a gravity model and estimate a gravity equation, although with different objectives. Chapter 2 employs a theoretical gravity model for migration derived from an individual discrete choice over potential locations. I use the gravity model here to infer international migration costs and to predict equilibrium changes of international migration flows with respect to changes in migration policy.

Chapter 3 actually includes two micro-founded gravity equations, one for trade flows and one for migration flows. Here we propose a full general equilibrium model of international trade where workers are mobile in the vein of new quantitative trade models. The twofold gravity structure guides on the one hand the structural estimation of trade and migration frictions and on the other hand the equilibrium welfare calculation with respect to changes in international trade and/or migration frictions.

One common property of (most) existing gravity models is that they cannot explain the stylized fact that some countries do not trade with all other countries in the world, but for infinite trade costs. In Chapter 4 we employ a gravity estimator which is derived from a heterogeneous firms model. Within this model, zero trade flow observations for some country-pairs are explained by insufficient productivity levels of firms to serve every foreign market due to fixed costs from exporting.

One other core result of the heterogeneous firms literature is that not all firms engage in international activity. This fact inspires the the general research design and the identification strategy of Chapter 5 using micro-data and methods from the treatment evaluation literature. At the same time the empirical specification in this chapter is guided by the general insights from the heterogeneous firms literature.

So, Chapters 4 and 5 share a close relationship to the heterogeneous firms literature. For a much broader overview on both, the gravity and new quantitative trade literature and the heterogeneous firms literature I refer to specific chapters of the most recent volume of the Handbook of International Economics. Head and Mayer (2014) summarize the status of gravity in international trade while Costinot and Rodríguez-Clare (2014) do this for new quantitative trade models. Melitz and Redding (2014) and Antràs and Yeaple (2014) review the importance of heterogeneous firms in international trade over the last decade.

All chapters of this thesis are self-contained and can be read autonomously. They provide own introductions, conclusions, and appendices.

2. Comparative Statics Quantification of Structural Migration Gravity Models

2.1. Introduction

International migration is subject to various frictions. Changes of these frictions result in complex changes of migration flows which are highly relevant for policy makers. In a world of more than two countries barriers to migration between two countries contain a multilateral component. If two countries bilaterally lower their migration barriers, migration from one of these two countries to a third country becomes relatively less attractive in terms of relative costs. The literature calls this property of a multi-country migration model multilateral resistance. Theoretical concepts of multilateral resistance to migration involve potential migration redirection effects from bilateral changes in migration frictions and thus potential immigration effects on third countries. In this paper I quantify the complex changes of migration flows in a structural gravity model of international migration. I apply the Anderson (2011) model to a data set of 33 European Union and OECD countries, estimate the model's migration cost parameters implied by the model's migration gravity structure and illustrate how this framework can be used for comparative statics. I demonstrate that comparative statics are quantitatively and qualitatively different from merely interpreted gravity regression coefficients. Specifically, I explore neglected properties of a Random Utility Model (RUM) based general equilibrium migration gravity model.[1] I focus on the multilateral resistance equilibrium nature of the model by Anderson (2011) which enables a quantification of migration flows in a comparative static analysis. Since multilateral resistance terms of this model depend on all bilateral migration barriers, their change must be accounted for in a quantification of

[1]The general idea of RUMs is to derive a discrete choice model under the assumption of utility maximization following to Marschak (1959). See Section 2.3.

the equilibrium impact of changes in bilateral migration barriers on migration flows. The structural model allows me to account for these changes, additional to a consistent structural effects estimation. The effects resulting from the comparative static quantification are called *conditional general equilibrium* (conditional GE) effects.[2] However, interpreting theory-consistent regression coefficients, given that they might already control for multilateral resistance to migration, does not deliver correct predictions of migration flows, since this does not account for equilibrium changes due to multilateral resistance. The comprehensive application of multilateral resistance to migration in a comparative static analysis is therefore crucial for a quantification of immigration effects. Another advantage of the conditional GE approach is that it sheds light on the heterogeneity of immigration changes across countries. Therefore, we gain a much more differentiated picture from this exercise.

The conditional GE effects of a change in bilateral migration barriers on immigration can be obtained as follows. First, I consistently estimate the structural parameters of the theoretical model. This gives theory-consistent parameters. Then, I use parameters and observed values of the model and calculate multilateral resistance terms for every country. Third, I recalculate multilateral resistance terms for counterfactual scenarios. These new values can then be used to calculate changes in bilateral migration flows for every country-pair. This delivers migration redirection and third country effects and a detailed picture of the heterogeneity of effects on immigration. In contrast to conditional GE effects, I refer to predictions of migration flows from consistently estimated coefficients from the empirical gravity equation as *partial effects*.

I demonstrate and compare the differences between partial effects and conditional GE effects of bilateral changes in migration barriers on migration flows for two counterfactual scenarios. The first example is an evaluation of Turkey becoming a full member of the European Union and the effects on multilateral migration flows. As one of the so-called four freedoms of the single market project, becoming a member of the European Union includes the free movement of workers within all member countries. Therefore this exercise serves as a prototypical example for a policy induced change of migration barriers.

[2]Multilateral resistance is a general equilibrium concept in the model by Anderson (2011) which means that it involves all bilateral changes of migration frictions in the world. The term conditional stems from the fact that multilateral resistance effects are conditional on the supplied labor force to a country. In the trade literature this term was coined by Anderson and Yotov (2010). See Section 2.2 for a discussion on the relation to the trade gravity literature of this approach and Section 2.3 for details on the model. Chapter 3 of this thesis shows how the supplied labor force to a country can be endogenized in a gravity model of migration and trade.

In the second scenario I hypothetically lower migration frictions between all European Union member countries in terms of language barriers. The literature on the determinants of migration costs documents the economic importance of language barriers for aggregate migration flows (Chiswick, 2015; Adserà and Pytlikova, 2015). The European Commission proposes the promotion of the integration process via the framework strategy for multilingualism. Effectively, language is seen as a long term policy variable for deepening integration, especially via increased job opportunities of migrants within the European Union. Thus, both scenarios lower existing migration barriers for a subset of countries.

The results can be summarized as follows. Lowering migration frictions increases migration. Partial and conditional GE effects on immigration deliver expected qualitative results for the countries which are directly involved in the bilateral reduction of frictions. This is in line with previous findings in the literature and our general intuition with respect to migration barriers. While the results do not differ with respect to the sign of the change in migration flows for directly affected countries, I document substantial quantitative and qualitative differences between interpreting consistent regression coefficients and comparative static results on immigration. For example, partial effects predict a bilateral relative increase in immigration of Turkey becoming a member of the European Union of around 113% for Turkey-EU country-pairs, whereas the comparative static analysis only predicts an increase of around 75% for bilateral immigration for the same country-pairs. Partial effects for bilateral immigration are constant, while conditional GE effects are heterogeneous with values ranging from 7% to 98% for the Turkey-EU country-pairs. Total immigration changes for the two counterfactuals are heterogeneous at the country-level, although due to very different reasons.[3] For the partial effects prediction of total immigration changes at the country-level, I must multiply the uniform estimate from the regression with the observed migration flows for every country-pair where a change in the migration cost vector is induced in the counterfactual scenario, i.e. EU-Turkey or EU-EU country-pairs respectively. Immigration from all other countries does not change for this exercise. Since the share of immigration from these countries in total immigration again differs at the country-level, I do observe some heterogeneity for total immigration changes across destination countries. In contrast to this, heterogeneity of the conditional GE effects on total immigration results from changes in multilateral resistance to migration which is endogenous in the model. The degree of heterogeneity is substantial for

[3] To be precise, bilateral immigration is the migration flow from *one* particular country to *one* other particular country. Total immigration is the aggregate migration flow from *all* countries to one particular country.

conditional GE effects, while it is minor for partial effects. The qualitative difference of partial effects and conditional GE effects is also documented for total immigration. On the one hand, partial effects on total immigration are zero for third countries and positive for all other countries. On the other hand, total immigration from the comparative static analysis – which accounts for multilateral resistance and delivers conditional GE effects on total immigration – are substantially negative for third countries. With this, I quantify causal migration redirection effects from multilateral resistance which cannot be detected by simply interpreting estimated coefficients of a gravity equation. To sum up, consistent estimates from a migration gravity model do not give a correct impact of migration frictions on migration flows. This paper is related to two strands of literature. First, recent contributions to the literature of international migration propose varieties of migration gravity equations to analyze international migration matrices and to estimate parameters of bilateral migration flow determinants. Multilateral resistance to migration is accounted for at the estimation stage in some works. As a result these studies provide consistently estimated coefficients. I briefly review this literature in Section 2.2. Beine et al. (2015) provide a broader guide through this young strand of literature. I contribute to this literature with the first application of the model by Anderson (2011) which includes a comparative static analysis to account for multilateral resistance comprehensively. I will refer in the following to the international trade literature. Most importantly, I transfer the insight from a comparative statics quantification of multilateral resistance to trade to multilateral resistance to migration. Relations to this literature are reviewed in Section 2.2. The remainder of the paper is as follows. Section 2.2 briefly reviews recent migration gravity studies and relates to the trade literature. Section 2.3 recaps the migration gravity model of Anderson (2011) on which I will base the structural estimation and the comparative static analysis. Section 2.4 presents the estimation stage, after which Section 2.5 provides information on the compiled data set. Section 2.6 discusses the results of the estimation, while Section 2.7 discusses the comparative static quantification for both counterfactual scenarios. Section 2.8 concludes.

2.2. Related Literature

2.2.1. Gravity Equations for Migration

The first connection of empirical regularities of migration flows to a law of gravity similar to Newton's law of gravity dates back to the 19$^{\text{th}}$ century. Early works which document the idea of gravitational forces à la Newton driving spatial interaction of economic entities include Carey (1858). Ravenstein (1885) is known for characterizing laws of migration following a gravity intuition. Only recently this idea has regained attention in the economic literature on international migration. Beine et al. (2015) blame the absence of (dyadic) migration data for a century without progress on migration gravity. However, recent contributions employ varieties of migration gravity estimations to establish bilateral determinants of migration flows (Grogger and Hanson, 2011; Beine et al., 2011; Ortega and Peri, 2013; Bertoli and Fernández-Huertas Moraga, 2013; Orefice, 2015; Figueiredo et al., 2016; Adserà and Pytlikova, 2015).

The common denominator of micro-foundations for a migration gravity equation which the literature proposes is a RUM. Generally, the maximized utility by individuals in a RUM consists of two parts. One which is observed by the researcher and one which is private information of the individual. The observed part of the utility is given by the payoffs from migration (usually income) reduced by the costs from migrating. To gain individual probabilities of migration from a discrete choice model, distributional assumptions about the unobserved part of individual utility are necessary. Migration gravity approaches in the literature differ by their specification of the observed part and by the distributional assumptions about the unobserved part of utility. Beine et al. (2015) give an overview on RUMs which are used for international migration gravity modeling. In the next step, one can derive an aggregate expression for migration from these probabilities. In Section 2.3, I explore this in more detail for the model proposed by Anderson (2011).

Specifying the payoffs of the observed part of the utility with bilateral variables already yields a partial equilibrium gravity model for aggregate migration flows. See Beine et al. (2015) for a general presentation of this approach. Existing studies use this RUM foundation either to establish empirical specifications of migration barriers or to clarify selection and sorting issues of migration with respect to payoffs and costs. For example Grogger and Hanson (2011) use such a framework with two skill groups to derive an empirical migration gravity equation which sheds light on migration costs and the international sorting of migrants across skill groups. Beine et al. (2011) document the importance of network effects measured via past stocks of bilateral migrants with a similar design of the analysis. Ortega and Peri (2013) construct a unique measure of migration policy tightness to establish that migration costs are considerably affected by policy regulations. Adserà and Pytlikova (2015) give a detailed picture of the effects of different language barriers on migration flows. Bertoli and Fernández-Huertas Moraga (2013) derive a concept of multilateral resistance to migration from a generalization of the distributional assumptions of the unobserved component of utility. They show that the error term of an empirical gravity equation of migration shares entails a multilateral component which generally depends on alternative migration destinations and bilateral migration barriers. This concept of multilateral resistance can then be controlled for in an estimation on data with higher frequency using recent advances of panel data estimators.

In contrast to Bertoli and Fernández-Huertas Moraga (2013), Anderson (2011) proposes a theoretical concept of multilateral resistance to migration in a general equilibrium model, which also builds on the canonical RUM. From this, he can obtain a structural migration gravity model where multilateral resistance to migration occurs for standard assumptions on the unobserved part of utility (see Section 2.3 for details on the model). Note that multilateral resistance to migration is a general equilibrium concept in Anderson (2011) while it is an assumption about the error term of an empirical gravity equation in Bertoli and Fernández-Huertas Moraga (2013). To quantify the effects of multilateral resistance to migration of Anderson (2011), a comparative static analysis of the model is necessary. Orefice (2015) and Figueiredo et al. (2016) refer to the model of Anderson (2011), although they do not use the model for a comparative static analysis but estimate partial effects. They estimate the model to establish regional trade agreements as a determinant of bilateral migration frictions.

I contribute to this literature by using the model of Anderson (2011) for a quantification of multilateral resistance consistent counterfactual migration flows. Some works in the literature present empirical specifications which already control for the concept of multilateral resistance to migration of Anderson (2011) at the estimation stage. So do Orefice (2015) and Figueiredo et al. (2016). Therefore, they present consistent estimated coefficients which can be used for a prediction of migration flows in form of partial effects. The theoretical model allows me to conduct a comparative static analysis which is consistent with changes of multilateral resistance terms in a new counterfactual equilibrium. The quantification I present here therefore entails for the first time endogenous equilibrium changes of multilateral resistance to migration.

2.2.2. Relations to Structural Trade Gravity

The importance of a comprehensive treatment of multilateral resistance in a general equilibrium model is well known for trade gravity approaches, although not commonly implemented. Anderson and van Wincoop (2003) introduce the concept of multilateral resistance to trade in a micro-founded general equilibrium trade gravity model. Over the last decade, such structural trade gravity models became fundamental in the trade literature.[4] Anderson and van Wincoop (2003) show the puzzling high negative effect of national borders on trade in goods to be driven by missing multilateral resistance to trade. Specifically, they show that a comparative static analysis of equilibrium changes of trade flows, which account for multilateral resistance comprehensively, does not show the puzzling effect of borders anymore. However, the trade gravity literature elucidates of the fact that interpreting consistent regression coefficients does not give a correct quantification of the impact of bilateral changes in trade costs on trade flows. Head and Mayer (2014) write that the estimation of empirical trade gravity models became "[...] just a first step before a deeper analysis [...]". I transfer this insight to the migration gravity literature by estimating the model of Anderson (2011) and conducting a comparative static analysis. My results show qualitatively a similar picture of the importance of multilateral resistance to migration compared to multilateral resistance to trade. Although structural gravity models are sometimes reviewed as applying to factor flows as well (Head and Mayer, 2014; Anderson, 2011), a comparable implementation and quantification seems to be missing in the migration literature.

The formal representation of the theoretical migration gravity model of Anderson (2011) (see Section 2.3) is analogous to the one in Anderson and van Wincoop (2003). This allows me to draw on recent insights from the trade gravity literature.

As for Anderson and van Wincoop (2003), the modularity of the structural migration gravity model by Anderson (2011) allows one to correct consistent estimates of bilateral changes in migration barriers to ones which account for the effects via a recalculation of the multilateral resistance module for a new equilibrium of migration flows. However, Head and Mayer (2014) call the interpretation of theory consistent estimates at the estimation stage of a trade gravity the Partial Trade Impact. I call the prediction of migration flows from this *partial effects*, as outlined in Section 2.1. For predicted migration flows which incorporate multilateral resistance term changes from bilateral changes in migration barriers, I use the term conditional GE effects. For the trade analog, Anderson and Yotov (2010) coin the term conditional general equilibrium technique. Compared to a full general equilibrium where GDPs and expenditures are recalculated in the comparative static analysis, the multilateral resistance terms can be recalculated separately

[4] I dare to say that the theoretical underpinnings of trade gravity models by Eaton and Kortum (2002), Anderson (1979), and Anderson and van Wincoop (2003) are initially accountable for the so-called literature of new quantitative trade models. Roughly, these models use micro-founded general equilibrium trade models to quantify economic impacts from changes in trade determinants on spatially linked economic entities. See Costinot and Rodríguez-Clare (2014) for an overview on this literature.

in the trade gravity model as well.[5] Head and Mayer (2014) therefore call conditional GE effects in a trade gravity the Modular Trade Impact. Importantly, they note that the difference of results for moderate trade cost changes between conditional GE and full general equilibrium effects are minor.

2.3. Migration Gravity Model

I briefly recap the structural migration gravity system proposed by Anderson (2011). In a multi-country setting emigration is a discrete choice from the set of countries in the world from the perspective of a single worker. A worker h migrates from country o (origin) to d (destination) only if her utility of choosing d is bigger than for all other possible choices. The utility in country o is given by her wage, w_o plus an idiosyncratic part of utility. Migration to country d involves country-pair specific costs of migration, $\delta_{od} > 1 \; \forall \; d \neq o$ and $\delta_{oo} = 1$, which reduce utility in country d in an iceberg type way, w_d/δ_{od}. Migration additionally involves a worker and country-pair specific factor of utility ϵ_{odh}. So a worker decides to migrate from country o to d iff $(w_d/\delta_{od})\epsilon_{odh} \geq w_o\epsilon_{ooh}$. In line with discrete choice theory, utility of a representative migrant is separated into two parts. One part which is observable and determined by characteristics at the country-pair-level, $V_{od} = \ln(w_d) - \ln(w_o) - \ln(\delta_{od})$. The second part of the utility, which is worker and country-pair specific, $\varepsilon_{odh} = \ln \epsilon_{odh}$, is not observable for the researcher. With distributional assumptions for ε_{odh}, one can derive the probability of a randomly drawn worker to migrate.[6]

From multiplying the number of people in country o with the migration probability of a randomly drawn worker of country o, $G(V_{od})$, we gain an aggregate multi-country migration flow equation,

$$M_{od} = G(V_{od})N_o, \tag{2.1}$$

where N_o is the number of natives in o and $G(V_{od})$ gives the proportion of migrants from o to d, which is given by

$$G(V_{od}) = \frac{e^{V_{od}}}{\sum_k e^{V_{ok}}}. \tag{2.2}$$

[5] Anderson (2011) highlights the general modularity of the gravity equation in more detail and with respect to a sectoral analysis.

[6] Adopted to the multi-country discrete choice of a representative worker, a derivation of the multinomial-logit probabilities is given in Appendix A.1.

Plugging in the V's yields a multilateral migration flow equation as

$$M_{od} = \frac{\frac{w_d}{\delta_{od}}}{\sum_k (\frac{w_k}{\delta_{ok}})} N_o. \tag{2.3}$$

The migration flow from country o to d is positively associated with the wage in the destination country d, bilateral migration barriers to all other potential countries than d, δ_{ok}, and the number of natives of the source country o, N_o.[7] Migration is negatively associated with bilateral migration barriers, captured by δ_{od}, and wages in all other countries than d, w_k. Note that the idiosyncratic or worker specific part of the utility is captured implicitly by the functional form of Equation (2.3). So the individual probabilities, which are derived in Appendix A.1, already capture the unobserved part of the migrant's utility, ε_{odh}.

Using accounting identities and the labor market clearance condition, Anderson (2011) provides the following migration gravity system:[8]

$$M_{od} = \underbrace{\frac{L_d N_o}{N^w}}_{\text{frictionless migration}} \underbrace{\frac{1/\delta_{od}}{\Omega_d W_o}}_{\text{migration frictions}}, \quad \text{with} \tag{2.4}$$

$$\underbrace{\Omega_d = \left[\sum_o \left(\frac{1/\delta_{od}}{W_o}\right) \frac{N_o}{N^w}\right]}_{\text{inward multilateral resistance}}, \quad \underbrace{W_o = \left[\sum_d \left(\frac{1/\delta_{od}}{\Omega_d}\right) \frac{L_d}{N^w}\right]}_{\text{outward multilateral resistance}}. \tag{2.5}$$

The masses which drive migration flows in this gravity model are given by N_o, the population of the origin country, and by L_d, the labor force supplied to country d. Both increase migration flows between a bilateral pair of countries and their product goes into the flow equation relatively to the world population N^w. Bilateral migration barriers, δ_{od}, decrease migration flows. Ω_d and W_o indicate the multilateral resistance terms to migration.

Section 2.4 estimates Equation (2.4) structurally to infer δ_{od}, and in Section 2.7 I use this system to conduct the comparative static analysis. This can be done by realizing that multilateral resistance terms can be solved for observed values of N_o, L_d, and δ_{od}.

Before I go on, several things are worth mentioning about this model. First of all, we can observe the hypothetical migration pattern of a frictionless world by the first part of Equation (2.4). In a world without any friction to migration, we would observe the mi-

[7]Beine et al. (2015) call the latter the potential of a country for sending migrants.

[8]For intermediate steps of the derivation see Appendix A.2.

grant share from country o of the labor force supplied to d to be equal to country o's share of the world population. From this we can nicely observe the general two-way migration nature of the model. The precise two-way migration pattern is additionally shifted by bilateral migration costs and multilateral resistance terms. The frictionless view already points to the second important fact, that the model would only imply a zero migration flow if the frictions between two countries o and d were infinitely large. Migration frictions are collected in the second part of Equation (2.4). Frictions are a composite of bilateral migration barriers, δ_{od}, and multilateral resistance terms. Bilateral migration costs affect bilateral migration flows relative to the multilateral resistance terms. We can already see that multilateral resistance terms depend on bilateral migration barriers. Therefore, a change in the bilateral migration cost vector for one country-pair affects all countries' multilateral resistance terms which has to be accounted for when it comes to a prediction of migration flows. Technically multilateral resistance terms are averages of inverse migration frictions weighted by the relative size of a country. The inward multilateral resistance term collects all barriers for migrants *to* a specific migration destination country, while the outward multilateral resistance term collects all barriers for migrants *from* a specific migration origin country. Anderson and Yotov (2010) give a nice intuition for these terms for trade flows. They suggest understanding inward multilateral resistance as the uniform markup a buyer pays for a bundle of goods from a hypothetical world market. Outward multilateral resistance is then understood as the average trade cost which an exporter faces when selling to this world market. Transferring this intuition to migration means that inward multilateral resistance captures migration barriers for every migrant to destination country d for migrants from a hypothetical world origin, i.e. irrespective of her origin country. Then, outward multilateral resistance measures the uniform costs every migrant faces for migration from country o to the hypothetical migrant's country, i.e. irrespective of her actual destination country.[9] Put differently, inward multilateral resistance of a country aggregates unilateral *immigration* barriers from a hypothetical world origin country and outward multilateral resistance of a country aggregates *emigration* barriers to a hypothetical world destination. Multilateral resistance terms are aggregate concepts. Migration flows at the aggregate (Equation (2.4)) are determined by bilateral migration barriers relative to multilateral resistance terms. Also, multilateral resistance terms vary across countries. A change in bilateral migration barriers results in heterogeneous migration effects. The multilateral resistance terms entail non-trivial, multilateral changes of the migration pattern from bilateral changes in migration barriers,

[9] How to transfer the incidence intuition to migration is not obvious since for migration it is not clear who is the hypothetical entity which is actually charged.

which can be inferred from the comparative static analysis in Section 2.7.

Also note that there is no term left in Equation (2.4) which explicitly captures wage differentials, since they were substituted out via the labor market clearance equation (see Appendix A.2). This explains the difference of the empirical specification of the migration gravity to other RUM based migration approaches like Grogger and Hanson (2011). Furthermore, the theoretical migration gravity model is, in a way, agnostic about the classical differentiation between push and pull factors and the importance of specific migration barriers. Simply put, δ_{od} is not specified by the model. The specification of migration barriers is an empirical question and oftentimes hinges on the availability of bilateral measures and data.[10] I leave the presentation of the empirical specification for Section 2.4.

2.4. Structural Estimation of the Migration Gravity System

The formally equal representation of the structural migration gravity model and the structural trade gravity model allows me to borrow several insights from the trade gravity literature for a structural estimation of Equation (2.4). With a stochastic error term, Equation (2.4) can be written as

$$M_{od} = \exp\left(\ln L_d + \ln N_o - \ln N^w + \ln(1/\delta_{od}) - \ln \Omega_d - \ln W_o\right) + \varepsilon_{od}. \tag{2.6}$$

Multilateral resistance to migration terms, $\ln \Omega_d$ and $\ln W_o$, are accounted for in the estimation with origin and destination fixed effects as do Orefice (2015) and Figueiredo et al. (2016). Anderson and van Wincoop (2003) and Feenstra (2004) are usually credited for the inclusion of importer and exporter fixed effects to capture multilateral resistance to trade. I follow Santos Silva and Tenreyro (2006) who show a bias from estimating a log-linearized gravity equation via OLS if data are heteroskedastic. Standard heteroskedasticity tests reject the Null hypotheses of a constant variance of residuals after an estimation of a correctly specified gravity also for migration data. The argument for the bias from estimating a log-linearized gravity via OLS then holds true. Therefore, I estimate Equation (2.6) via Poisson Pseudo Maximum Likelihood (PPML). I control for $\ln L_d$ and $\ln N_o$ via the inclusion of the correct set of fixed effects to capture the multilateral resistance terms. Note also that with included origin and destination fixed

[10] The same is true for any structural trade gravity model.

effects, all unilaterally varying determinants of migration flows and many classical push and pull factors of migration are accounted for. World population, $\ln N^w$, is captured by a constant.

PPML allows me to include migration flows in levels instead of logged migration flows in a log-linearized version of the model for a linear estimation via OLS. Thus, zero migration flow observations do not drop out during the estimation.[11] Also following Santos Silva and Tenreyro (2006), PPML estimates Equation (2.6) consistently for a sample which includes many zero observations. Remember that the theoretical model only predicts zero migration flows between a pair of countries if their migration barriers are infinite. Zero observations in the data thus are assumed to occur randomly or due to measurement errors in form of rounding errors.[12]

For the purpose of this paper, I stick to a fairly simple specification of δ_{od}. I specify bilateral migration barriers as

$$\delta_{od}^{-1} = \exp(\gamma_1 \ln DIST_{od} + \gamma_2 CONTIG_{od} + \gamma_3 COLONY_{od} + \gamma_4 LANG_{od} + \gamma_5 EU_{od}), \quad (2.7)$$

where $\ln DIST_{od}$ is the log of distance between country o and d. $CONTIG_{od}$ and $COLONY_{od}$ indicate contiguity and a common colonial history of country-pairs. $LANG_{od}$ is equal to one if a country-pair shares at least one common official language and EU_{od} is one if a country-pair belongs to the European Union.

As usual I have to assume regressors to be exogenous to collect consistent estimates of the γ coefficients and consistent estimated migration barriers for the comparative static analysis. This assumption might not be plausibly fulfilled for the EU_{od} indicator variable due to a selection bias. One might argue that the inclusion of distance and origin and destination fixed effects already captures a lot of the selection process of becoming a European Union member. However, to overcome a potentially left selection bias, as Figueiredo et al. (2016), I follow Baier and Bergstrand (2007) and include directional bilateral fixed effects in an auxiliary estimation. Augmenting data by the time dimension allows me to infer γ_5 less prone to a bias from selection. I then estimate Equation (2.7) with the constrained coefficient from the auxiliary estimation to infer δ_{od}.

There are further observations one might make with respect to the specification. As

[11] Ortega and Peri (2013) add a small value to all observations to circumvent the problem of zero observations. In general, this leads to biased estimates. See Santos Silva and Tenreyro (2006).

[12] This is also true for structural trade gravity estimations. Egger et al. (2011) use a two part model to allow for a different data generating process for zero observations of bilateral trade flows. See also Helpman et al. (2008) and Chapter 4 of this thesis on zero observations in trade gravity estimations.

previously mentioned, wages are substituted out by the labor market clearance condition, and therefore bilaterally varying wage differentials do not show up in the empirical specification. Note also that the classical distinction between push and pull factors of migration is perfectly in line with a correct specification of migration barriers in a structural gravity estimation. Most of these factors are already captured by the origin and destination fixed effects. An obviously missing determinant of bilateral migration barriers is the restrictiveness of migration policies. A bilaterally varying measure for migration policy is simply not yet available. An already launched data project, the IMPALA database, might solve this missing data problem for future research.[13] With the free movement of labor within the European Union, the EU-pair dummy variable captures at least a part of this potential variation.

To sum up, my preferred estimation includes origin and destination fixed effects, specifies migration costs according to (2.7) with a constrained coefficient for γ_5 and employs PPML. I present the results of the auxiliary regression and the outlined estimation in Section 2.6.

2.5. Data

As a measure for M_{od} I use the yearly inflow of foreign population by nationality. The meta source for this information here is the International Migration Database (IMD) compiled and freely provided by the OECD.[14] To my knowledge the IMD offers the most extensive coverage in terms of origin and destination country combinations of aggregate and dyadic migration flow data. The IMD collects data which are initially gathered at the national level, mainly by statistical offices and official registers who try to maintain consistent definitions of immigrants over time. I use the inflows of foreign population by nationality from the IMD. National information are either derived from population registers and residence and/or work permits or by special surveys for some countries.[15] Countries rarely use specific methods to collect data on migration, especially when it comes to migrant outflows. Even if there might be a legal obligation to report out migration in a specific country, there is no obvious incentive for individuals to indicate emigration. Therefore, I only use migrant inflows and follow the literature to construct

[13] See http://www.impaladatabase.org/.

[14] See https://stats.oecd.org/Index.aspx?DataSetCode=MIG.

[15] The countries which use different special survey approaches are Ireland, United Kingdom, Australia and New Zealand. Detailed Information on methods and sources by country can be found at the website given in Footnote 14.

a dyadic data set on migration flows.[16] Standard geographical information stem from the GeoDist data set provided by CEPII.[17] I extracted population data from World Development Indicators provided by the World Bank.[18] For the auxiliary estimation I compile data over the period from 2000 to 2012. For the main regressions I keep the cross section of 2010 because coverage in this year is most extensive. Potentially the IMD offers a set of 210 origin regions and 34 destination countries. For some specifications in 2.6 I employ the largest possible sample, excluding duplicates due to regional aggregations. The main sample is defined by the countries which belong to the OECD and/or to the European Union. Due to missing migration data, I provide the comparative static results on a subsample of 33 countries of these.[19]

2.6. Estimation Results

As discussed in Section 2.4, I provide two sets of estimation results. The auxiliary estimation from which I gain a consistent coefficient for the EU-pair dummy is given in Table 2.1. Table 2.2 provides estimation results of Equation (2.6), including my preferred specification, from which I predict migration barriers for the comparative static analysis. For both tables I provide OLS and the preferred PPML results for different samples and for different sets of included fixed effects. I also indicate whether the PPML regressions include zero observations or whether I use the corresponding sample of the OLS estimates which does not include zero observations. All depicted standard errors are heteroskedasticity robust. For Table 2.2 I also present regression results which do not constrain the EU-pair coefficient.

Table 2.1 reads as follows. From the left to right, I reduce the sample size to achieve a set of countries where PPML estimation converges and where the singularity condition of the variance matrix for the huge set of dummy variables is fulfilled. All regressions include origin-year and destination-year fixed effects to capture multilateral resistance terms. Columns (1)-(3) show OLS results, where column (1) does not include directional country-

[16] Other studies use migration stock data either to construct flow data from these or to directly use stock data as a long term equivalent to flows (see Figueiredo et al. (2016)).

[17] See `http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6`.

[18] See `http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators`.

[19] The 33 countries are: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Republic of, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

pair fixed effects. Column (3)-(5) present results on a reduced sample of 15 destination countries.[20] Column (4) presents PPML results without zero observations and column (5) presents results of my preferred PPML specification including zero observations.

Except for column (1), which does not control for selection, the EU-pair coefficient is positive as expected and highly significant in all specifications. The preferred specification of column (5) reports a coefficient of 0.76 which translates to an average percent effect of $(exp(0.760) - 1) * 100\% = 113.83\%$. This means that, conditional on all other regressors, becoming a member of the European Union increases immigration between country-pairs on average by around 113%. For the specification of the main estimation, which I use to predict migration barriers to use in the comparative static analysis, I will constrain the EU-pair dummy to this estimate.

In Table 2.2 the most right column (8) reports the estimates which I use for the prediction of migration barriers for the comparative static analysis in Section 2.7. All other columns report results for variations in the sample and contrast (constrained) OLS to (constrained) PPML results. The overall picture for this migration gravity is as expected. I estimate a negative and highly significant effect of bilateral distance on migration flows, where coefficients are lower for the EU-OECD sample and for PPML results in general. Contiguity of countries is either insignificant or increases migration significantly in column (7) and (8). A common colonial past of countries leads also to significantly higher migration between countries and seems to be less pronounced, but still very high in economic terms, for the EU-OECD-sample. This picture is repeated for the common language dummy. The coefficients are highly significant with a coefficient of 0.578 in the preferred specification. This translates to an average partial effect of sharing a common language of $(exp(0.578) - 1) * 100\% = 78.24\%$.

Both estimated coefficients, which are of interest for the comparison to conditional GE effects to partial effects in Section 2.7, are substantial in driving migration flows. The European Union formulates four freedoms as a basis for the single market project. One of these four freedoms is the free movement of workers including working permissions in all member countries without any disadvantages for migrants. Therefore the EU-pair dummy is prototypical for a policy change influencing migration flows. With a partial effect of around 113% this is already indicated here for partial effects. The same is true for the common language dummy with around 78%. I confirm the result of the literature (Chiswick, 2015) that language and correlated cultural barriers are economically

[20]The 15 destination countries are Australia, Belgium, Canada, Denmark, Finland, Germany, Italy, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, United States. The same set of destination countries is used in Ortega and Peri (2013).

Table 2.1.: Auxiliary Migration Gravity Estimation for Years 2000 to 2012

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) PPML	(5) PPML
log(Distance)	-1.153***				
	(0.0143)				
Contiguity	-0.299***				
	(0.0517)				
Colony	1.388***				
	(0.0460)				
European Union	-0.201***	0.426***	0.723***	0.742***	0.760***
	(0.0348)	(0.0406)	(0.0503)	(0.0669)	(0.0670)
Common Language	1.159***				
	(0.0263)				
Observations	44,464	44,464	7,054	7,054	7,089
Origin-Year FE	Yes	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	Yes	Yes	Yes
Including zeros	No	No	No	No	Yes
Sample	Full	Full	Reduced	Reduced	Reduced

Notes: Dependent variable for OLS columns is the log of migration flows from country o to country d, $\ln M_{od}$. Dependent variable for PPML columns is migration flows in levels, M_{od}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

Table 2.2.: Migration Gravity Estimation for the Year 2010

VARIABLES	(1) OLS	(2) OLS	(3) PPML	(4) PPML	(5) OLS	(6) OLS	(7) PPML	(8) PPML
log(Distance)	-1.209***	-1.037***	-0.954***	-0.922***	-0.784***	-0.651***	-0.574***	-0.589***
	(0.0500)	(0.0456)	(0.0719)	(0.0666)	(0.0840)	(0.0743)	(0.107)	(0.0919)
Contiguity	-0.234	-0.230	0.169	0.166	0.165	0.242	0.500**	0.500**
	(0.184)	(0.193)	(0.208)	(0.211)	(0.188)	(0.193)	(0.217)	(0.217)
Colony	1.160***	1.250***	1.012***	1.042***	0.718***	0.821***	0.532***	0.518**
	(0.158)	(0.158)	(0.140)	(0.139)	(0.192)	(0.193)	(0.205)	(0.208)
European Union	-0.385***	0.760	0.444**	0.760	0.0842	0.760	0.858***	0.760
	(0.112)	(-)	(0.216)	(-)	(0.156)	(-)	(0.302)	(-)
Common Language	1.133***	1.165***	0.988***	0.994***	0.694***	0.699***	0.586***	0.578***
	(0.0851)	(0.0863)	(0.126)	(0.128)	(0.151)	(0.149)	(0.220)	(0.223)
Observations	4,160	4,160	4,940	4,982	1,095	1,095	1,145	1,205
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Including zeros	No	No	Yes	Yes	No	No	Yes	Yes
Sample	Full	Full	Full	Full	OECD-EU	OECD-EU	OECD-EU	OECD-EU

Notes: Dependent variable for OLS columns is the log of migration flows from country o to country d, $\ln M_{od}$. Dependent variable for PPML columns is migration flows in levels, M_{od}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

important migration flow shifters.

2.7. Counterfactuals

In this section, I present selected results of the counterfactual scenarios with a special emphasis on specific gains from the conditional GE approach.[21] Therefore I first contrast the estimated partial effects from Section 2.6 to their counterpart from the conditional GE analysis. Second, I want to shed light on the heterogeneity of effects in contrast to the average effects from estimation. Third, I show multilateral migration redirection and its effects on countries which are not directly affected by the induced bilateral changes. I do this for both scenarios.

To be clear about the counterfactual scenarios, I outline the involved steps to gain the subsequent results. Once I collect observed values for N_o, N_d and N^w, and estimated migration barriers δ^{-1} (Equation (2.7)), I can solve for the multilateral resistance terms (Equation (2.5)) and gain migration flows for the baseline b, M_{od}^b (Equation (2.4)). The next step is to change the world to a counterfactual scenario and to resolve the multilateral resistance terms. The resulting migration flows are defined M_{od}^c and vary across counterfactual scenarios, c. The two changes of the world which I induce are the following. For the Turkey counterfactual I change the EU-pair dummy variable to one between Turkey and current European Union member countries. For the language counterfactual I set the dummy variable of a common official language equal to one for all European Union member countries.

Simply interpreting the consistently estimated coefficients would lead us to a conclusion like 'if Turkey becomes a member of the European Union, we expect an bilateral increase in immigration for Turkey from European Union member countries and vice versa of 113.83% on average'. The interpretation of the common language dummy variable would imply an on average higher bilateral immigration between countries which share a common official language of 78.24%. Table 2.3 contrasts these two results with the conditional GE effects. The counterpart to the estimated partial effects is obviously given by the average relative change of immigration in Turkey from European Union member countries and in European Union member countries from Turkey. I calculate $\Delta M_{od}^c\% = (\frac{M_{od}^c}{M_{od}^b} - 1) * 100\%$ for the respective countries and take the average, indicated by $\overline{\Delta M_{od}^c\%}$. With 74.63% we observe a substantially lower immigration effect from the

[21] Note that potentially this simulation exercise delivers changes for every bilateral migration flow.

comparative static analysis. For the language counterfactual I calculated an average relative change in bilateral immigration between European Union member countries of 25.8%, which is an even bigger drop from partial effects to the conditional GE impact.

Table 2.3.: Partial vs Conditional GE Effects on Immigration

	(1)	(2)
	$[exp(\gamma_i) - 1] * 100\%$	$\overline{\Delta M^c_{od}\%}$
	Partial Effects	Conditional GE Effects
Turkey Counterfactual	113.83 %	74.63 %
Language Counterfactual	78.24 %	25.8 %

Notes: Values in column (1) use estimates from Table 2.2. Column (2) reports average bilateral immigration between Turkey and European Union countries from the comparative static analysis.

Table 2.4 documents the heterogeneity of the bilateral effects, simply by showing bilateral flows, $\Delta M^c_{od}\%$, for the Turkey counterfactual.[22] As the estimated coefficient is the same for all European Union country-pairs, partial effects would be uniform here and are depicted in Table 2.3. I can document substantial heterogeneity in the bilateral immigration effects of Turkey becoming a member of the European Union.[23] Immigration changes for Turkey range from 7.37% from Slovenia to 85.72% from Belgium. Immigration changes for European Union countries from Turkey are much more homogenous around 98%. Although this indicates that the heterogeneity might be driven mainly by origin country characteristics, there is no obvious pattern apparent for all other bilateral changes.

In contrast to Table 2.3, Table 2.5 includes all multilateral changes of immigration in and from all countries, and not only the bilateral immigration changes of directly involved countries of the respective counterfactual.[24] Column (1) depicts the partial effects and column (2) the results from the comparative static analysis. Remember that I reduce

[22]For convenience of the presentation, I do not report the corresponding bilateral heterogeneity for the language counterfactual in a table since it would consist of $23 * 22 = 506$ entries. For details on the heterogeneity of effects from this counterfactual I refer to Table 2.7 where I report average changes of total immigration at the country-level.

[23]Since I only observe heterogeneity at the second decimal place if Turkey is the origin country for the percentage changes, I report the average for immigration from Turkey to European Union countries in the last row of Table 2.4. This indicates that the heterogeneity seems to be driven by the country of origin characteristics here.

[24]Partial effects are given by $\Delta M^p_d\% = (\frac{M^p_d}{M^b_d} - 1) * 100\%$, where $M^b_d = \sum_o M^b_{od}$ and $M^p_d = \sum_o M^b_{od} + (M^b_{od} * (exp(\gamma_i) - 1) * \mathbb{1}_{cf})$, where γ_i is either the estimated coefficient of the EU-pair, or of the common language dummy variable. $\mathbb{1}_{cf}$ is an indicator function which is either one for European Union country-pairs including Turkey in the Turkey counterfactual scenario, or it is one for European Union country-pairs in the language counterfactual.

Table 2.4.: Bilateral Immigration Changes from Turkey Counterfactual for Turkey and EU Member Countries

Origin	Destination	$\Delta M_{od}^{c}\%$
Austria	Turkey	47.19
Belgium	Turkey	85.72
Czech Republic	Turkey	29.17
Germany	Turkey	80.48
Denmark	Turkey	25.20
Spain	Turkey	82.52
Estonia	Turkey	12.92
Finland	Turkey	54.11
France	Turkey	77.87
United Kingdom	Turkey	84.11
Greece	Turkey	49.42
Hungary	Turkey	42.27
Ireland	Turkey	62.47
Italy	Turkey	83.92
Lithuania	Turkey	22.47
Luxembourg	Turkey	61.29
Latvia	Turkey	19.61
Netherlands	Turkey	51.51
Poland	Turkey	63.71
Portugal	Turkey	55.65
Slovakia	Turkey	19.32
Slovenia	Turkey	7.37
Sweden	Turkey	58.08
Turkey	EU Countries	98.11

Notes: Values in column (1) report bilateral immigration changes at the country-pair-level of the comparative static analysis which includes conditional GE effects.

migration barriers in both counterfactuals. On average, we observe a substantial increase in immigration for both counterfactual scenarios. For both counterfactuals, we observe on average a much lower immigration effect from the comparative static analysis, compared to the partial effects calculation. Obviously, simply interpreting the coefficients does not give the whole picture of immigration changes since immigration changes due to changes in multilateral resistance terms are not captured.

Table 2.5.: Average Percent Immigration Changes for 33 EU and OECD Countries

	(1)	(2)
	$\Delta M_d^p\%$	$\Delta M_d^c\%$
	Partial Effects	Conditional GE Effects
Turkey counterfactual	49.89	6.44
Language counterfactual	31.44	4.55

Notes: Values in column (1) report average immigration changes for and from all 33 countries from calculations using partial effects. Column (2) reports average immigration changes for and from all 33 countries of the comparative static analysis which includes conditional GE effects.

Tables 2.6 and 2.7 give a more detailed picture of immigration changes at the country-level. Again, I contrast partial (column (1)) and conditional GE changes (column (2)) of total immigration. The picture of Table 2.5 is repeated at the country-level. Partial effects are much higher than the immigration effects which account for multilateral resistance equilibrium changes. Naturally, partial effects are zero for countries which are not directly affected by the counterfactual change of the world. Most important for Tables 2.6 and 2.7 are the non-zero third country effects measured by $\Delta M_d^c\%$. For all countries which are not directly affected by a decrease in migration frictions, we observe a substantial decrease in immigration. These negative immigration changes for third countries nicely show migration redirection effects. For example, in both counterfactual scenarios, Norway loses the most in terms of immigration with -19% if Turkey becomes a European Union member, and -17.76% if European Union countries hypothetically were to share at least one common language. Norway, for example, is geographically closely linked to the European Union without being a member, which perfectly in line with my expectation. For the counterfactual concerning Turkey, Turkey would gain most with an immigration increase of 40.06%, which is also in line with the intuition that Turkey would receives many migrants from the European Union if it were to join the single market. For the language counterfactual, we observe the highest increase in immigration of 33.37% for Portugal, which does not share a common language with any European Union country. Belgium has the lowest increase of 0.62% which is consistent with the above results, since it shares its official languages with contiguous neighboring countries France, Germany, and Netherlands. We also see a substantial degree of heterogeneity for conditional GE effects at the country-level. Note that this heterogeneity is driven by the model structure and is therefore endogenous. The minor heterogeneity we see for partial effects is only driven by the exogenous heterogeneity of baseline migration flows. The quite substantial drop of immigration changes from partial effects to the comparative static results, the substantial negative third country effects, and the heterogeneity of effects from this exercise document the importance of multilateral resistance for migration gravity.

2.8. Conclusion

To my knowledge, I present the first comparative static analysis of changes of migration flows which builds on a general equilibrium migration gravity model. There are multiple gains from this analysis compared to existing migration gravity works. First, I document that partial effects estimations cannot recover the full impact of changes in migration barriers on migration flows. This holds true even if the estimation is consistent with the theory and controls for multilateral resistance to migration. Second, the analysis documents a substantial endogenous degree of heterogeneity of immigration effects across countries in contrast to uniform consistent estimates. Third, the comparative static analysis accounts for non-trivial conditional general equilibrium changes via multilateral resistance to migration. These changes uncover indirect third country and migration redirection effects. I show that a simple interpretation of estimated coefficients of a migration gravity are qualitatively and quantitatively misleading. Researchers who want to use the gravity equation in the context of international migration are to be made aware of these effects.

Table 2.6.: Percent Immigration Changes from Turkey Counterfactual for 33 EU and OECD Countries

Destination	(1) $\Delta M_d^{p}\%$	(2) $\Delta M_d^{c}\%$	EU Member
Australia	0.00	-9.07	No
Austria	101.81	14.74	Yes
Belgium	102.67	0.64	Yes
Canada	0.00	-8.60	No
Chile	0.00	-11.40	No
Czech Republic	101.70	31.27	Yes
Denmark	97.56	32.06	Yes
Estonia	98.65	27.65	Yes
Finland	98.44	10.27	Yes
France	92.74	25.87	Yes
Germany	99.34	20.74	Yes
Greece	95.08	38.14	Yes
Hungary	100.40	25.78	Yes
Iceland	0.00	-15.27	No
Ireland	97.98	11.56	Yes
Israel	0.00	-14.64	No
Italy	91.90	27.63	Yes
Japan	0.00	-10.12	No
Korea, Republic of	0.00	-10.37	No
Latvia	97.98	26.76	Yes
Lithuania	98.64	28.65	Yes
Luxembourg	101.49	6.70	Yes
Mexico	0.00	-10.51	No
Netherlands	99.97	26.49	Yes
New Zealand	0.00	-8.48	No
Norway	0.00	-19.00	No
Poland	100.54	29.46	Yes
Portugal	95.22	34.03	Yes
Slovakia	103.38	27.25	Yes
Slovenia	100.78	31.03	Yes
Spain	90.53	28.84	Yes
Sweden	96.02	13.85	Yes
Switzerland	0.00	-17.48	No
Turkey	78.65	40.06	Yes, hypothetically
United Kingdom	89.23	21.24	Yes
United States	0.00	-8.53	No

Notes: Values in column (1) report total immigration changes for and from all 33 countries from calculations using partial effects. Column (2) reports average immigration changes for and from all 33 countries of the comparative static analysis which includes conditional GE effects.

Table 2.7.: Percent Immigration Changes from Language Counterfactual for 33 EU and OECD Countries

Destination	(1) $\Delta M_d^{p}\%$	(2) $\Delta M_d^{c}\%$	EU Member
Australia	0.00	-8.31	No
Austria	68.67	14.82	Yes
Belgium	69.61	0.62	Yes
Canada	0.00	-7.98	No
Chile	0.00	-10.48	No
Czech Republic	68.35	31.14	Yes
Denmark	65.38	31.68	Yes
Estonia	66.02	27.45	Yes
Finland	66.08	9.97	Yes
France	62.40	25.69	Yes
Germany	66.93	20.73	Yes
Greece	53.51	24.58	Yes
Hungary	66.88	25.10	Yes
Iceland	0.00	-14.17	No
Ireland	66.19	11.37	Yes
Israel	0.00	-12.91	No
Italy	60.81	26.64	Yes
Japan	0.00	-9.25	No
Korea, Republic of	0.00	-9.47	No
Latvia	65.34	26.25	Yes
Lithuania	65.67	27.93	Yes
Luxembourg	68.73	6.67	Yes
Mexico	0.00	-9.69	No
Netherlands	67.28	26.29	Yes
New Zealand	0.00	-7.77	No
Norway	0.00	-17.76	No
Poland	67.21	28.96	Yes
Portugal	63.56	33.37	Yes
Slovakia	69.53	27.41	Yes
Slovenia	67.46	30.75	Yes
Spain	60.44	28.31	Yes
Sweden	64.42	13.58	Yes
Switzerland	0.00	-16.42	No
Turkey	0.00	-15.81	No
United Kingdom	59.91	20.97	Yes
United States	0.00	-7.91	No

Notes: Values in column (1) report total immigration changes for and from all 33 countries from calculations using partial effects. Column (2) reports average immigration changes for and from all 33 countries of the comparative static analysis which includes conditional GE effects.

3. International Trade and Migration: A Quantitative Framework[1]

3.1. Introduction

Most generally, economists are aware of the fact that frictions commonly prevent an efficient allocation of goods and factors. Most recently, economic evaluations of changes in frictions to international trade gained a considerable amount of attention. A dominating topic in international economic policy these days is the potential mega deal of economic integration between the European Union and the United States, the Transatlantic Trade and Investment Partnership (TTIP). The general economic intention of such a deal is to reduce frictions. As is negotiated for TTIP, most international economic integration agreements intend to reduce frictions for trade in goods, at least as a first step. The international trade literature proposes various methodical frameworks and methods to evaluate welfare effects of such agreements with respect to changes in trade frictions. So-called *new quantitative trade models* represent a strand of literature which aims at a rigor empirical quantification of welfare effects from globalization in terms of trade frictions (Costinot and Rodríguez-Clare, 2014). However, one potentially important aspect of economic integration is the integration of labor markets in terms of lowered migration frictions. This seems to be mainly missed by both, the political debate and the literature of new quantitative trade models. The latter assume workers to be immobile across countries. Therefore, welfare evaluations from these models are agnostic about the potential impact of lowered migration frictions and the interdependency of goods and factor flows. And in the public debate on economic integration agreements like TTIP, mobility of workers appears to be of minor importance, although for the example of the European Union as another economic integration mega deal integration is defined by the

[1] This chapter bases on joint work with Mario Larch. All remaining errors in this volume are mine.

freedom of movement of workers.[2] We formulate the broad research question of this paper as "How does a state-of-the-art welfare evaluation of trade liberalizations change if we allow workers to be mobile?". In the vein of new quantitative trade models, we therefore propose a micro-founded general equilibrium trade model where workers endogenously decide about migration.

Specifically, we first set up a one sector one factor trade model with Armington (1969) preferences and link it to a recently developed Random Utility Model (RUM) migration model borrowed from Anderson (2011). Thus, we incorporate in an established trade model an endogenous individual migration decision. From this we gain a tractable and quantifiable framework of international trade and international migration. An important property of the framework is that we can derive two structurally estimable gravity equations for the estimation of core parameters of the model. From a gravity equation for international trade we can infer trade frictions and from a migration gravity equation we can infer migration frictions. The structural estimation ensures that we account for both, multilateral resistance to trade and multilateral resistance to migration. With observed values of the model and consistently estimated frictions of international trade and migration, we can use the model for a full general equilibrium quantification of trade, migration, and welfare effects from changes in bilateral frictions. The framework generally allows us to quantify ex post or ex ante changes in trade and/or migration frictions.

Arkolakis et al. (2012) show for different existing micro-foundations and resulting rationales for trade of new quantitative trade models that changes in the import penetration ratio due to changes in barriers to trade and the trade elasticity are sufficient to calculate welfare effects. However, a priori it is not clear whether this result holds true if for example a fundamental assumption like immobile workers is relaxed. For example Heid and Larch (2016) show for relaxing another common assumption in this class of models, namely the assumption of full employment, that the welfare formula of Arkolakis et al. (2012) changes qualitatively. We show that the sufficient statistic of Arkolakis et al. (2012) for welfare calculation in terms of GDP per labor force does not change for the consideration of workers being mobile across countries. However, we document that for an actual welfare quantification the precise change in the import penetration ratio due to changes in frictions is partly driven by the immobile workers assumption. To show the resulting effects of this, we use our model to ex ante evaluate TTIP. That means for a

[2]One notable exception in the political debate in Germany is Klaus F. Zimmermann. As the director of the director of the Institute for the Study of Labor (IZA), Zimmermann repeatedly puts the view of an increasing potential of TTIP if it would additionally include reduced frictions for workers into the policy debate. See for example Zimmermann (2014).

counterfactual analysis we change the world to a scenario where the European Union and the United States have already signed a regional trade agreement. For a sample of 36 European Union and OECD countries we compare the evaluation of TTIP of our model which allows migration to an evaluation where migration costs are prohibitive. Additionally we compare welfare effects from TTIP to a TTIP scenario where at the same time migration frictions between the European Union and United States of America are reduced by the average effect of a free movement policy between countries. Our results show subtle changes in welfare effects if we allow for migration.

The remainder of the paper is structured as follows. Section 3.2 presents the structural gravity model including an individual, explicit and multilateral migration decision. The following Section 3.3 derives two estimable gravity equations, one for trade and one for migration from which we extract structural parameters. The bilateral trade and migration data we use are described in Section 3.4 and Section 3.6 presents the results and the design of the counterfactual analysis. Section 3.7 concludes.

3.2. The Model

Our quantifiable general equilibrium framework consists of a trade system à la Anderson and van Wincoop (2003) and a linked migration system following Anderson (2011). That means we link an established multi-country perfect competition trade model to a multi-country Random Utility Model of migration. The indirect utility function of a representative consumer drives the individual probability to migrate and therefore the aggregate migration flow. From the model we derive two gravity equations, one for trade and one for migration. A crucial general equilibrium feature for both gravity equations is multilateral resistance which we have to account for in the empirical analysis. We start with a one sector one factor multi-country perfect competition trade model.

3.2.1. Aggregate Bilateral Trade Flows

The utility of a representative consumer in country j is denoted U_j. We assume goods to be differentiated by country of origin following Armington (1969). The quantity of purchased goods from country i is given by c_{ij}, leading to the following utility function

$$U_j = \left[\sum_{i=1}^{n} \beta_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{3.1}$$

where n is the number of countries in the world, σ is the elasticity of substitution in consumption across goods, and β_i is a positive preference parameter indicating the product appeal for goods from country i. With the factory gate price of the good in country i, p_i, and a trade costs factor $t_{ij} > 1$ of goods from i to j following (Samuelson, 1952), profit maximization implies $p_{ij} = p_i t_{ij}$. The representative consumer maximizes Equation (3.1) subject to the budget constraint $Y_j = \sum_{i=1}^{n} p_{ij} c_{ij}$. The value of aggregate sales of goods from country i to country j can then be expressed as

$$X_{ij} = p_i t_{ij} c_{ij} = \left(\frac{\beta_i p_i t_{ij}}{P_j} \right)^{1-\sigma} Y_j, \tag{3.2}$$

where P_j is a standard CES price index given by $P_j = [\sum_{i=1}^{n} (\beta_i p_i t_{ij})^{1-\sigma}]^{1/(1-\sigma)}$. In general equilibrium, total sales of a country correspond to its nominal income, i.e., $Y_i = \sum_{j=1}^{n} X_{ij}$. Assuming labor to be the only factor of production and full employment, we can express GDP also by total factor income, i.e., $Y_i = w_i L_i$, where w_i is the uniform wage in country i and L_i is the number of people working in country i, so the labor force. Note already that L_i changes if we allow for migration in the model. As we assume perfect competition and one unit of labor produces one unit of output, it holds that $p_i = w_i$.

The individual decision of a worker to migrate from an origin country j to a destination i will crucially depend on the net attainable utility in every alternative. We therefore derive the indirect utility of a representative consumer in country j, U_j^*, given by

$$\begin{aligned} U_j^* &= \frac{1}{L_j} \left[\sum_{i=1}^{n} \beta_i^{\frac{1-\sigma}{\sigma}} \left(\left(\frac{\beta_i}{P_j} \right)^{1-\sigma} (p_i t_{ij})^{-\sigma} Y_j \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \\ &= \frac{Y_j P_j^{\sigma-1}}{L_j} \left[\sum_{i=1}^{n} (\beta_i p_i t_{ij})^{1-\sigma} \right]^{\frac{\sigma}{\sigma-1}} = \frac{Y_j P_j^{\sigma-1}}{L_j} \left[P_j^{1-\sigma} \right]^{\frac{\sigma}{\sigma-1}} \\ &= \frac{Y_j}{L_j P_j} = \frac{w_j}{P_j}. \end{aligned} \tag{3.3}$$

Hence, the decision to migrate will, among other factors, depend on the real wage differences between country j and potential destinations i.

3.2.2. The Trade Gravity Equation

From the set up so far, we derive a gravity equation for bilateral trade flows following Anderson and van Wincoop (2003). We first use $Y_i = \sum_{j=1}^{n} X_{ij}$ which summarizes the general equilibrium nature of our model and implies market clearing, i.e.,

$$Y_i \quad = \quad \sum_{j=1}^{n} X_{ij} = (\beta_i p_i)^{1-\sigma} \sum_{j=1}^{n} \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} Y_j. \tag{3.4}$$

Solving for scaled prices, $\beta_i p_i$, defining world income by $Y^W \equiv \sum_j Y_j$, and income shares $\theta_j \equiv Y_j/Y^W$, we can write bilateral trade flows as given in Equation (3.2) as

$$X_{ij} = \frac{Y_i Y_j}{Y^W} \left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma}, \tag{3.5}$$

where

$$\Pi_i \equiv \left(\sum_{j=1}^{n} \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \theta_j\right)^{1/(1-\sigma)}, \tag{3.6}$$

and

$$P_j \equiv \left(\sum_{i=1}^{n} \left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma} \theta_i\right)^{1/(1-\sigma)}. \tag{3.7}$$

We substituted equilibrium scaled prices into the definition of the price index to obtain the multilateral resistance to trade.

Note that this system of equations is anlagous to the system given in Equations (9)-(11) in Anderson and van Wincoop (2003). Since migration only will change the number of people working in country i, L_i, we take as a first result that the estimation of trade costs does not change when relaxing the assumption of immobile workers. Migration therefore influences trade flows, X_{ij}, via changes in prices respectively in total production which we take as second results from the model so far.

3.2.3. Aggregate Bilateral Migration Flows

Following the presentation of Anderson (2011), a worker h migrates from country j to i if her utility is bigger in i. Since we assume that migration also involves country-pair specific costs modeled as a factor $\delta_{ji} > 1$, the individual decision of a worker to migrate is given by

$$(U_i^*/\delta_{ji})\epsilon_{jih} \geq U_j^* \epsilon_{jjh}, \tag{3.8}$$

where ϵ_{jih} indicates a worker specific, unobserved utility factor. That means, when it comes to the individual migration decision, we allow workers' preferences about origin-destination-pairs to be heterogeneous. In our multi-country setting, and with log utility, and separating the observed and the unobserved part of Equation (3.8) we can write the probability of the decision of worker h to migrate from j to i, P_{jih}, as

$$\begin{aligned} P_{jih} &= \text{Prob}(V_{ji} + \varepsilon_{jih} > V_{jk} + \varepsilon_{jkh} \,\forall\, k \neq i) \\ P_{jih} &= \text{Prob}(\varepsilon_{jkh} < \varepsilon_{jih} + V_{ji} - V_{jk} \,\forall\, k \neq i), \end{aligned} \tag{3.9}$$

where $V_{ji} = \ln(U_i^*) - \ln(U_j^*) - \ln(\delta_{ji})$ and $\varepsilon_{jih} = \ln \epsilon_{jih}$. Assuming ε_{jih} to be distributed independently, identically extreme value, we can derive now (see Appendix B.1) the multinomial-logit probabilities à la McFadden (1974) given by

$$P_{jih} = \frac{e^{V_{ji}}}{\sum_k e^{V_{jk}}}. \tag{3.10}$$

This gives the probability of a randomly drawn worker h to migrate from country j to country i. Obviously, this probability coincides in the aggregate with the share of migrants from country j to country i. From multiplying this share with the number of natives in country j, N_j, we get an equation for the aggregate migration flow from j to i as

$$M_{ji} = P_{jih} N_j. \tag{3.11}$$

Now, inserting (3.3) in (3.11), we can write the bilateral migration flow equation as

$$M_{ji} = \frac{\frac{w_i}{w_j} \frac{P_j}{P_i} \frac{1}{\delta_{ji}}}{\sum_k \left(\frac{w_k}{w_j} \frac{P_j}{P_k} \frac{1}{\delta_{ji}}\right)} N_j \tag{3.12}$$

$$= \frac{\frac{w_i}{P_i} \frac{1}{\delta_{ji}}}{\sum_k \left(\frac{w_k}{P_k} \frac{1}{\delta_{jk}}\right)} N_j. \tag{3.13}$$

The migration flow from country j to i is positively associated with the real wage in the destination country i, bilateral migration barriers to all other potential countries than i, δ_{jk} and the number of natives of the source country j. Migration is negatively associated with bilateral migration barriers captured by δ_{ji} and the real wage in all other countries than i. Importantly note that the idiosyncratic or worker specific part of the utility is captured implicitly by the functional form of Equation (3.10). So, the multinomial-logit probabilities already capture the heterogeneity of workers. In Section 3.2.6 we illustrate this specific property of the model.

3.2.4. Migration Gravity Equation

We now derive a gravity equation for bilateral migration flows following Anderson (2011). First, we define $\omega_i \equiv \frac{w_i}{p_i}$. Then Equation (3.13) boils down to the aggregate migration flow expression in Anderson (2011) given by

$$M_{ji} = \frac{\frac{\omega_i}{\delta_{ji}}}{\sum_k (\frac{\omega_k}{\delta_{jk}})} N_j. \tag{3.14}$$

We note that $\sum_i M_{ji} = N_j$ and $L_i = \sum_j M_{ji}$. With the world labor supply $N^w \equiv \sum_j N_j = \sum_i L_i$, labor market clearance is given by

$$L_i = \omega_i \sum_j \left(\frac{1/\delta_{ji}}{W_j} \right) N_j, \tag{3.15}$$

where $W_j \equiv \sum_k \omega_k / \delta_{jk}$. Solving for ω_i, it follows that

$$\omega_i = \frac{L_i}{\Omega_i N^w}, \tag{3.16}$$

where $\Omega_i \equiv \sum_j \frac{1/\delta_{ji}}{W_j} \frac{N_j}{N^w}$. Using Equation (3.16), we can write W_j as $W_j = \sum_k \frac{L_k}{\Omega_k \delta_{jk} N^w}$. Again using Equation (3.16), we can write bilateral migration flows as given in Equation (3.14) as

$$M_{ji} = \frac{L_i N_j}{N^w} \frac{1/\delta_{ji}}{\Omega_i W_j}, \tag{3.17}$$

with

$$\Omega_i = \sum_j \frac{1/\delta_{ji}}{W_j} \frac{N_j}{N^w}, \tag{3.18}$$

and

$$W_j = \sum_i \frac{1/\delta_{ji}}{\Omega_i} \frac{L_i}{N^w}. \tag{3.19}$$

Analogously to the derivation of the trade gravity model where we substituted equilibrium scaled prices into the price index, we substituted equilibrium real wages into Equation (3.13) to obtain multilateral resistance to migration. As a result we keep in mind that we can derive from our model a gravity equation for international migration to infer migration costs. As for the trade gravity equation, the migration gravity equation incorporates specific multilateral resistance to migration terms which we have to capture in the estimation of migration costs.

3.2.5. Welfare

Probably the most important goal of new quantitative trade models is to quantify welfare effects of trade liberalizations. A specific difference of our framework to existing models is that the number and the composition of workers with respect to their origin country might change due to changes in trade and migration costs. As it is sufficient with immobile workers to measure welfare as the real income of a country which is usually done in the literature, obviously it is not for our model. To take the number of workers in a country into account we therefore measure welfare via the per capita equivalent of the real income, namely the real wage or the production per labor force. Although this welfare measure does not take into account the mentioned composition of the labor force, so far we abstain from calculating welfare effects for migrants and natives separately due to a missing appropriate alternative.[3] So, if we are willing to assume for the moment that every worker – migrant or native – is the same, we can compare our welfare results directly to other new quantitative trade models.

Arkolakis et al. (2012) provide a sufficient statistic to calculate welfare effects for a class of trade models including the basic Armington set up we use for the trade side. Specifically, they show that the domestic expenditure or the import penetration ratio respectively, and the trade elasticity are sufficient to calculate welfare changes from trade shocks in these models. We show in Appendix B.2 that welfare changes in our model with mobile workers and welfare measured as real GDP per labor force can also be calculated using the statistic of Arkolakis et al. (2012). However, we show for a numerical example of our model in Section 3.2.6 and also for the counterfactual analysis in Section 3.6.1 that domestic sales and therefore welfare implications change if we allow workers to be mobile in a new quantitative trade model.

3.2.6. Numerical Example

With the aggregate equation for trade flows, $X_{ij} = (\beta_i w_i / P_j)^{(1-\sigma)} t_{ij} Y_j$, and the aggregate equation for migration flows, $M_{ji} = \frac{(w_i/P_i)/\delta_{ji}}{\sum_k ((w_k/P_k)/\delta_{jk})} N_j$, the nominal wage, $w_i = Y_i / L_i$, the

[3] Also note that our welfare measure does not take into account the idiosyncratic part of the migration utility, yet.

price index, $P_i^{1-\sigma} = \sum_k (\beta_k w_k t_{ki})^{(1-\sigma)}$, the income level, $Y_i = \sum_j X_{ij}$, and the labor force, $L_i = \sum_j M_{ji}$, the equilibrium is determined. In Section 3.6.1 we use this model to provide empirical comparative statics results of counterfactual scenarios where we change trade and migration costs. These changes involve non-trivial changes of the system. In the following we present some basic properties of the model with a numerical three country example. As we will do for the counterfactual scenarios, for the baseline calculation we insert in this system Y_i, N_i, $t_{ij}^{(1-\sigma)}$, δ_{ji}^{-1}, and σ.

The first numerical example (Table 3.1) is characterized by completely symmetric countries in terms of size and income levels ($N_i = 100$; $Y_i = 100$), no trade frictions ($t_{ij}^{(1-\sigma)} = 1$), and no migration frictions ($\delta_{ji}^{-1} = 1$). So we look at the trade, migration and welfare pattern of a symmetric, frictionless world. For all examples we set $\sigma = 5$ and $\beta_i = 1$. Internal trade and migration costs are set to zero, i.e. $t_{ii}^{(1-\sigma)} = 1$ and $\delta_{ii}^{-1} = 1$.

Table 3.1.: Numerical Illustration 1

Country	Y_i	N_i	$t_{ij}^{(1-\sigma)}$	$\delta_{ij}^{(-1)}$	w_i	L_i	X_{ii}	M_{ii}	β_i	P_i	w_i/P_i
A	100	100	1	1	1	100	33.33	33.33	1	0.76	1.316
B	100	100	1	1	1	100	33.33	33.33	1	0.76	1.316
C	100	100	1	1	1	100	33.33	33.33	1	0.76	1.316

For this frictionless and symmetric set up we can see that the model predicts a symmetric world where every country sells one-third of its goods nationally and trades one-third of its goods with every other country respectively. This result nicely illustrates the love of variety property of the micro-structure behind the aggregate trade flow equation. The same is true for migration. One-third of the natives of every country stay, while the other two-thirds migrate in equal proportion to the rest of the world. This result stems from the unobserved part of the utility from migration. The functional form of the multinational logit probability captures this part of the utility and drives the general two-way migration structure of the model. Importantly note here, that we observe migration even if the world is symmetric and thus there are no real wage differences.

For the next example (Table 3.2) we introduce some level of trade costs with country B being less remote to trade compared to countries A and C, and we make migration infinitely costly. Note also that we assume trade and migration costs to be symmetric, i.e. $t_{ij}^{(1-\sigma)} = t_{ji}^{(1-\sigma)}$ and $\delta_{ji}^{-1} = \delta_{ij}^{-1}$. Specifically we set $t_{AB}^{(1-\sigma)} = t_{BA}^{(1-\sigma)} = 0,5$, $t_{BC}^{(1-\sigma)} = t_{CB}^{(1-\sigma)} = 0,5$, $t_{AC}^{(1-\sigma)} = t_{CA}^{(1-\sigma)} = 0,4$, and $\delta_{ij}^{(1-\sigma)} = 0$ for all countries.

The less trade remote country B sells less goods domestically and trades more vice versa

Table 3.2.: Numerical Illustration 2

Country	Y_i	N_i	$t_{ij}^{(1-\sigma)}$	$\delta_{ij}^{(-1)}$	w_i	L_i	X_{ii}	M_{ii}	β_i	P_i	w_i/P_i
A	100	100	0.5	0	1	100	53.106	100	1	0.854	1.171
B	100.867	100	0.4	0	1.009	100	49.563	100	1	0.845	1.194
C	100	100	0.5	0	1	100	53.106	100	1	0.854	1.171

compared to A and C in equilibrium. Note also that country B realizes higher gains from trade which translate into a higher welfare or real wage respectively. The increased real wage gives rise to migration in the next setting where we allow for free migration in the world (see Table 3.3). We keep the trade cost setting as before, i.e. we set $t_{AB}^{(1-\sigma)} = t_{BA}^{(1-\sigma)} = 0,5$, $t_{BC}^{(1-\sigma)} = t_{CB}^{(1-\sigma)} = 0,5$, $t_{AC}^{(1-\sigma)} = t_{CA}^{(1-\sigma)} = 0,4$, and set $\delta_{ij}^{(1-\sigma)} = 1$ for all country-pairs.

Table 3.3.: Numerical Illustration 3

Country	Y_i	N_i	$t_{ij}^{(1-\sigma)}$	$\delta_{ij}^{(-1)}$	w_i	L_i	X_{ii}	M_{ii}	β_i	P_i	w_i/P_i
A	100	100	0.5	1	1.006	99.425	52.948	33.142	1	0.858	1.172
B	102.32	100	0.4	1	1.012	101.15	50.573	33.717	1	0.848	1.193
C	100	100	0.5	1	1.006	99.425	52.948	33.142	1	0.858	1.172

The higher real wage in country B now drives migration into country B. But even costless migration does not equalize real wages or leads to a full agglomeration here. As for Table 3.1, workers migrate to every destination and only a part of the workers react to the real wage differences and therefore increase the labor force in country B. Compared to the setting before with infinite migration costs, this translates into less trade and a lower real wage for country B, but a higher real wage for the other two countries. For the counterfactual analysis in Section 3.6.1 we will see that the general equilibrium nature of the model with asymmetric countries in terms of size and income and with estimated trade and migration costs involves non-trivial changes in trade flows, migration flows, and hence in welfare due to counterfactual changes in trade and migration costs. The most important lesson from the last numerical example where workers can migrate is that we observe a change in domestic sales of the countries. Due to a change in the number of people working in a country, a welfare prediction from trade liberalization changes compared to a model where workers are assumed to be immobile.

3.3. Structural Estimation of Trade and Migration Frictions

The parameters we want to estimate from the data are namely the trade costs, t_{ij}, and the migration costs, δ_{ij}. We describe in this section how we infer these from the structurally estimating the two gravity equations.

3.3.1. Trade Gravity Estimation

The gravity equation for international trade is given by the system of Equations (3.5) to (3.7). With a stochastic error term, ν_{ij}, we rewrite Equation (3.5) as

$$X_{ij} = \exp\left(\ln Y_i + \ln Y_j - \ln Y^W + (1-\sigma)\ln t_{ij} - \ln \Pi_i^{1-\sigma} - \ln P_j^{1-\sigma}\right) + \nu_{ij}, \quad (3.20)$$

Hence, a structural estimation of the aggregate, bilateral trade flow expression implies that one accounts in the estimation for the total production levels of countries i and j, the world income, trade costs and the respective multilateral resistance to trade terms. Since we are able to derive the same gravity equation here as for models without migration (see for example (Anderson and van Wincoop, 2003)), we can base the estimation of Equation (3.20) on the recent developments which the literature proposes.[4] Most importantly we have to care about the unobservable multilateral resistance to trade terms in the estimation to prevent a bias from omitting these. From Equations (3.6) and (3.7) we see that these terms depend on all bilateral trade costs and income shares. Therefore missing these terms for estimation would lead to an omitted variable bias as is shown in Anderson and van Wincoop (2003). We follow the standard approach and include importer and exporter fixed effects to capture $\ln \Pi_i^{1-\sigma}$ and $\ln P_j^{1-\sigma}$ respectively. At the same time, these fixed effects capture $\ln Y_i$ and $\ln Y_j$, while $\ln Y^W$ is captured by a constant in the regression. The actual goal of the estimation stage here is to consistently recover trade costs. Therefore we have to specify bilateral trade costs empirically. We again follow the literature here and specify $t_{ij}^{1-\sigma}$ as

$$t_{ij}^{1-\sigma} = \exp(\beta_1 \ln DIST_{ij} + \beta_2 RTA_{ij} + \beta_3 CONTIG_{ij} + \beta_4 LANG_{ij} + \beta_5 COLONY_{ij}), \quad (3.21)$$

[4]See Head and Mayer (2014) for a very useful summary of recent insights on the estimation of gravity equations.

where $DIST_{ij}$ measures the distance between countries i and j, RTA_{ij} indicates whether the two countries are jointly part of at least one regional trade agreement, and $CONTIG_{ij}$, $LANG_{ij}$, and $COLONY_{ij}$ indicate whether the countries are contiguous, share a common language or an historic colonial relationship.

Santos Silva and Tenreyro (2006) argue that the standard approach of estimating the above multiplicative model of Equations 3.20 by simply taking logarithms and estimate the resulting linear model with OLS yields inconsistent parameter estimates due to heteroskedasticity in bilateral trade data. In addition, log-linearization would drop all zero observations from the trade and migration matrices, which is clearly not theoretically justified and will in general lead to biased estimates. Thus, we do not rely on OLS estimates for the bilateral trade and migration costs respectively but chose the Poisson pseudo-maximum-likelihood estimator.[5] By now this approach is standard for trade gravity estimation and below we estimate the migration gravity as well via PPML for the same reason.

As the regional trade agreement indicator is a policy variable which potentially violates the exogeneity condition caused by self selection of specific countries into such an agreement, we have to account for that in the estimation, too. To do so we follow Baier and Bergstrand (2007) and estimate Equation 3.20 in two steps. In the first step we add the time dimension to our bilateral trade data set and include in our specification directional country-pair fixed effects to capture potential self selection effects which are constant over time at the country-pair-level. We then restrict the RTA coefficient in the main regression to the estimated RTA coefficient from this auxiliary first step regression.

3.3.2. Migration Gravity Estimation

To infer migration costs, we similarly estimate the migration gravity equation from Section 3.2.4. We add a stochastic error term, μ_{ji}, to Equation (3.17) and rewrite it to

$$M_{ji} = \exp\left(\ln L_i + \ln N_j - N^w + \ln(1/\delta_{ji}) - \ln \Omega_i - \ln W_j\right) + \mu_{ji}. \tag{3.22}$$

As for the trade gravity estimation, we capture the multilateral resistance to migration terms, $\ln \Omega_i$ and $\ln W_j$, and the labor supply and the number of natives, $\ln L_i$ and $\ln N_j$ by

[5] As argued by Santos Silva and Tenreyro (2006), PPML is also likely to be a more sensible choice than other consistent non-linear estimators (such as non-linear least squares or Gamma PML), because it gives equal weight to all observations. They additionally demonstrate that the PPML estimator is generally well behaved in the context of constant elasticity models by conducting Monte Carlo simulations (see Santos Silva and Tenreyro (2011)).

i and j specific fixed effects. World population, N^w, is captured by a constant. Since all unilaterally varying drivers of migration in this structural estimation are already captured by the fixed effects, we come up with a parsimonious specification of migration costs. The gravity variables we include are given by

$$\delta_{ji}^{-1} = \exp(\gamma_1 \ln DIST_{ji} + \gamma_2 EU_{ji} + \gamma_3 CONTIG_{ji} + \gamma_4 LANG_{ji} + \gamma_5 COLONY_{ji}), \quad (3.23)$$

where $DIST_{ji}$, $CONTIG_{ji}$, $LANG_{ji}$, and $COLONY_{ji}$ correspond to the regressors of the trade gravity estimation. EU_{ji} indicates whether the two countries belong to the European Union. We include the common European Union membership indicator as a measure for free movement of workers. Within the European Union, workers are generally allowed to move freely and to work in any member country. Hence, we expect migration costs within the European Union to be significantly lower and vice versa. However, as for the RTA indicator in the trade regression, the membership of the European Union is not randomly assigned and therefore potentially involves a selection process which might not be captured by the other regressors. We circumvent this problem in the same way as we do for the RTA indicator and follow again the two step procedure of Baier and Bergstrand (2007). As mentioned, we also follow for the estimation of the migration gravity the recommendation of Santos Silva and Tenreyro (2006) for the same arguments and estimate our preferred specification via PPML.

3.4. Data

We compile the data we use for the estimation of the gravity equations and for the baseline calculation in the comparative static analysis from different freely available sources. The bilateral trade flows as a measure for, X_{ij}, are originally compiled by Head et al. (2010) and provided by the CEPII.[6] The data set includes bilateral trade flows for all world pairs of countries from the year 1948 to 2006. Bilateral migration flows, M_{ji} stem from the International Migration Database available from the OECD.[7] For the measure of bilateral migration flows we use here, namely the yearly inflows of foreign population by nationality, the OECD provides these information for a broad coverage of origin and destination countries from 2000 to 2012. For the information on RTAs we use Mario Larch's Regional Trade Agreements Database from Egger and Larch (2008).[8] This ready-

[6]See `http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8`.

[7]See `https://stats.oecd.org/Index.aspx?DataSetCode=MIG`.

[8]See `http://www.ewf.uni-bayreuth.de/en/research/RTA-data/index.html`.

to-use data set includes all multilateral and bilateral trade agreements as notified to the World Trade Organization from 1950 to 2014. There is a total of 453 such agreements in the data set. At last, we source the population information for N_j and N^W from World Development Indicators.[9] From all the data we merge, we keep a sample of 36 EU-OECD for the estimation on which we also provide the comparative static results.[10]

3.5. Estimation Results

We present the results of the estimation of the gravity equations in Tables 3.4 to 3.7. Table 3.4 presents the results for the auxiliary trade gravity estimation where we include directional country-pair fixed effects to control for potential self selection into RTAs. Our preferred specification using the PPML estimator in column (4) reports a coefficient of 0.344 which translates to an average increase in bilateral trade flows of a country-pair signing a regional trade agreement of $(exp(0.344) - 1) * 100\% = 41.05\%$. We use this coefficient in our preferred specification in the main estimation of trade costs in Table 3.5 as the restriction. Besides this RTA coefficient which we expected to have a positive effect on bilateral trade flows, Table 3.5 reports expected qualitative and quantitative coefficients for the other gravity variables. Again, for our preferred specification in column (4) we find the logarithm of distance between two countries to have a negative and significant effect on bilateral trade flows, while two contiguous countries and countries which share a common language trade more with each other. For our sample of countries for the year 2005, the dummy which indicates a common colonial history of countries turns out to have a negative but only slightly significant effect on trade flows.

Turning to the auxiliary estimation of the migration gravity from which we estimate the the coefficient for the European Union indicator, we also find an expected result. All else equal, country-pairs joining the European Union observe on average $(exp(0.760) - 1) * 100\% = 113.83\%$ more bilateral migration. This effect is estimated on a reduced sample of destination countries, since the variance-covariance matrix for the full set of countries turns out to be highly non-singular.[11] And again for the main estimation of the migration

[9] See http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators.

[10] The included countries are: Australia, Austria, Belgium, Canada, Chile , Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Republic of, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and the United States.

[11] The 15 destination countries are Australia, Belgium, Canada, Denmark, Finland, Germany, Italy, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, United States.

Table 3.4.: Auxiliary Trade Gravity Estimation for Years 1948 to 2006

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) PPML
log(Distance)	0.0279**	-1.088***		
	(0.0112)	(0.00964)		
RTA	2.258***	0.337***	0.215***	0.344***
	(0.0217)	(0.0152)	(0.0136)	(0.0135)
Contiguity	2.132***	0.215***		
	(0.0440)	(0.0222)		
Common Language	0.680***	0.201***		
	(0.0434)	(0.0191)		
Colony	0.896***	0.562***		
	(0.0620)	(0.0248)		
Observations	63,395	63,395	63,395	66,110
Exporter-Year FE	No	Yes	Yes	Yes
Importer-Year FE	No	Yes	Yes	Yes
Country-pair FE	No	No	Yes	Yes
Zeros included	No	No	No	Yes
Sample	OECD-EU	OECD-EU	OECD-EU	OECD-EU

Notes: Dependent variable for OLS columns is the log of trade flows from country i to country j, $\ln X_{ij}$. Dependent variable for PPML columns is trade flows in levels, M_{ij}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

Table 3.5.: Trade Gravity Estimation for the Year 2005

VARIABLES	(1) OLS	(2) OLS	(3) PPML	(4) PPML
log(Distance)	-1.318***	-1.341***	-0.833***	-0.882***
	(0.0519)	(0.0608)	(0.0605)	(0.0424)
RTA	0.459***	0.344	0.489***	0.344
	(0.129)	(0)	(0.128)	(0)
Contiguity	0.381***	0.364**	0.390***	0.374***
	(0.115)	(0.141)	(0.0744)	(0.0742)
Common Language	0.0567	0.0698	0.142	0.167*
	(0.110)	(0.139)	(0.0963)	(0.0949)
Colony	0.308**	0.292	-0.149	-0.203*
	(0.132)	(0.183)	(0.109)	(0.106)
Observations	1,558	1,558	1,560	1,560
Exporter FE	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes
Zeros included	No	No	Yes	Yes
Sample	OECD-EU	OECD-EU	OECD-EU	OECD-EU

Notes: Dependent variable for OLS columns is the log of trade flows from country i to country j, $\ln X_{ij}$. Dependent variable for PPML columns is trade flows in levels, M_{ij}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

gravity in Table 3.7, we do find expected effects. The negative effect of bilateral distance on bilateral migration flows is repeated for the migration gravity as is the positive effect of contiguity and common language. The colonial history indicator has a positive and significant effect in the migration gravity.

Table 3.6.: Auxiliary Migration Gravity Estimation for Years 2000 to 2012

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) PPML
log(Distance)	-0.812***			
	(0.0231)			
European Union	0.156***	0.457***	0.723***	0.760***
	(0.0431)	(0.0411)	(0.0503)	(0.0670)
Contiguity	0.146***			
	(0.0520)			
Common Language	0.757***			
	(0.0428)			
Colony	0.773***			
	(0.0597)			
Observations	12,472	12,472	7,054	7,089
Year FE	No	No	No	No
Origin-Year FE	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	Yes	Yes
Zeros included	No	No	No	Yes
Sample	OECD-EU	OECD-EU	Reduced	Reduced

Notes: Dependent variable for OLS columns is the log of migration flows from country i to country j, $\ln X_{ij}$. Dependent variable for PPML columns is migration flows in levels, M_{ij}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

3.6. Comparative Static Quantification

With observed measures for model parameters from the data (Section 3.4) and theory consistent estimated trade and migration costs (Section 3.5), we are now able to use the model for comparative static calculations. As for the numerical example in Section

The same set of destination countries is used in Ortega and Peri (2013).

Table 3.7.: Migration Gravity Estimation for the Year 2010

VARIABLES	(1) OLS	(2) OLS	(3) PPML	(4) PPML
log(Distance)	-0.784***	-0.651***	-0.574***	-0.589***
	(0.0840)	(0.0743)	(0.107)	(0.0919)
European Union	0.0842	0.760	0.858***	0.760
	(0.156)	(0)	(0.302)	(0)
Contiguity	0.165	0.242	0.500**	0.500***
	(0.188)	(0.193)	(0.217)	(0.217)
Common Language	0.694***	0.699***	0.586***	0.578***
	(0.151)	(0.149)	(0.220)	(0.223)
Colony	0.718***	0.821***	0.532***	0.518***
	(0.192)	(0.193)	(0.205)	(0.208)
Observations	1,095	1,095	1,145	1,205
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Zeros included	No	No	No	No
Sample	OECD-EU	OECD-EU	OECD-EU	OECD-EU

Notes: Dependent variable for OLS columns is the log of migration flows from country i to country j, $\ln X_{ij}$. Dependent variable for PPML columns is migration flows in levels, M_{ij}. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For information on the different samples see text.

3.2.6, we can solve the model for given information on Y_i, N_i, $t_{ij}^{(1-\sigma)}$, δ_{ji}^{-1}, and σ. We use GDPs of 2005 as a measure for Y_i, total population of a country in 2005 measures N_i, while trade and migration costs, $t_{ij}^{(1-\sigma)}$ and δ_{ji}^{-1}, are estimated as described in Section 3.5. We set $\sigma = 5$ as in Anderson and van Wincoop (2003). We provide results for two different counterfactual scenarios. In the first scenario we evaluate the welfare effects of TTIP if migration is infinitely costly and therefore workers are immobile and compare them to welfare effects if migration costs are as estimated. In a second scenario we highlight differences between the standard TTIP scenario and a TTIP scenario where we additionally reduce migration frictions.

3.6.1. Results

For the first counterfactual exercise, we hypothetically change the true RTA vector to one where an agreement would be signed between the European Union and United States in addition to existing RTAs in 2005. We focus on how workers mobility would change the results. Therefore we compare the welfare results once if migration involves infinite migration costs to the situation where we plug in estimated migration costs according to Table 3.7. Table 3.8 reports the results for the selection of 36 EU-OECD countries where the upper panel provides the results for non-signing countries and the lower panel for signing countries. The first column reports the change in welfare measured as real income per labor force if migration costs are infinite. We allow for migration in the second column and compare the welfare changes in the third column. The last two columns report the changes in total trade respectively. First of all, we observe the usual qualitative picture from new quantitative trade models that countries that do not take part directly in the trade liberalization suffer in terms of welfare and signing countries increase their welfare. The same is true for the effects if workers are potentially mobile but with different quantitative results. For all non-signing countries the negative third country effects of TTIP are lower if migration is possible. The reduction of the negative welfare effects vary across these countries from around 4 up to 40 percent. For the signing countries, allowing workers to be mobile is a mixed blessing. Still, all signing countries win from signing TTIP but some are better off and some loose if migration is introduced in the model. The percent changes in the welfare changes vary here from around -8 to 9 percent. Note that at the same time all signing countries' total trade increases with migration. So the potential for welfare increasing trade effects are even higher if we allow for migration. But this potential cannot be exploited by all countries due to either immigration which reduces welfare measured as real income per capita directly and/or non-trivial changes in prices induced by migration.

Table 3.9 presents the results for the comparison of the above TTIP scenario with estimated migration costs to a TTIP scenario where we additionally reduce migration costs by the estimated effect of the EU indicator. So, we simulate the welfare effects from TTIP if it would have been signed with the free movement of workers as within the European Union. We call this scenario TTIPplus. The additional reduction of migration costs is a mixed blessing for non-signers. While some of these countries' negative welfare effects are lower, for some other the negative effects are even increased. This picture is repeated for the signing countries. The qualitative result of the trade liberalization holds, that means all signers still win in terms of welfare from signing TTIP but some winners win less and some more if migration frictions are reduced additionally. The exact quantitative change depends on non-trivial changes in prices, the multilateral migration and trade costs pattern, the resulting change in L_i.

3.7. Conclusion

We present in this paper a new quantitative trade frameworks where labor is immobile. We provide a structurally estimable trade model where we include an explicit and endogenous migration decision at the individual level. From the model we derive two structural gravity equations which we use to infer trade and migration costs empirically. We rigorously estimate the model and account for multilateral resistance to trade and to migration. Trade and migration interact via price effects which lead to quantitatively different comparative static effects if we use the model to ex ante evaluate the negotiated Transatlantic Trade and Investment Partnership between the European Union and the United States. The general equilibrium welfare changes if migration is allowed in such a model are subtle. A further investigation on the exact relationship between trade and migration has yet to be provided. So far, we conclude that considering the mobility of workers in new quantitative trade models change welfare implications quantitatively.

Table 3.8.: Comparison of Welfare Effects from Signing TTIP for 36 EU and OECD Countries

	Welfare			Trade	
	$\%\Delta(w_i/P_i)$ with $\delta_{ij}=\infty$	$\%\Delta(w_i/P_i)$ with $\delta_{ij}=\hat{\delta}_{ij}$	$\%\Delta\%\Delta(w_i/P_i)$	$\%\Delta X_i$ with $\delta_{ij}=\infty$	$\%\Delta X_i$ with $\delta_{ij}=\hat{\delta}_{ij}$
Non-Signers					
Australia	-0.40	-0.36	-9.18	-0.59	-0.54
Canada	-0.95	-0.74	-21.82	-0.34	-0.79
Chile	-0.66	-0.56	-13.95	-0.35	-0.47
Iceland	-0.75	-0.53	-29.43	-0.21	-0.82
Israel	-0.53	-0.42	-20.63	-0.47	-0.73
Japan	-0.12	-0.12	-3.92	-0.98	-0.82
Korea, Republic of	-0.15	-0.14	-8.31	-0.28	-0.19
Mexico	-0.87	-0.75	-13.98	-0.40	-0.54
New Zealand	-0.33	-0.28	-15.82	-0.16	-0.25
Norway	-0.78	-0.57	-27.70	-0.34	-0.85
Switzerland	-0.75	-0.53	-29.77	-0.28	-0.92
Turkey	-0.67	-0.53	-21.40	-0.38	-0.66
Signers					
Austria	0.87	0.89	2.54	0.70	0.91
Belgium	0.56	0.61	9.04	0.54	0.58
Czech Republic	0.99	1.00	0.59	0.69	0.97
Denmark	1.24	1.20	-3.08	1.00	1.45
Estonia	2.21	2.02	-8.50	1.22	2.35
Finland	2.27	2.09	-7.81	1.74	2.78
France	0.77	0.80	3.35	0.97	1.15
Germany	1.01	1.01	-0.32	1.58	1.83
Greece	1.87	1.73	-7.42	1.84	2.69
Hungary	1.40	1.34	-3.99	0.97	1.50
Ireland	1.91	1.79	-6.53	1.41	2.25
Italy	1.37	1.33	-2.86	1.82	2.22
Netherlands	0.75	0.78	3.31	0.83	0.98
Poland	1.46	1.41	-3.14	1.10	1.60
Portugal	2.25	2.11	-6.21	1.60	2.54
Slovakia	1.30	1.26	-3.09	0.81	1.29
Slovenia	1.04	1.03	-0.31	0.68	1.02
Spain	1.49	1.44	-2.81	1.73	2.21
Sweden	1.88	1.76	-6.50	1.46	2.23
United Kingdom	1.27	1.25	-1.06	2.48	2.85
United States	1.39	1.36	-2.13	11.44	11.59

Notes: Welfare and trade effects for 36 EU and OECD countries from signing TTIP comparing infinite migration costs and estimated migration costs.

Table 3.9.: Comparison of Welfare Effects from Signing TTIP to Signing TTIPplus for 36 EU and OECD Countries

	Welfare			Trade	
	$\%\Delta(w_i/P_i)$ of TTIP	$\%\Delta(w_i/P_i)$ of TTIPplus	$\%\Delta\%\Delta(w_i/P_i)$	$\%\Delta X_i$ of TTIP	$\%\Delta X_i$ of TTIPplus
Non-Signers					
Australia	-0.36	-0.62	72.36	-0.54	-0.93
Canada	-0.74	-1.34	80.68	-0.79	-2.47
Chile	-0.56	-0.94	66.25	-0.47	-0.54
Iceland	-0.53	-0.28	-46.53	-0.82	-1.07
Israel	-0.42	-0.37	-11.96	-0.73	-0.69
Japan	-0.12	-0.17	45.02	-0.82	-0.96
Korea, Republic of	-0.14	-0.20	40.66	-0.19	-0.01
Mexico	-0.75	-1.55	108.08	-0.54	-0.93
New Zealand	-0.28	-0.43	51.52	-0.25	-0.59
Norway	-0.57	-0.16	-71.74	-0.85	-0.47
Switzerland	-0.53	-0.10	-80.63	-0.92	-0.42
Turkey	-0.53	-0.37	-30.71	-0.66	-0.03
Signers					
Austria	0.89	0.65	-26.88	0.91	3.53
Belgium	0.61	0.53	-12.93	0.58	3.33
Czech Republic	1.00	0.75	-25.43	0.97	3.68
Denmark	1.20	0.74	-38.55	1.45	5.09
Estonia	2.02	1.22	-39.49	2.35	7.00
Finland	2.09	1.44	-31.14	2.78	6.66
France	0.80	0.60	-24.66	1.15	3.95
Germany	1.01	1.07	5.93	1.83	3.69
Greece	1.73	1.43	-17.09	2.69	4.95
Hungary	1.34	1.03	-22.98	1.50	4.34
Ireland	1.79	0.97	-45.76	2.25	8.11
Italy	1.33	1.29	-3.57	2.22	3.99
Netherlands	0.78	0.57	-27.17	0.98	4.07
Poland	1.41	1.38	-2.32	1.60	3.27
Portugal	2.11	1.63	-22.72	2.54	5.83
Slovakia	1.26	0.94	-25.28	1.29	4.24
Slovenia	1.03	0.57	-44.45	1.02	4.61
Spain	1.44	1.23	-14.72	2.21	4.83
Sweden	1.76	1.22	-30.87	2.23	5.84
United Kingdom	1.25	0.68	-45.49	2.85	7.19
United States	1.36	1.73	27.25	11.59	9.87

Notes: Welfare and trade effects for 36 EU and OECD countries from signing TTIP comparing signing of TTIP to signing of TTIP with an additional reduction in migration costs.

4. Heterogeneous Firms, Globalization and the Distance Puzzle[1]

4.1. Introduction

> *"From the telegraph to the Internet, every new communications technology has promised to shrink the distance between people, to increase access to information, and to bring us ever closer to the dream of a perfectly efficient, frictionless global market." ((Friedman, 2005), p. 204)*

The many facets of globalization like the increased trade in final goods, intermediate inputs and services, or the increased international mobility of capital and labor, are perceived to bring countries closer together, shrinking the impediments of distance. However, gravity estimations regressing bilateral trade on distance, inter alia, tell us the opposite. Disdier and Head (2008) undertake a meta analysis of the magnitude of the distance coefficient based on 103 empirical studies and find that (i) the mean effect of the distance coefficient is about $|-0.9|$ across studies, and (ii) the negative impact of distance on trade rose around the middle of the century and has remained persistently high ever since.[2]

A stable or rising distance coefficient over time is puzzling because the distance coefficient has the structural interpretation of the elasticity of bilateral trade with respect to distance (e.g. Anderson and van Wincoop (2003)). Transport technology is known to be biased in favor of long distances (see Hummels (2007)), which should lead to a decrease of the

[1] This chapter bases on joint work with Mario Larch, Pehr-Johan Norbäck and Dieter Urban. A version of this chapter is published as Larch et al. (2015). All remaining errors in this volume are mine.

[2] This paper also provides a good collection of references for the "distance puzzle". Hence, we here dispense with a discussion of all relevant papers and with providing all references.

distance effect. Hence, the elasticity of bilateral trade with respect to distance should fall with increasing globalization.

In this paper, we use the recently developed framework from Helpman et al. (2008), henceforth HMR, which accounts for zero trade flows and firm heterogeneity. To explain the increasing OLS bias, we formally derive how the bias of the OLS distance coefficient evolves over time if the true data generating process is the HMR model and the elasticity of trade with respect to distance decreases during globalization through, for instance, improved transport and communication technologies.[3]

OLS estimates suffer from two biases. First, there is a *sample selection bias* because bilateral trade is measured as a logarithm and zero values of bilateral trade turn into missing values. As small or distant countries are more likely to have small trade flows, measurement errors in export flows will more likely lead to zero trade flows for those countries. This leads to a positive correlation of the error term with distance, causing a downward bias in the distance coefficient, i.e. the value of the distance coefficient is too small in absolute terms. Hence, accounting for zero trade flows does not explain the large distance coefficients of OLS.

Second, there is an *omitted variable bias* from ignoring that firms are heterogeneous in productivity. If an index of the number of exporting firms in an industry is not included as a control in the gravity estimation, then it appears in the regression error causing a negative correlation between error and distance because there are fewer exporters to more distant destinations. Hence, the distance coefficient is upward biased through omitting a control on firm productivity, i.e. the value of the distance coefficient is too large in absolute terms. As the sample selection bias and the omitted variable bias work in opposite directions, the overall bias from OLS estimates is ambiguous theoretically.

Assuming that globalization can be associated with a fall in the elasticity of trade with respect to distance, we show how the two biases evolve over time. We first show that ceteris paribus, the downward bias through sample selection, must decrease over time. Intuitively, as trade costs decrease, ever fewer country-pairs have zero trade flows and eventually all countries trade with each other. However, then the sample selection bias

[3]We follow this interpretation of the distance coefficient throughout the paper. Assuming decreasing distance costs would lead to a *flatter world* without relative differences of trade volumes across trading partners w.r.t to distance. However, Buch et al. (2004) argue that the distance puzzle is not that puzzling when the effect of distance is interpreted in absolute terms. Under the assumption of linear dependency of trade costs with respect to distance, they show that a potential decline in the impact of distance would be caught by the constant term in the gravity equation. But still we should—but do not—observe a decline in the relative impact of distance on bilateral trade, which is exactly measured by the elasticity we look at.

disappears, i.e. the distance coefficient rises. We then show that, ceteris paribus, the upward bias from omitting the number of exporting firms also becomes smaller over time when the elasticity of trade with respect to distance falls. Intuitively, at a lower trade elasticity, most firms will export, reducing the upward bias, i.e. the distance coefficient decreases. Thus, globalization – as captured by a smaller impact of distance – has an ambiguous effect on the bias of OLS in general as both biases decrease with a fall in the elasticity of trade with respect to distance.

To investigate the empirical success of the HMR estimator in solving the distance puzzle, we use three different trade data sets, two aggregate and one at industry-level, over different time periods.[4] We find that the HMR estimates of the distance coefficient (in absolute value) are decreasing on average over time as expected.

Having empirically shown that the HMR estimator does produce decreasing distance coefficients over time, we compare the outcome with OLS estimates. We first confirm the finding of HMR that OLS produces larger distance coefficients (in absolute value). More importantly, we show that these distance coefficients increase over time. Hence, the distance puzzle arises due to the fact that *the bias of OLS increases over time.* Employing the HMR estimator instead of OLS solves the distance puzzle.

We then disentangle the estimated OLS bias in its two components, the omitted variable bias and the sample selection bias. From theory, if globalization only reduces the elasticity of distance on trade, this is consistent with the downward bias from sample selection decreasing faster than the upward bias from not controlling for the number of exporting firms. Thus, the HMR model suggests that the distance puzzle arises from firm heterogeneity having become relatively more important over time. When decomposing the bias terms empirically, we find evidence on the importance of firm heterogeneity. Contrary to the predictions of changes of the distance elasticity over time, the omitted variable bias increases over time. We show that this empirical finding is in line with the theoretical model if globalization not only decreases the elasticity of trade with respect to distance, but also increases the elasticity of trade with respect to firm heterogeneity over time. This is also nicely in line with empirical evidence provided by Poschke (2014) suggesting that as countries develop the distribution of firm sizes becomes more dispersed. We finally show that the estimated coefficients from HMR are also strongly correlated with the time patterns in freight costs reported by Hummels (2007) and Brun et al. (2005), which in turn depend on fluctuations of oil prices.

[4] Berthelon and Freund (2008) document the distance puzzle on bilateral industry data rather than on bilateral country data.

For future work, we also suggest a linearization of the HMR estimator, which is comparable to the non-parametric approach of Helpman et al. (2008). This approach is easy to implement with standard econometric programs because it is estimable via OLS. We show that such a simplified estimator performs just as well as the original non-linear least squares version.

We also show that a Heckman estimator deviates from the HMR estimates and produces bigger distance coefficients and increasing differences to the OLS estimates over time. The Heckman correction results lead to the conclusion that taking into account zero trade flows cannot solve the distance puzzle, as expected from our theoretical results.

Alternative attempts to solve the distance puzzle stem from Felbermayr and Kohler (2006) using Tobit estimates to take zero trade flows into account.[5] Other studies explain why the substitution elasticity may have been rising over time (Glaeser and Kohlhase, 2004; Krautheim, 2012; Lawless and Whelan, 2007; Berthelon and Freund, 2008), possibly overcompensating for the fall in trade costs, which both determine the distance coefficient in theory. Duranton and Storper (2008) provide an alternative model to rationalize rising overall trade costs besides falling transport costs. They assume vertically linked industries in which the quality of inputs is not contractible and where providing a given level of quality to suppliers becomes more costly with distance. Their main finding is that lower transport costs imply that higher quality inputs are traded in equilibrium, and the effect of this higher quality is that there is an increase in trade costs. Yotov (2012) proposes to measure the effects of distance on international trade relative to the effects of distance within national borders as a simple and useful solution to the distance puzzle. He finds a drop in the impact of distance on trade of roughly 50% from the mid-sixties to 2005. Finally, using bilateral country data for the year 1986, HMR find that their estimated distance coefficient represents a drop of roughly one third as compared to

[5] There is ample evidence from micro-data for particular countries that the extensive margin matters. Bernard et al. (2006) use firm-level data to distinguish the entry and exit of firms into and out of exporting (extensive margin) from the export volumes of exporting firms (intensive margin). They find that a reduction in trade costs may increase industry productivity through changes on the extensive margin. Hummels and Klenow (2005) use disaggregated product-level data to distinguish between the variety dimension (extensive margin) and the quality as well as the quantity dimension (intensive margin). One of their main results is that adverse terms-of-trade effects occur more frequently if growth takes place mainly at the extensive margin. Similarly, Baldwin and Harrigan (2011) use product-level data on bilateral U.S. exports demonstrating that a large part of potential export flows are zero, and showing that the incidence of these zero export flows is strongly correlated with distance and importing country size. Hillberry and Hummels (2008) analyze trade at the five-digit zip codes and decompose the extensive and intensive margins of shipments. Their main finding is that distance reduces aggregate trade values primarily by reducing the number of commodities shipped and the number of shipping establishments. However, the extensive margin is important over very short distances.

OLS. However, HMR do not examine the evolution of the distance coefficient over time. Hence, none of the mentioned papers discuss the role of the omitted variable problem of firm heterogeneity in creating an increasing bias over time which, in contrast, is the contribution of our paper.

The remainder of the paper is organized as follows. Section 4.2 is divided into 2 subsections where we in Subsection 4.2.1 derive the gravity equation controlling for zero trade flows and firm-level heterogeneity. This is done following HMR, and then in Subsection 4.2.2 we calculate the biases of OLS estimates. Section 4.3 presents our estimation equation in Subsection 4.3.1, describes the data in Subsection 4.3.2, and gives the results in Subsection 4.3.3. The last section concludes.

4.2. HMR and the Distance Puzzle

In this section, we use the HMR model to examine the distance puzzle. We will assume the HMR model to be the data generating process and examine to what extent the OLS estimates are biased and in what direction this bias goes. Then, we will examine how the bias of OLS is affected by globalization.

4.2.1. The Gravity Equation from HMR

The HMR model is a multi-country monopolistic competition model with heterogeneous firms and identical consumers with CES "love-of-variety"utility functions à la Dixit and Stiglitz (1977). HMR assume that firm productivity, $1/a$, follows a truncated Pareto distribution, $G(a) = (a^k - a_L^k)/(a_H^k - a_L^k)$, where $k > (\varepsilon - 1)$ is the shape parameter and a_L and a_H are the lower (highest productivity) and upper support (lowest productivity). HMR obtain the following gravity equation (their Equation (9)):

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \omega_{ij} + u_{ij}, \tag{4.1}$$

where m_{ij} is logged aggregate imports of country i from country j. $\beta_0 = (\varepsilon - 1)\ln(\alpha) + \ln(\psi)$, where ε is the substitution elasticity between any two varieties, $1/\alpha = \varepsilon/(\varepsilon - 1)$, and $\psi = ka_L^{k-\varepsilon+1}/[(k-\varepsilon+1)(a_H^k - a_L^k)]$. Exporter country-fixed effects, $\lambda_j = (1-\varepsilon)\ln c_j + n_j$, contain the country-specific minimum cost of a bundle of inputs in country j, c_j, and the log of the number of firms from country j is denoted by n_j. Importer country-fixed effects, $\chi_i = (\varepsilon - 1)p_i + \ln \mu_i + y_i$, p_i denote the logged consumer price index in i, μ_i

denote the (constant) share of income spent by consumers of country i, and y_i is the income in country i. d_{ij} is the log of the distance between i and j and γ the elasticity of bilateral trade with respect to distance. ω_{ij} captures the number of exporters from j exporting to i given by $\omega_{ij} = \ln[(a_{ij}/a_L)^{k-\varepsilon+1} - 1]$, where a_{ij} denotes the inverse of the cutoff productivity level of exporting firms. Note that ω_{ij} is the only new in the gravity equation as compared to Anderson and van Wincoop (2003). u_{ij} is an iid remainder error term with variance σ_u^2.

The estimation of (4.1) is hampered by two problems. First, it is only estimated on data with positive trade flows since the dependent variable, the log of trade volume (m_{ij}), is not defined for zero import values. Second, there is an omitted variable problem through ω_{ij} which captures the degree of firm heterogeneity in country j, information which is typically not available for gravity estimations on world trade data sets.[6]

HMR note that both problems are related to the extensive margin of trade. They use the zero-profit condition for exporting from country j to country i and define a latent variable for the cutoff productivity for positive exports, z_{ij}:

$$z_{ij} = E\left[z_{ij}\left|d_{ij,}\xi_j, \zeta_i, \phi_{ij}\right.\right] + \eta_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij}, \tag{4.2}$$

where γ_0 collects constant terms, $\xi_j = -\varepsilon \ln c_j - \phi_{EX,j}$ is an exporter fixed effect capturing, in addition to c_j, a measure of fixed export costs common across all export destinations ($\phi_{EX,j}$). $\zeta_i = (\varepsilon - 1)\, p_i + y_i + \ln \mu_i - \phi_{IM,i}$ is an importer fixed effect that captures, besides the consumer price index, income and income shares, a fixed trade barrier imposed by the importing country on all exporters ($\phi_{IM,i}$). ϕ_{ij} is an observed measure of any additional country-pair specific fixed trade costs and κ the corresponding parameter. $\eta_{ij} = u_{ij} + \nu_{ij}$, where $\nu_{ij} \sim N(0, \sigma_\nu^2)$ is an error term in the fixed trade costs specification. σ_η^2 is the variance of η_{ij}. Using this latent variable, the omitted variable of the number of exporting firms, ω_{ij}, can be expressed as

$$\omega_{ij} = \ln\left[\exp\left[\delta z_{ij}\right] - 1\right], \tag{4.3}$$

where $\delta = (k - \varepsilon + 1)\,/\,(\varepsilon - 1)$.

While the latent variable, z_{ij}, cannot be observed, one can observe if trade takes place. Thus, an indicator variable, $T_{ij} = \mathcal{I}_{[z_{ij}>0]}$, can be defined from which the selection equa-

[6]Flam and Nordström (2011) have recently included a proxy variable for ω_{ij}, which is available for Swedish exports. However, they did not estimate the distance coefficient over time, which is the focus of this paper.

tion for the probability of strictly positive exports is obtained:

$$\begin{aligned}\Pr\left(T_{ij}=1\left|d_{ij},\xi_j^*,\zeta_i^*,\phi_{ij}\right.\right) &= \Pr\left(z_{ij}^*>0\left|d_{ij},\xi_j^*,\zeta_i^*,\phi_{ij}\right.\right) \\ &= \Pr\left(\gamma_0^*+\xi_j^*+\zeta_i^*-\gamma^*d_{ij}-\kappa^*\phi_{ij}>-\eta_{ij}^*\left|d_{ij},\xi_j^*,\zeta_i^*,\phi_{ij}\right.\right) \\ &= \Phi\left(\gamma_0^*+\xi_j^*+\zeta_i^*-\gamma^*d_{ij}-\kappa^*\phi_{ij}\right) \\ &= E\left[z_{ij}^*\left|d_{ij},\xi_j^*,\zeta_i^*,\phi_{ij}\right.\right],\end{aligned} \tag{4.4}$$

where $\Phi(.)$ is the cumulative distribution function of the unit normal distribution and every starred coefficient represents the original coefficient divided by σ_η.[7]

One can now in a first stage estimate (4.4) by a probit estimation. Inverting the predicted probability from (4.4) yields an estimate of the underlying latent variable, $\hat{z}_{ij}^*$.

Defining $\delta=\sigma_\eta(k-\varepsilon+1)/(\varepsilon-1)>0$, HMR use $\hat{\omega}_{ij}^*\equiv\ln\left\{\exp\left[\delta\left(\hat{z}_{ij}^*+\hat{\bar{\eta}}_{ij}^*\right)\right]-1\right\}$ as an estimate for $E[\omega_{ij}|.,z_{ij}^*>0]$,[8] where $\hat{\bar{\eta}}_{ij}^*=\phi(\hat{z}_{ij}^*)/\Phi(\hat{z}_{ij}^*)$ is the inverse Mills ratio from the first-stage probit estimation, which itself is well-known to be a consistent estimate of $E[u_{ij}|.,z_{ij}^*>0]$.[9] Inserting these terms into (4.1), HMR show that estimation of the gravity model requires estimation of the following specification:

$$m_{ij}=\beta_0+\lambda_j+\chi_i-\gamma d_{ij}+\ln\left\{\exp\left[\delta\left(\hat{z}_{ij}^*+\hat{\bar{\eta}}_{ij}^*\right)\right]-1\right\}+\beta_{u\eta}\hat{\bar{\eta}}_{ij}^*+e_{ij}, \tag{4.5}$$

where $\beta_{u\eta}\equiv corr(u_{ij},\eta_{ij})(\sigma_u/\sigma_\eta)=corr(u_{ij},u_{ij}+\nu_{ij})(\sigma_u/\sigma_\eta)>0$. The term $\ln\{\exp[\delta(\hat{z}_{ij}^*+\hat{\bar{\eta}}_{ij}^*)]-1\}$ corrects for the omitted variable ω_{ij} in the presence of sample selection[10] and $\beta_{u\eta}\hat{\bar{\eta}}_{ij}^*$ is the well-known correction of the error term u_{ij} in the presence of sample selection. As a result, e_{ij} is an i.i.d. error term satisfying $E[e_{ij}|.,T_{ij}=1]=0$. Therefore, one can estimate (4.5) using NLS and obtain an estimate of the distance coefficient, γ, having the structural interpretation of the elasticity of bilateral trade with respect to distance for all country-pairs in the population, i.e. for positive and zero trade flows.

[7] As in every discrete choice model, the scale can be arbitrarily chosen, i.e. the model must be properly normalized. We normalize by dividing through σ_η, following HMR. This leads the error term $\eta_{ij}^*=\eta_{ij}/\sigma_\eta$ to be distributed unit normal.

[8] Santos Silva and Tenreyro (2006) and Santos Silva and Tenreyro (2015) note that this is not a consistent estimate because of Jensen's inequality. However, Santos Silva and Tenreyro (2015) also note that it is a reasonably accurate approximation in many practical situations. The similarity of our results from the linear approximation of HMR below supports this claim (see Section 4.3.3).

[9] This term is also known as Heckman's lambda (Heckman, 1979).

[10] In the absence of a sample selection bias but in the presence of the omitted variable bias, the correction term would simplify to $\ln\left\{\exp\left(\delta\hat{z}_{ij}^*\right)-1\right\}$, since plim$\hat{\bar{\eta}}_{ij}^*=E[u_{ij}|.,z_{ij}^*>0]=0$ in this case.

4.2.2. The Bias of OLS

Let us now start to examine the properties of an OLS estimate of the distance coefficient, $\hat{\gamma}^{OLS}$, from estimating gravity Equation (4.1) without a sample selection correction and when not controlling for the omitted variable bias due to firm heterogeneity by ω_{ij}. To gain some intuition on these two biases and their direction, we first look at them individually before considering them simultaneously. We begin by discussing the sample selection bias and then continue with the omitted variable bias.

Selection Bias By taking logs of imports, all zero trade flows are omitted from the sample. This is the selection bias. The effect on the estimates of the distance elasticity are summarized in the following Lemma:

Lemma 1. *The selection bias resulting from ignoring zero values of bilateral trade leads to an underestimation of the elasticity of bilateral trade with respect to distance.*
Proof see Appendix C.

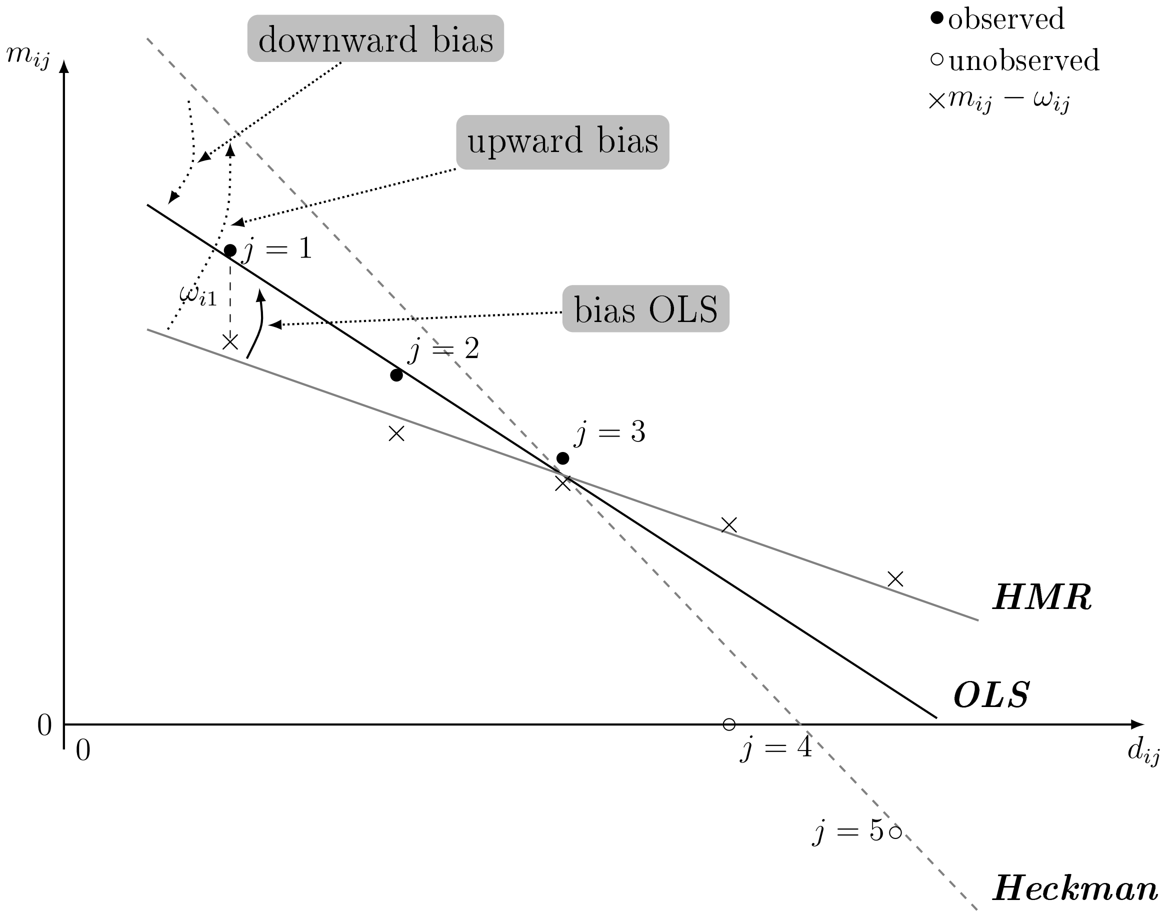
Intuitively, this result is due to the fact that distant countries are more likely to have small trade flows. Hence, measurement errors will more likely lead to zero trade flows for those distant countries. Countries that are distant but remain in the sample will have positive measurement errors, leading to a positive correlation between distance and the error term. This explains the downward bias in the distance coefficient of the selection bias, i.e. a too small value of the distance coefficient in absolute terms.

We illustrate this result by using Figure 4.1 which contains distance, d_{ij}, on the horizontal axis and imports, m_{ij}, on the vertical axis. We depict by circles imports to country i from countries $j = 1, 2, 3, 4, 5$, holding the control variables constant over countries j for the purpose of graphical illustration. From the selection equation for the probability of strictly positive exports (4.4), we know that distance has a negative effect on the probability of exporting.[11] Thus, missing observations are more likely the larger is the distance. In addition, the smaller the error term, u_{ij}, the more likely trade is to be predicted to be zero. For this reason, we draw potential imports between countries i and $j = 4$ and $j = 5$ such that the distance is large and the error terms u_{i4} and u_{i5} are negative, causing these two observations to drop out of the sample, which we indicate by hollow circles. Since the negative u_{i4} and u_{i5} are not only contained in the selection Equation (4.4), but also in the gravity Equation (4.1), the imports that drop out do

[11] $\partial \Pr(T_{ij} = 1 \,|\cdot) / \partial d_{ij} = -\gamma^* \phi(\cdot) < 0$, where $\phi(\cdot)$ is the normal density function.

not only occur at a large distance but also at unusually low values of imports.[12] The non-missing imports at large distances, indicated by filled circles, are those with positive values of u_{ij}, i.e. $E[u_{ij}|d_{ij}, T_{ij} = 1] > 0$ if the distance, d_{ij}, is large.

Figure 4.1.: Illustrating the Bias of OLS



Notes: d_{ij} denotes distance between i and j and m_{ij} imports from j to i. ***HMR*** is given by $E\left[m_{ij} - \omega_{ij}|d_{ij}\right]$ with distance coefficient γ. ***Heckman*** is given by $E\left[m_{ij}|d_{ij}\right]$ and ***OLS*** by $E\left[m_{ij}|d_{ij}, T_{ij} = 1\right]$. The *bias OLS* corresponds to $Bias(\hat{\gamma}^{OLS}) = \gamma\delta - \Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}^*_{ij}$, where $(\gamma\delta)$ is denoted by ***upward bias*** and $(-\Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}^*_{ij})$ by ***downward bias*** in the figure. ω_{ij} controls for the omitted variable due to firm heterogeneity.

Since the unconditional expected value of u_{ij} is zero by construction of the OLS estimator,[13] i.e. $E[u_{ij}|T_{ij} = 1] = 0$, the conditional expected value of u_{ij} is negative,

[12] Note that we have drawn negative values of m_{ij}. Naturally, negative values of m_{ij} can never exist, but are generated by the gravity Equation (4.1), since shocks are, by assumption, normally distributed on a range from $-\infty$ to $+\infty$. However, whenever m_{ij} is negative, it is not observed.

[13] The estimated regression constant will always ensure that the unconditional expected value of the error term is zero in an OLS regression, whereas the conditional expected value of the error term is only zero for a correctly specified model, i.e. a model without endogeneity problems.

$E[u_{ij}|d_{ij}, T_{ij} = 1] < 0$, if the d_{ij} is small. But then the error term in the outcome equation u_{ij} and d_{ij} are positively correlated.

Lemma 2. *Accounting for zero trade flows will lead to distance coefficient estimates that are larger in absolute value.*
Proof see Appendix C.

As we have shown in Lemma 1, omitting zero trade flows leads to an underestimation of the distance elasticity. When now accounting for zero trade flows by employing a Heckman estimator, the resulting distance elasticities will be larger in absolute values.

This can be seen by decomposing the expected value of positive trade flows as follows:[14]

$$E[m_{ij}|d_{ij}, T_{ij} = 1] = E[m_{ij}|d_{ij}] + E[u_{ij}|d_{ij}, T_{ij} = 1], \tag{4.6}$$

where $E[m_{ij}|d_{ij}, T_{ij} = 1]$ is fitted by an OLS regression on the remaining three strictly positive import data points from $j = 1, 2, 3$ and $E[m_{ij}|d_{ij}]$ is fitted by an OLS regression on the entire population including $j = 4$ and $j = 5$. This in turn, is asymptotically equivalent to an OLS regression with a sample selection correction denoted by *Heckman* in Figure 4.1. Hence, the positive slope of $E[u_{ij}|d_{ij}, T_{ij} = 1]$ in d_{ij} results in a flatter declining slope of an OLS regression without sample selection, $E[m_{ij}|d_{ij}, T_{ij} = 1]$, than one with sample selection, $E[m_{ij}|d_{ij}]$.

Omitted Variable Bias Recent literature starting with Melitz (2003) highlights the importance of firm heterogeneity for explaining trade patterns. Ignoring empirically firm heterogeneity leads to an omitted variable bias. The effect on the estimates for the distance elasticity are summarized in the following Lemma:

Lemma 3. *The omitted variable bias resulting from ignoring firm heterogeneity leads to an overestimation of the elasticity of bilateral trade with respect to distance.*
Proof see Appendix C.

The intuition is that if an explicit control for the number of exporting firms is missing in the regression, it will end up in the error term. As there are fewer firms that export to more distant countries, there is a negative correlation between the error term and the distance. This results in an upward bias, i.e. the value of the distance coefficient

[14]See, e.g., Equation (16.34) in Cameron and Trivedi (2005), p. 549.

is too large in absolute terms. Note that this implies that accounting for the omitted variable bias will lead to lower distance coefficients in absolute values. A detailed formal derivation of this result is given in Appendix C.

We again illustrate the bias with help of Figure 4.1. Assume that ω_{ij} is known (and other controls are kept constant). Then a linear OLS regression of import values, m_{ij}, on distance, d_{ij}, controlling for ω_{ij} is like a regression of $(m_{ij} - \omega_{ij})$ on d_{ij}.[15] An OLS estimator fitting the regression line $E\left[m_{ij} - \omega_{ij}|d_{ij}\right]$ then has the same slope in d_{ij} as one fitting $E\left[m_{ij}|d_{ij}, \omega_{ij}\right]$ or, indeed, one using a consistent correction factor that controls for ω_{ij}, i.e. the *HMR* estimator (while at the same time controlling for the sample selection effect). To obtain $(m_{ij} - \omega_{ij})$ in Figure 4.1, indicated by crosses, we read off the difference between m_{ij} and ω_{ij}, an example of which is given for ω_{i1}. As can be seen, the crosses indicating $(m_{ij} - \omega_{ij})$ are systematically located below the circles indicating m_{ij} at small distances and above at large distances.[16] Hence, a fit of the crosses by the solid *HMR* line, $E\left[m_{ij} - \omega_{ij}|d_{ij}\right]$, rather than the circles by the Heckman line, $E\left[m_{ij}|d_{ij}\right]$, is flatter implying an upward bias of the distance coefficient. Hence, the OLS estimator omitting ω_{ij} overestimates the elasticity of bilateral trade with respect to distance.

Interacting the Two Biases Let us now consider both biases simultaneously, formally taking into account the interaction of the two. For this purpose, we need to draw on an approximation of (4.3),

$$\omega_{ij} \approx \delta z_{ij}, \tag{4.7}$$

where $\delta = \partial\omega_{ij}/\partial z_{ij}$ evaluated at the mean of z_{ij}.

We then have the following proposition:

Proposition 1. *When assuming that the HMR model is the data generating process, the OLS estimate of γ in (4.1) may then be (asymptotically) upward or downward biased, depending on whether the omitted variable bias from the share of exporting firms or the sample selection bias due to the omission of zero trade flows dominates, respectively.*
Proof see Appendix C.

[15] This follows from the fact that the regression coefficient of d_{ij} explains the remaining variation of the corresponding variable that is not at the same time common variation with another control variable (Frisch-Waugh theorem) and ω_{ij} enters the regression Equation (4.1) with coefficient one.

[16] Note that the *HMR* regression line fits all crosses for $j = 1, 2, 3, 4, 5$, because it does not only correct for the omitted variable bias, but also for sample selection simultaneously. If only the omitted variable bias was controlled for but not the sample selection bias, such a regression line would only fit the crosses corresponding to $j = 1, 2, 3$.

We derive the simultaneous bias term in Appendix C, which is given by the following simple expression:

$$Bias(\hat{\gamma}^{OLS}) = \gamma\delta - \Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}^{*}_{ij} \gtreqless 0, \tag{4.8}$$

where $\Xi = \sum_i \sum_j d_{ij} / \sum_i \sum_j (d_{ij})^2$.

Thus, as shown in Figure 4.1, the term $\gamma\delta > 0$ in (4.8) represents an upward bias in OLS (and Heckman) from not controlling for the number of exporting firms, and the last two terms measure a downward bias from sample selection in OLS, when omitting zero trade flows, as $\beta_{u\eta}$, $\bar{\eta}^{*}_{ij}$ and Ξ are positive. Overall, it is then indeterminate whether the OLS line $E\left[m_{ij}|d_{ij}, T_{ij} = 1\right]$ is flatter or steeper than the HMR line $E\left[m_{ij} - \omega_{ij}|d_{ij}\right]$. In anticipation of our empirical results, we have drawn it such that the OLS line is steeper than the HMR line, which implies that the omitted variable bias dominates the sample selection bias in levels. We depict this as *Bias OLS* in Figure 4.1.

4.2.3. Globalization

How would the bias of OLS evolve over time when globalization reduces the responsiveness of bilateral trade flows with respect to distance, due to new and better communication and transport technologies? Make the following assumption:

Assumption Increased globalization implies that $\frac{\partial\gamma}{\partial t} < 0$.

We then have the following proposition:

Proposition 2. *When assuming that the HMR model is the data generating process, both the downward bias from sample selection due to zero trade flows and the upward bias from omitting the number of exporting firms decrease in the pace of globalization, i.e. when the elasticity of bilateral trade with respect to distance (γ) falls over time.*
Proof see Appendix C.

The change in the bias of the distance coefficient $\partial\text{Bias}(\hat{\gamma}^{OLS})/\partial t$ can once more be understood intuitively looking at the two biases separately. Beginning with the change of the sample selection bias over time, we first notice that the bias depends on how the slope of $E[u_{ij}|d_{ij}]$ changes when γ changes over time. To understand this, we need to first look at how the selection process is influenced by a reduction in γ. An observation is missing whenever $z^{*}_{ij} < 0$ according to (4.4). Obviously, a reduction in γ decreases

z_{ij}^* $\left(\partial z_{ij}^*/\partial\gamma = -d_{ij} < 0\right)$, where some missing trade links turn positive. Eventually, all missing trade links turn into positive ones at sufficiently low γ. Hence, the true line fitting the data after globalization becomes flatter.

Turning to the change of the omitted variable bias over time, we once more need to understand how the slope of the conditional expectation function, $E[\omega_{ij}|d_{ij}]$, changes with a reduction of γ. For this purpose, it is sufficient to look at how ω_{ij} changes for each observation when γ falls. From (4.2) and (4.3) we immediately obtain

$$\frac{\partial\omega_{ij}}{\partial\gamma} = -d_{ij}\delta\frac{e^{\delta z_{ij}}}{e^{\delta z_{ij}}-1} < 0, \tag{4.9}$$

for all ω_{ij} that are non-missing. Hence, the share of exporting firms of a country j exporting to country i is increasing for each country-pair when γ falls. More importantly, this share increases less for increasingly distant trading partners:

$$\frac{\partial^2\omega_{ij}}{\partial\gamma\partial d_{ij}} = -\delta\frac{e^{\delta z_{ij}}}{e^{\delta z_{ij}}-1} - d_{ij}\gamma\delta^2\frac{e^{\delta z_{ij}}}{\left(e^{\delta z_{ij}}-1\right)^2} < 0, \tag{4.10}$$

for all ω_{ij} that are non-missing.

Since $\mathrm{E}[\omega_{ij}|d_{ij}]$ is flatter after globalization than before globalization, the upward bias in the distance coefficient from omitting the variable ω_{ij} also becomes smaller.

Considering changes in both biases simultaneously, we cannot tell whether the difference in slopes between the HMR-line and the OLS line will increase or decrease over time, because the downward bias from sample selection decreases and the upward bias from the omitted variable ω_{ij} also decreases. Since we cannot tell how the bias of OLS will behave under globalization, the OLS estimate of the distance coefficient may also increase or decrease over time.

The HMR Estimator and the Distance Puzzle Let us now show how the HMR estimator can be used to explain the distance puzzle. Suppose that the omitted variable bias dominates in levels at the beginning of the data period such that there is an overall upward bias in the distance coefficient (see the estimates of Helpman et al. (2008)), i.e. the OLS estimated schedule is steeper than the true line (HMR) just as in Figure 4.1. Then we have the following result:

Proposition 3. *If the downward bias from sample selection due to the omission of zero trade flows decreases faster than the upward bias from omitting the share of exporting*

firms, then, overall, the estimated OLS schedule will become steeper, explaining the distance puzzle. Controlling for both biases solves the distance puzzle.
Proof see Appendix C.

Hence, to capture the change of the bias in the distance coefficient, we need a larger decrease in the zero trade flows bias as compared to the omitted variable bias due to firm heterogeneity. A first glance at the data and anecdotal evidence cope with these facts. Whereas there has been a dramatic decrease in zero trade flows over the last two decades, firm sizes and productivities are still heavily dispersed (Poschke, 2014) and the share of exporting firms remains small.

Note also that the sample selection bias alone cannot solve the distance puzzle if the HMR model is the data generating process, as was suggested by Felbermayr and Kohler (2006) without being specific about the underlying data generating process. As the sample selection bias leads to a downward bias, the importance of distance will be underestimated. Hence, the level cannot be correctly captured accounting for sample selection alone.

4.3. Econometric Analysis

4.3.1. Base-line Estimation Equation and Alternative Estimators

Our baseline estimation equation is the HMR gravity Equation (4.5). Since our main interest rests on the coefficient of the distance variable, γ, and how it evolves over time, we will estimate this equation separately by year and industry. We use the following augmented specification:

$$m_{ij} = \beta_0 - \gamma d_{ij} + \alpha \mathbf{X}_{ij} + \lambda_j + \chi_i + \ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\right) - 1\right]\right\} + \beta_{u\eta}\hat{\bar{\eta}}^*_{ij} + e_{ij}, \quad (4.11)$$

where we explain the additional variables below. Once more, note that $\ln\Big\{\exp\Big[\delta\Big(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\Big) - 1\Big]\Big\}$ captures the omitted variable bias due to firm-level heterogeneity in the presence of sample selection, whereas $\hat{\bar{\eta}}^*_{ij}$ captures the sample selection bias of the error term from estimating (4.11) for non-zero trade. To estimate these correction terms, we add a first-stage equation in order to estimate (4.4), where:

$$z^*_{ij} = \varphi^*_0 - \gamma^* d_{ij} + \vartheta^* \mathbf{X}_{ij} + \varphi^*_1 COMM\_REL_{ij} + \varphi^*_2 COMM\_LANG_{ij} + \xi^*_j + \zeta^*_i + \eta_{ij}. \quad (4.12)$$

Other Estimators

We have shown that the distance puzzle can be studied by systematically comparing the estimates from HMR with corresponding estimates obtained with OLS. The OLS estimator estimates Equation (4.11), omitting the correction terms for firm-level heterogeneity and sample selection, i.e. excluding $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$ and $\hat{\bar{\eta}}^*_{ij}$. By comparing the HMR and OLS estimators, we can evaluate how the bias of OLS evolves over time as predicted by Propositions 1 and 2 . We will also compare our estimates with HMR with a number of other estimators.

Heckman The usual Heckman estimator estimates Equation (4.11) omitting the correction terms for firm-level heterogeneity but including that for sample selection, i.e. excluding $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$ but including $\hat{\bar{\eta}}^*_{ij}$.

Linear Approximation of HMR As δ enters the estimation equation non-linearly, we first estimate Equation (4.11) via non-linear least squares, as proposed by HMR. However, as discussed in Santos Silva and Tenreyro (2015), this correction term is biased if their theoretical model is the data generating process. However, for a wide range of $\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}$, the term $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$ may be well approximated by $\bar{\delta}\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)$ for some appropriate parameter $\bar{\delta}$, which can be estimated by OLS (see our discussion in Section 4.2.1). Hence, we also estimate the model via OLS and include $\varpi_{ij}=\bar{\delta}\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)$ instead of $\ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)-1\right]\right\}$.[17]

4.3.2. Data

The first of three data sets which we employ is borrowed from the original HMR paper (Helpman et al., 2008). Despite that HMR provide their main results for the year 1986, they also offer results for 1980s, adding year fixed effects to a panel. A comprehensive description of this data can be found in *Appendix I* in the HMR paper; the data are available at http://scholar.harvard.edu/helpman. The second data set is the standard CEPII gravity data set.[18] A full description can be found in the appendix of Head et al. (2010). The CEPII data enables us to explore the distance coefficients for a longer period than with the original HMR data set. Although the CEPII data set already starts in the

[17] HMR use a polynomial of degree 3 in the score variable in one of their robustness checks. We will point out that even a linear approximation works well in practice.

[18] See `http://www.cepii.fr/anglaisgraph/bdd/gravity.asp`.

1940s, due to the number of observations, we use data from 1980 to 2006 which is the latest available year. Thirdly, we use an industry-level data set where imports are taken from Nicita and Olarreaga (2001), who have compiled an industry data set corresponding to the 3-digit ISIC, revision 2, level that contains 28 manufacturing industries for up to 100 countries during 1976-2004. Because there is a large number of missing values in the early years and we are lacking a control variable in the last year, we have restricted the sample to 1978-2003. This data set is available for downloading from the World Bank (www.worldbank.org\trade). In turn, this data set draws its bilateral industry import data from COMTRADE of the UN which is based on the Standard International Trade Classification (SITC) and then transformed into ISIC. Production data are taken from UNIDO (International Yearbook of Industrial Statistics).

Dependent Variable

The dependent variable, m_{ij}, in Equation (4.11) is the natural logarithm of bilateral imports of country i from country j at a given year t; for the industry-level data additionally in a given industry l, measured in million US$ converted by the Penn World Tables 6.0 purchasing power parity exchange rate (PPP) and deflated by the U.S. consumer price index.

Explanatory Variables

The original HMR data set and the CEPII data set contain geographical information. The industry-level trade data set is merged into a balanced geography data set covering 170 countries. Thus, all three data sets contain geographical variables common to gravity estimations. These geography variables appear in Equations (4.11) and (4.12) and the different data sets as follows. Common to all data sets, d_{ij} is the log of the distance between countries i and j. λ_j and χ_i are full sets of exporter and importer dummy variables, respectively, controlling for, among others, the multilateral resistance terms pointed out by Anderson and van Wincoop (2003). $\mathbf{X}_{ij}$ contains a dummy variable indicating a common border between i and j in all data sets as well as an indicator for whether there is a common trade agreement between exporter i and importer j. Dummy variables for a common legal system, a common colonial history, a currency union and bilateral membership within GATT/WTO are only available and included for the HMR and the CEPII data sets. Common island and landlock status indicators are included in the HMR and the industry-level data sets. All these variables are captured by $\mathbf{X}_{ij}$ in

Equations (4.11) and (4.12).

Exclusion Restriction Variables

To overcome the weak identification simply through functional form, HMR propose at least three exclusion restriction variables for their procedure.

HMR prefer a specification where, in the first stage probit, a proxy variable of bilateral fixed export costs is employed. This variable—measuring the bilateral number of procedures needed to start exporting—might not influence the intensive margin but the probability of a positive trade flow. Since this variable does not cover a rich country sample they offer alternative exclusion restrictions. Beside the coverage issues of this variable, we suspect that the fixed exporting costs might significantly change over time. Therefore, using this variable which is, at best, available for periods after year 2000 would not fit our multi-period trade data sets that start in the seventies.

Alternatively HMR use the bilateral measures *common religion* and *common language* and do not find a qualitative differences in their results across any employed exclusion restrictions. The *common religion* variable measures to what extent the population of the importing country and the exporting share a common religion according to data from the Christian Research Association for the year 2003. In particular, the measure is calculated by first summing the number of people that belong to each existing religion in an importing country and then calculating each group's share of that country's total population. This share is then multiplied by the corresponding share of the exporting country. The measure is bounded between 0 and 1, with large numbers indicating a large degree of overlap in the religious structures of the country. The second excluded variable indicates whether the importer and the exporter share a common language. Below we stick to this choice of exclusion restrictions and use the same control variables as in (4.11) (including the importer and exporter fixed effects) in addition to both excluded variables to estimate the probability of exporting in the first stage. We do so for all three data sets.

4.3.3. Results

To explore the distance puzzle, we thus estimate (4.11) for all three data sets by year and additionally by industry for the industry-level data set. With ten years from the original HMR data set, 27 years from the CEPII data set and data for 28 industries over 26 years from the industry-level data set and with four specifications respectively, this amounts to estimating 765 first-stage regressions and 3060 second-stage regressions. For expositional reasons, we show our results graphically.

HMR versus OLS

Figure 4.2 depicts distance coefficients estimated with OLS and the non-linear method from HMR for the original HMR data set. For each year, the distance coefficient is calculated and is then plotted over the available time period from 1980-1989. To indicate the time pattern for each estimator, we have added a quadratic trend. Several interesting features are present in Figure 4.2.

Note that the trend of the distance coefficient, when estimated by OLS, $\hat{\gamma}^{OLS}$, is slightly *increasing* over time. This confirms the puzzling result in previous studies that the negative impact of distance on trade seems to increase rather then decrease over time, which would be expected from the globalization process. Turning to the HMR distance coefficient, $\hat{\gamma}^{HMR}$, we note that $\hat{\gamma}^{HMR}$ is indeed decreasing over time. Examining the bias of OLS, $\hat{\gamma}^{OLS} - \hat{\gamma}^{HMR}$, we note that this is positive. From Proposition 1, this is consistent with the upward bias from omitting the number of exporters dominating the selection bias from omitting zero trade flows. In addition, the bias grows over time. From theory, this suggests that globalization and reduced trade costs seem to decrease the downward bias from selection more than they reduce the upward bias from the number of exporters, see Proposition 2. Hence, the omitted variable bias seems to dominate the selection bias, and becomes relatively more important than the selection bias over time.

In Figure 4.3, we compare OLS with the linear approximation of HMR. We note that the results are qualitatively the same as in Figure 4.2. The HMR distance coefficient is decreasing over time, whereas the OLS coefficient increases with the associated bias of OLS increasing. Comparing Figures 4.2 and 4.3 we note that the linear approximation of HMR gives very similar results to the non-linear version of HMR. That the linear approximation of the HMR works satisfactorily is useful information for a future application of the linear approximation of the HMR methodology, given the cumbersome estimation of the non-linear version of HMR.

This main empirical finding holds for all three data sets as can be seen from Figures 4.4-4.7. Figures 4.5 show for the CEPII data qualitatively the same results as Figures 4.2 and 4.3 do for the original HMR data set. Again, we find this for the non-linear method of HMR and the linear approximation we propose. When we estimate (4.11) by year and industry and then average the estimated distance by year, we find a very similar pattern shown in Figures 4.6-4.7.[19]

Figure 4.2.: Comparing Estimates of HMR with OLS for Original HMR Data.

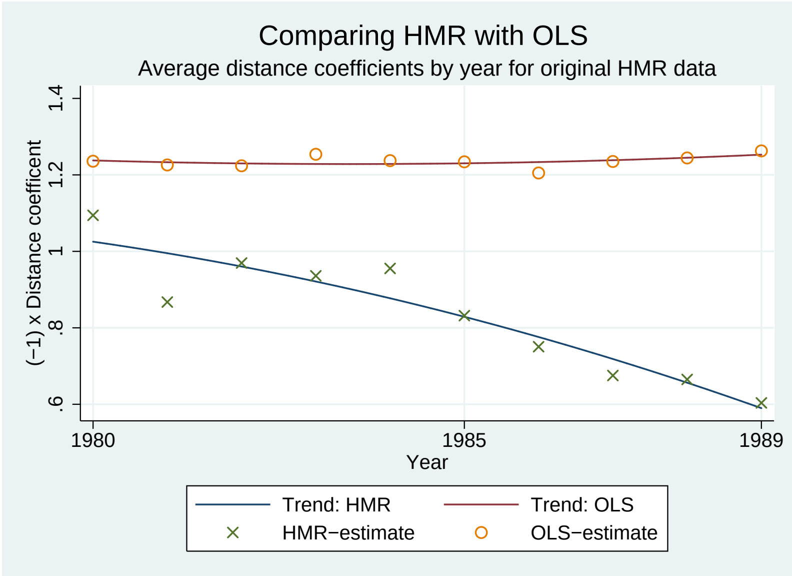
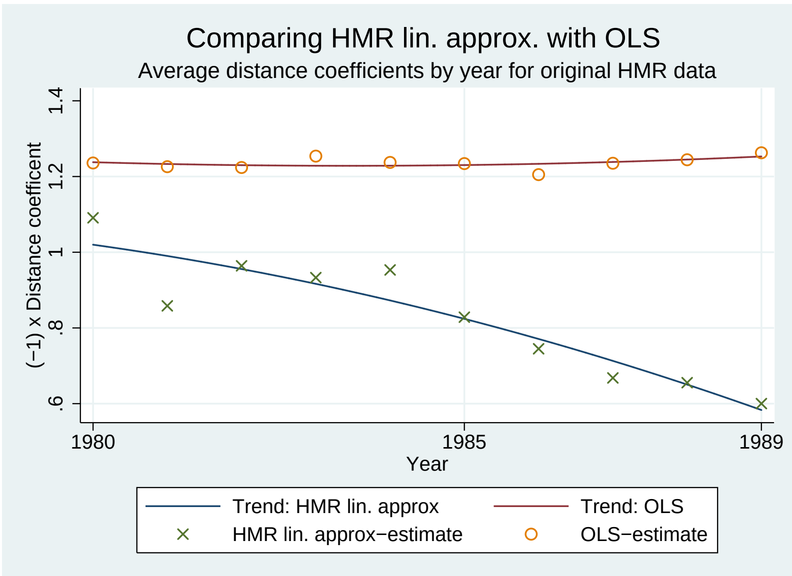


Figure 4.3.: Comparing Estimates of Linear Approximation of HMR with OLS for Original HMR Data.



[19]Note here that, although the linear approximation works best for values of δ around 1 (see Footnote 2), it still performs well for different values of correction factors.

Figure 4.4.: Comparing Estimates of HMR with OLS for CEPII Data.

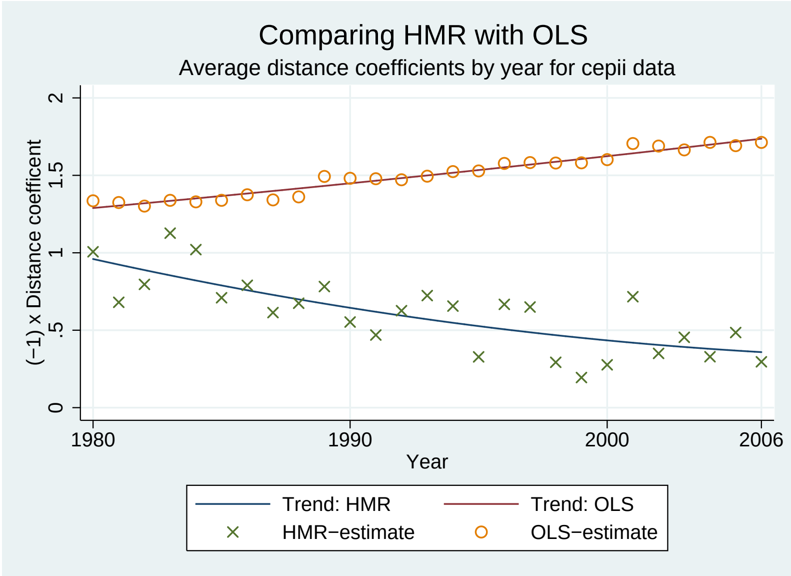
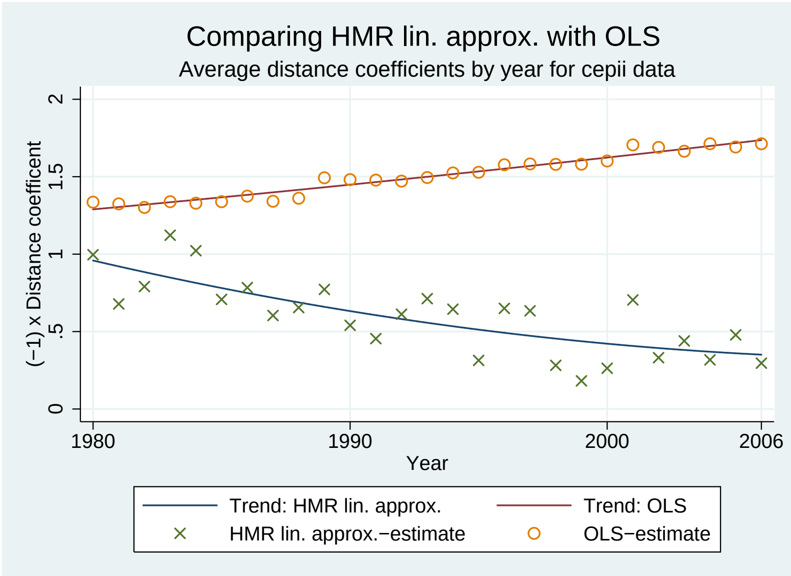


Figure 4.5.: Comparing Estimates of Linear Approximation of HMR with OLS for CEPII data.



Heckman versus OLS

Next, we make a comparison by results obtained with the usual Heckman procedure. Since Heckman does not correct for the omitted variable bias, but the sample selection, we expect it's estimated distance coefficients to be larger in absolute values than those from OLS. This is exactly what our results in Figure 4.8 for the original HMR data depict. The estimated distance coefficients are bigger than those estimated from OLS

Figure 4.6.: Comparing Estimates of HMR with OLS for Industry-Level Data (Averaged).

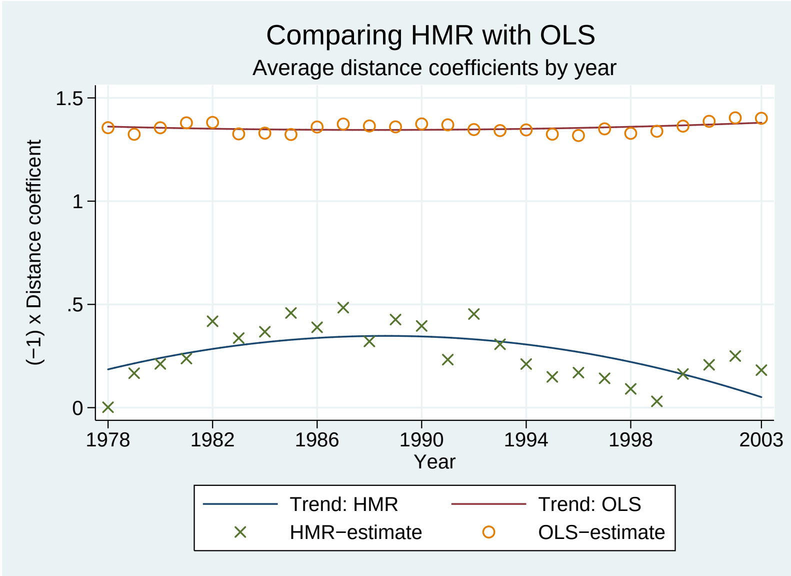
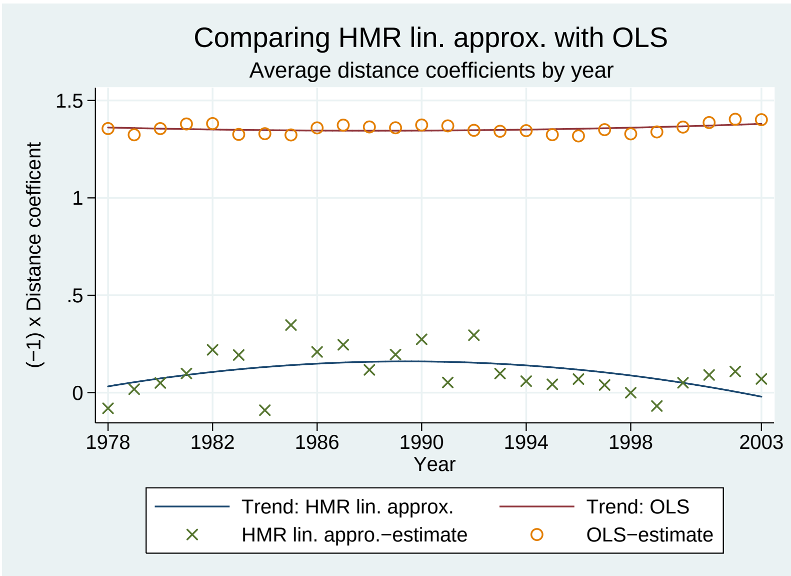


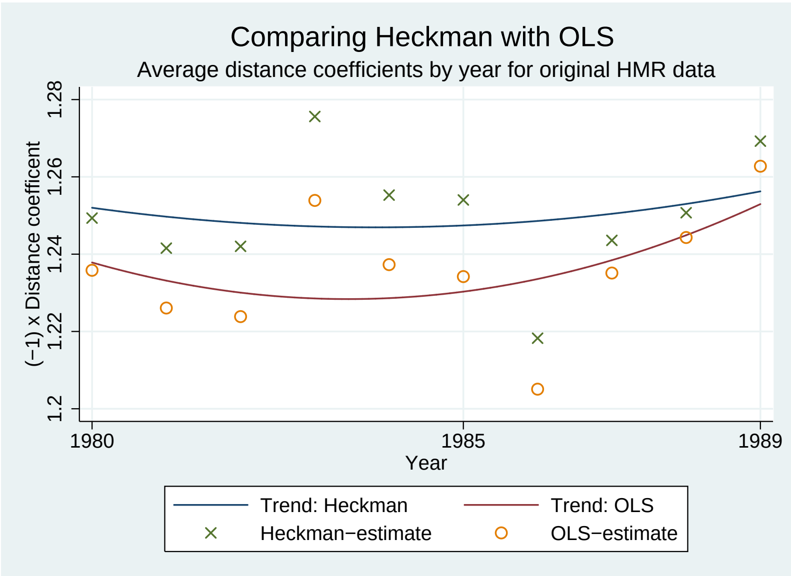
Figure 4.7.: Comparing Estimates of Linear Approximation of HMR with OLS for Industry-Level Data (Averaged).



in every single year in our data. This empirical finding is very much in line with our theoretical result that accounting for zero trade flows cannot solve the distance puzzle when HMR is the data generating process. The results for the CEPII data (Figure 4.9) and the averaged distance coefficients from the industry-level estimates (Figure 4.10) again support this theoretical result. We find no evidence for a reduction of estimated distance coefficients when accounting for sample selection from ignoring zero trade flows compared to OLS estimates. Figure 4.10 also shows bigger distance coefficients in every single year and an increasing trend for the Heckman estimates. The importance of zero trade flows seems to be less for the CEPII data set given that the Heckman estimates are very similar to the OLS results. This is reasonable since Head et al. (2010) fill up many zero trade flows which actually have not been zero while generating the CEPII data set (see appendix of Head et al. (2010)).

To sum up our results up until here, we do not find a qualitative difference between the three data sets. Some quantitative differences are quite reasonable since for example the results for the industry data are averaged over industries with equal weights.

Figure 4.8.: Comparing Estimates of Heckman with OLS for Original HMR Data.



Industries

Figure 4.11 show changes over time in the level of the distance coefficient for each of the 28 industries from HMR and OLS. Most industries show a similar pattern, where the distance coefficient with OLS is increasing over time and the HMR distance coefficient is decreasing over time, producing an increasing bias of the OLS estimates.[20] In particular, these patterns are present in industries that are characterized by intra-industry trade (e.g. “Footwear“ or “Manufacture of machinery”), whereas the patterns seem weaker in

[20] Actually, the bias can be identified visually from Figure 4.11. Therefore we added again quadratic fits over time to our estimates. We mostly observe an increase in the difference between the quadratic fit of the OLS estimates and the quadratic fit of the HMR estimates over time, at least for the second half of our data period. Note that this difference is always significant and never converges to the end of our data period, except for “petrolium refineries”.

Figure 4.9.: Comparing Estimates of Heckman with OLS for CEPII Data.

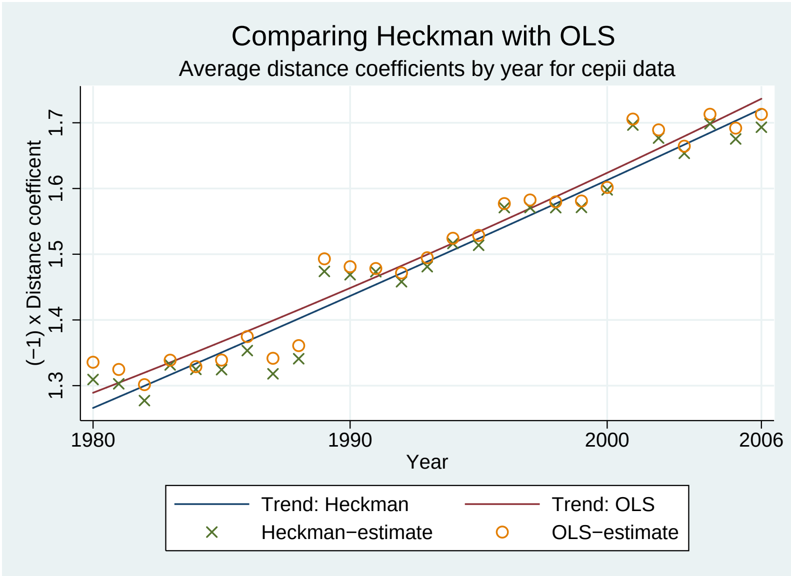
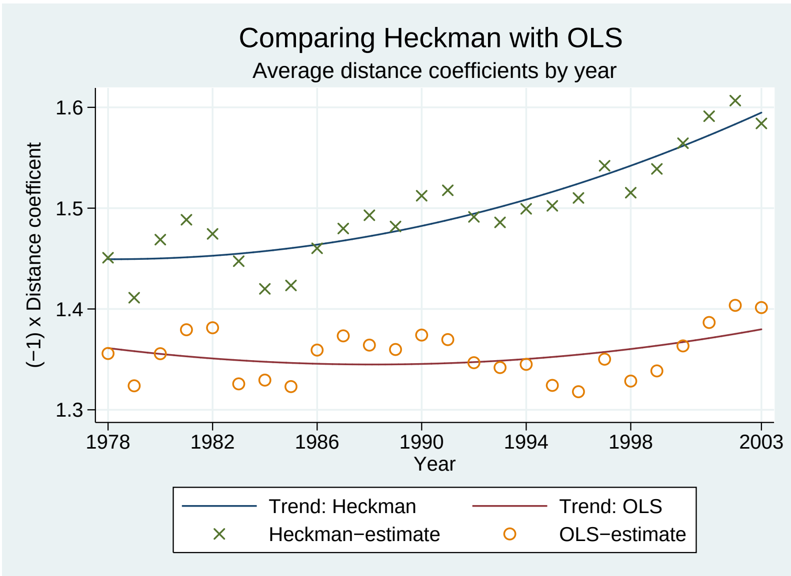


Figure 4.10.: Comparing Estimates of Heckman with OLS for Industry-Level Data (Averaged).



industries where the pattern of trade is to a larger extent explained by comparative advantage (e.g. “Tobacco manufactures” or “Petrolium refineries”). This is also what should be expected since trade in the HMR model generates intra-industry trade.

Descriptive evidence of these results is shown in Table 4.1 where the ISIC classification is linked to the industry classification with respect to product differentiation according to Rauch (1999) and the information of whether OLS bias increases or not. Rauch classifies industries at the SITC 4-digit level as differentiated or not. However, we first subsume these SITC 4-digit classifications into our ISIC classification which actually aggregates the SITC 4-digit industries at a higher level, i.e. the ISIC codes consist of more than one SITC 4-digit code. We then calculate the share of differentiated SITC 4-digit industries according to Rauch (1999) within our 28 ISIC industries (*Share of differentiated industries*).

In Table 4.1 we do find a correlation of 0,34 between the dummy *Increase in bias* and *Share of differentiated industries*.[21] The mean *Share of differentiated industries* within the 23 industries where we do find an increasing bias is 0.75 which is much higher than the 0.40 within the 5 industries where we do not find an increase in the bias. If we draw an arbitrary cutoff for differentiated versus homogeneous industries at a *Share of differentiated industries* of 0.5 we would see that 17 out of 19 cases are differentiated according to the Rauch classification. Since the sizes of the SITC 4-digit industries are not accounted for when subsuming them into the ISIC classification, we now concentrate on ISIC codes where we calculated a clear-cut *Share of differentiated industries* of either 0 or 1. Within these 15 observations we find 12 matches, either between no increase in the bias and a clear-cut *Share of differentiated industries* of 0 or between an increase in the bias and a clear-cut *Share of differentiated industries* of 1.

We also link our industry-level results in Table 4.1 to the estimated industry specific elasticities of substitution from Broda and Weinstein (2006). We then take the average of the estimated elasticity of substitution from Broda and Weinstein (2006) over all SITC rev. 2 3-digit industries which sum up to the ISIC-level. Generally, we expect that industries with high elasticities of substitution are less differentiated (more homogeneous) and that an increase in the bias for these industries is less likely with respect to the HMR model. Dropping the suspicious average elasticity of paper products which is far more than 2 standard deviations different from the average of the industry elasticities, we find the following result. There is a small negative correlation (-0.08) between the dummy which indicates an increase in the bias from our results and the elasticity calculated from the results of Broda and Weinstein (2006). In 13 out of 22 cases where industry elasticity is smaller than the average industry elasticity, we also find an increase in the bias of OLS. Examining the clear-cut cases, where we either observe 100 percent differentiated goods or 0 percent differentiated goods according to Rauch (1999), we find 10 out of 15 of these

[21] However, left with 28 industries/observations, the regression results lack in their precision, but can serve as additional descriptives. Point estimates of regressions (probit, logit or linear probability) of the dummy which indicates a bias increase on *Share of differentiated industries* give results in our favor (positive) and are significant at the 10% level.

observations to match with our expectation; the three zero percent differentiated goods industries all have an above average elasticity and 7 out of 10 100 percent differentiated goods industries have a below average elasticity. All in all the picture for the industry elasticities is consistent with our expectation. A more comprehensive look at the industry dimension of the HMR model is not the focus of this paper. However, it is also not necessary to support our result that the distance puzzle cannot be solved only by accounting for zeros.

Globalization and Transport Costs

Additionally, we provide evidence that the HMR data generating process fits the data well and that Equation (4.11) might consistently estimate the distance coefficient. Figures 4.12 to 4.17 show the results of relating the estimated distance coefficient $\hat{\gamma}^{HMR}$ to actual trade costs. Firstly, Figures 4.14 and 4.16 show that the estimated distance coefficients are strongly positively correlated with shipping costs in data recently published by Hummels (2007). Figure 4.12 does not support this finding, which we suspect to happen because of the low number of observations here. Secondly, Figures 4.13, 4.15 and 4.17 shows that the $\hat{\gamma}^{HMR}$ is also positively correlated with oil prices, which should be an important determinant of transport costs. Additionally, we note that the OLS estimate of the distance coefficient is negatively correlated with these data on transport costs. Once more, this non-intuitive correlation can be explained because OLS neither controls for the omitted variable of the number of exporters nor for the omission of zero trade flows.

Auxiliary Estimates

Finally, we underpin our results by plotting the bias terms in more detail and provide more evidence on the mechanisms in play and how they evolve over time. Actually, our estimates show that the bias of OLS increases over time. Hence, it must either be the case that the downward bias from sample selection decreases faster than the upward bias from not controlling for the number and size of exporting firms or that the downward bias stays stable over time and the upward bias increases or that the downward bias decreases and the upward bias stays stable, or anything in between these cases. Figures 4.18 to 4.20 depict the actual bias terms for the three data sets respectively. In these figures we plot the combined bias term of the OLS estimator (Equation (4.8)) and the two separated bias terms from sample selection ($\Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}^{*}_{ij}$) and omitting the measure of heterogeneity of the HMR estimator ($\gamma\delta$), respectively. All bias terms are averaged over all bilateral

Table 4.1.: Indusry Composition and the Bias of OLS

ISIC	Industry	Increase in bias	Share of differentiated industries	Elasticity
311	Food	yes	0.28	3.94
313	Beverage	yes	0.2	4.18
314	Tobacco	no	0	9.77
321	Textiles	yes	0.68	5.93
322	Wearing apparel	yes	0.96	5.83
323	Leather	yes	0.88	1.94
324	Footwear	yes	1	1.74
331	Wood and cork	no	0.7	2.13
332	Furniture	no	1	1.64
341	Paper	yes	0.18	45.81
342	Printing	yes	1	5.6
351	Industrial chemicals	no	0.19	4.33
352	Chemical products	yes	0.92	1.92
353	Petroleum refineries	no	0.13	9.63
354	Products of petroleum and coal	yes	0	15.64
355	Rubber products	yes	1	2.47
356	Plastic products	yes	1	2.7
361	Pottery, china and earthenware	yes	1	1.92
362	Glass	yes	1	1.92
369	Non-metallic mineral	yes	0.8	1.89
371	Iron and steel	yes	0.33	4.16
372	Non-ferrous metal	yes	0	5.09
381	Fabricated metal	yes	1	7.06
382	Machinery	yes	1	8.97
383	Electrical machinery	yes	1	6.63
384	Transport equipment	yes	1	10.31
385	Scientific equipment	yes	1	2.05
390	Other manufacturing	yes	0.92	2.69

Notes: 28 ISIC Rev. 2 manufacturing industries, where *yes* corresponds to a dummy which is equal to 1 if we do find an increase in the bias $\hat{\gamma}^{OLS} - \hat{\gamma}^{HMR}$ from Figure 4.11. *Share of differentiated industries* is the share of differentiated SITC 4-digit industries according to Rauch (1999) within the ISIC industry. *Elasticity* corresponds to the average elasticity of substitution from Broda and Weinstein (2006) over all SITC rev. 2 3-digit industries which sum up to the ISIC-level.

observations by year.

For all data sets the bias from zero trade flows is almost stable over time and only decreases slightly. $\gamma\delta$ actually increases over time. Solely important for our results is that the relative change over time fits in with Proposition 3. The bias from omitting heterogeneity dominates the bias from sample selection due to zero trade flows. These figures also highlight that the bias from omitting heterogeneity ($\gamma\delta$) drives the changes over time since the sample selection bias ($\Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}^*_{ij}$) changes only slightly. This is in line with the fact that zero trade flows alone cannot explain the distance puzzle.

Contrary to Proposition 2, figures 4.18 and 4.19 depict an increase in the heterogeneity. This difference between our theory and the estimates can be explained by the ceteris paribus assumption of the theory which is not met for the estimates. Specifically, recognize that we estimate the coefficients for each year so that the heterogeneity of firms can increase over time due to reasons other than trade. To demonstrate this point, we allow both the distance coefficient, γ, and the impact of firm heterogeneity on trade, δ, to change over time. Specifically, we again take the derivative with respect to time of Equation (4.8) but now take into account changes in γ and δ simultaneously:

$$\frac{\partial \text{Bias}(\hat{\gamma}^{OLS})}{\partial t} = \underbrace{\delta \underset{(-)}{\frac{\partial \gamma}{\partial t}} + \gamma \underset{(+)}{\frac{\partial \delta}{\partial t}}}_{\text{Omitted variable bias}} - \underbrace{\Xi \left(\left(\delta + \beta_{u\eta}\right) \underset{(-)}{\frac{\partial \bar{\eta}^*_{ij}}{\partial t}} + \bar{\eta}^*_{ij} \right)}_{\text{Selection bias}} \gtreqless 0. \tag{4.13}$$

We again can decompose the bias of OLS into the omitted variable bias and the selection bias. Sticking to the assumption that increased globalization implies a decreasing γ over time and assuming the selection bias to be small (as figures 4.18 to 4.20 suggest) and $\frac{\partial \delta}{\partial t}$ to be positive and bigger in absolute value than $\frac{\partial \gamma}{\partial t}$, we end up with an increase of the bias of the OLS estimates of the distance coefficient which is in line with the figures 4.18 to 4.20. This implies that an increase of the heterogeneity in the data due to other reasons can actually explain why the omitted variable bias can increase over time simultaneously as a decreasing elasticity of distance, while the selection bias only changes marginally.

4.4. Conclusions

Globalization has advanced rapidly during the last two decades. In contrast, the influence of distance in empirical estimates of bilateral trade flows has remained high and has not declined. In this paper, we use the model by Helpman et al. (2008), emphasizing zero trade flows and firm heterogeneity, to resolve this "distance puzzle".

Using different trade data sets, the non-linear estimation of HMR leads to declining distance coefficients over time. These coefficients also reflect the variation in "true trade costs" as the estimated HMR distance coefficients are also strongly correlated with the variation in freight costs and oil prices. When estimating the effect of distance on trade with OLS, we do not only find a larger distance coefficient but also that it increases over time. Thus, the distance puzzle arises from a growing bias of OLS estimates.

We show how the growing bias of OLS estimates can be explained from the two sources of bias generated from applying OLS to a gravity estimation when the HMR model is the data generating process. The upward bias of the OLS estimates implies that the omitted variable bias (from the number of heterogeneous exporting firms) must dominate the sample selection bias (due to the omission of zero trade flows). When relating globalization to a fall of the true distance coefficient, both the downward bias from sample selection from omitting zero trade flows and the upward bias from omitting the number and size of exporting firms will decrease with increasing globalization (in absolute value). We find that the bias of OLS increases over time. Decomposing the bias of OLS into its two componenets, the omitted variable bias and the sample selection bias, we empirically find an increase over time of the omitted variable bias while the sample selection bias hardly changes over time. This result implies an increasing importance of firm heterogeneity over time.

On a final note, the gravity equation is perhaps the most widely used tool in empirical work using aggregate international trade data. While firm-level data is becoming more frequent, applying gravity equations on aggregate trade data will also remain common in the future when various policy issues are investigated. In this paper, we have shown how taking sample selection and exporter firm heterogeneity into account is crucial for understanding the effect of distance on international trade when aggregate trade data is used. Then, we showed the usefulness of a linear approximation of the HMR estimator. As this estimator is much simpler to apply than the non-linear estimator of HMR, we suggest that the linear approximation could be fruitfully used for many other research questions.

Figure 4.11.: Comparing Estimates of HMR with OLS for Different ISIC Rev. 2 Industries

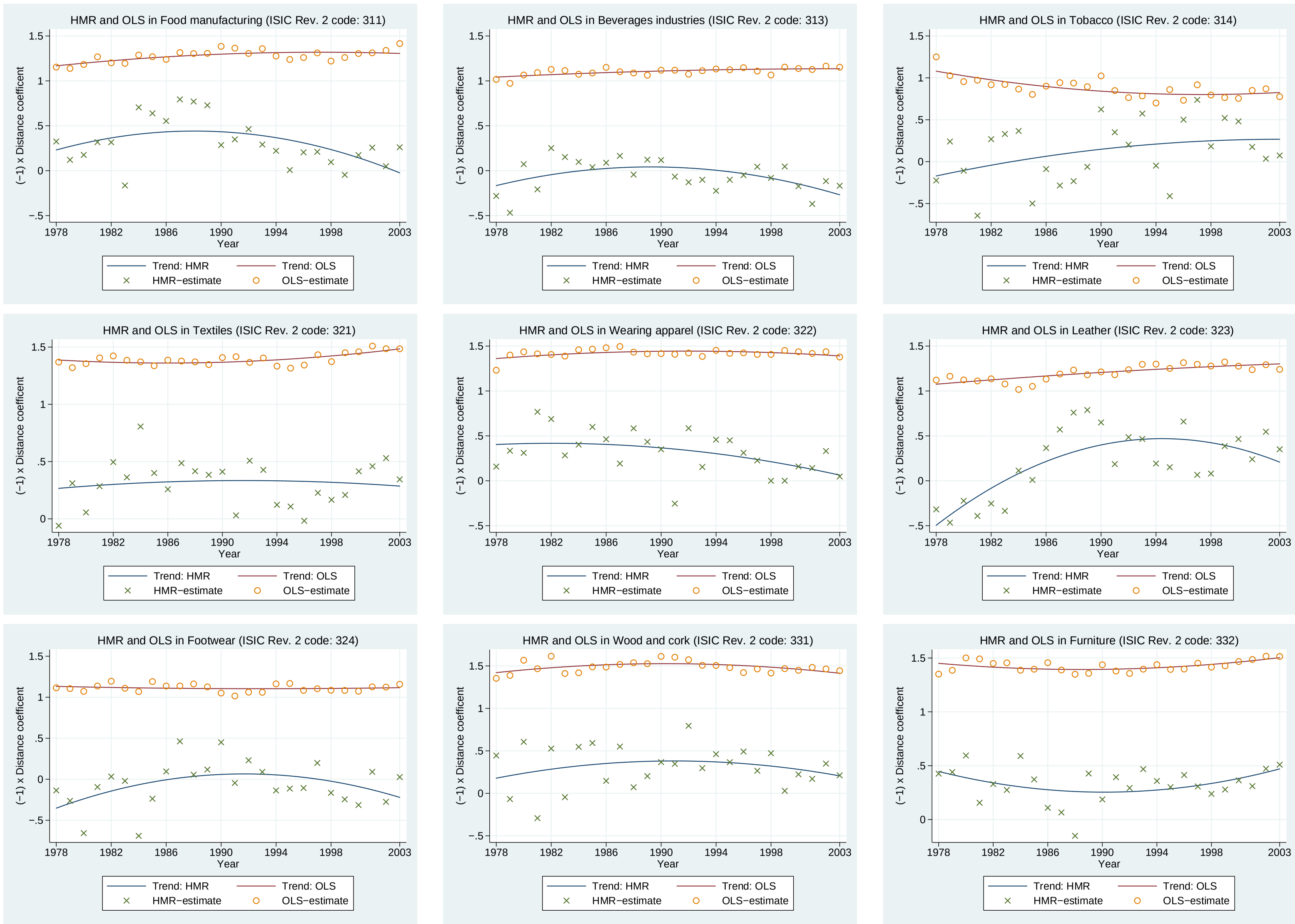


Figure 4.11.: Comparing Estimates of HMR with OLS for Different ISIC Rev. 2 Industries (Continued)

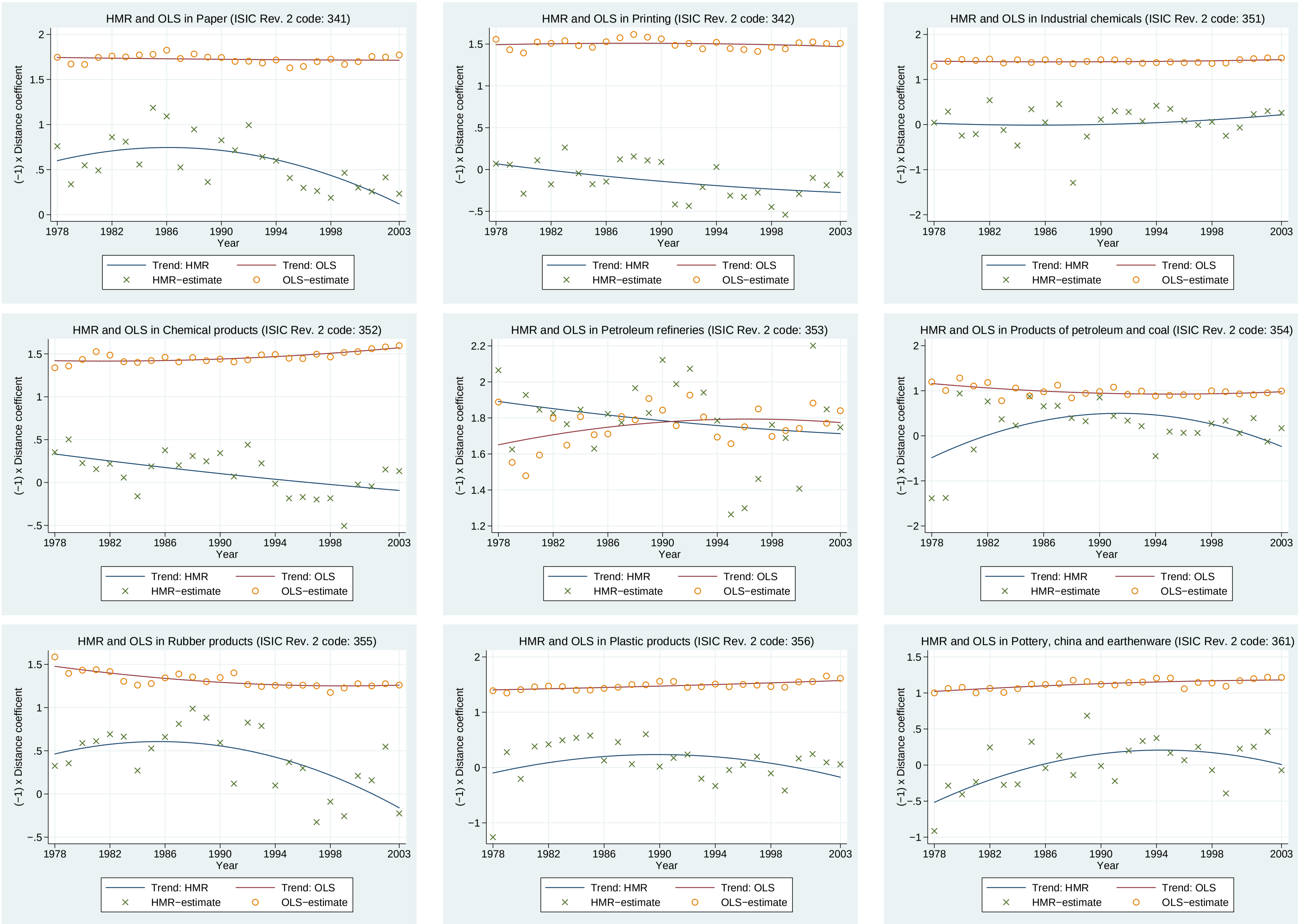


Figure 4.11.: Comparing Estimates of HMR with OLS for Different ISIC Rev. 2 Industries (Continued)

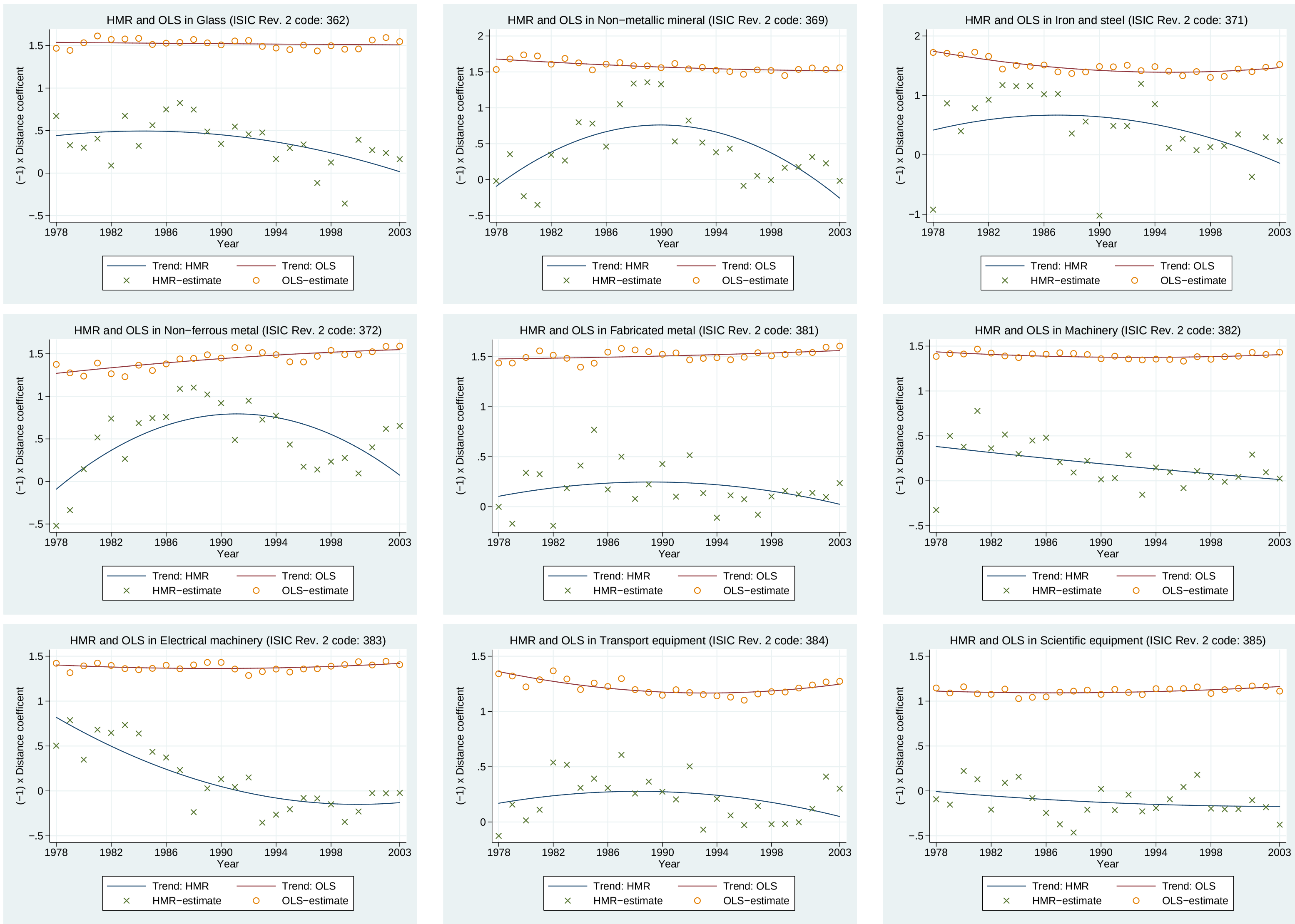


Figure 4.11.: Comparing Estimates of HMR with OLS for Different ISIC Rev. 2 Industries (Continued)

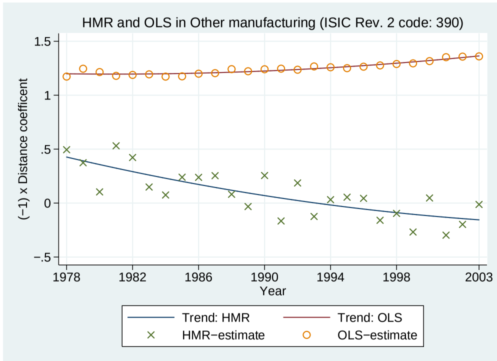


Figure 4.12.: HMR, OLS and Freight Costs for Original HMR Data.

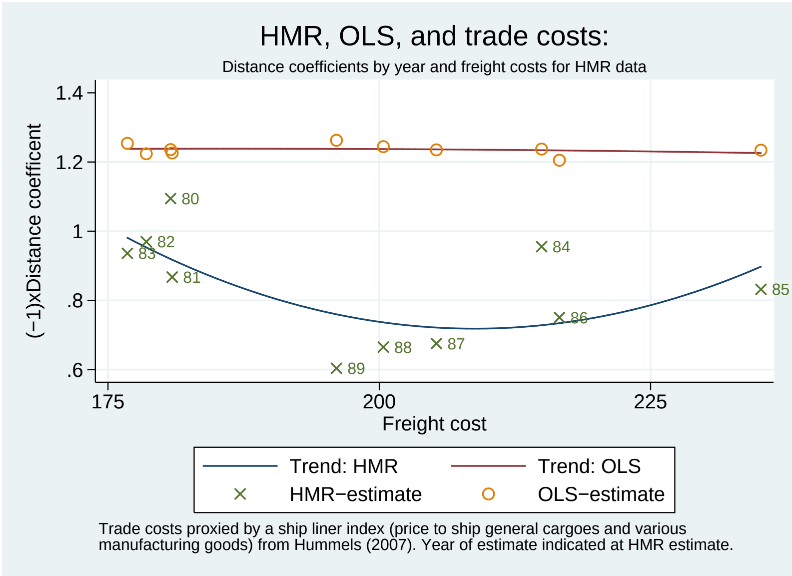


Figure 4.13.: HMR, OLS and Oil Prices for Original HMR Data.

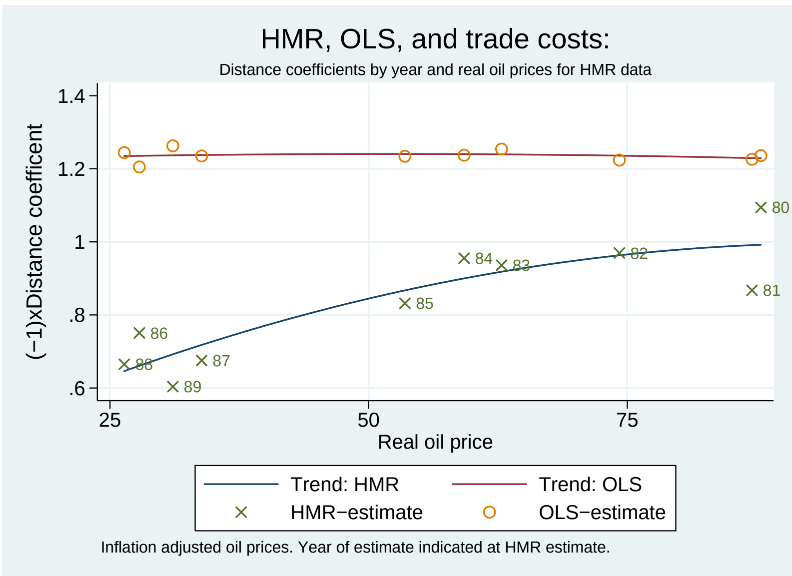


Figure 4.14.: HMR, OLS and Freight Costs for CEPII Data.

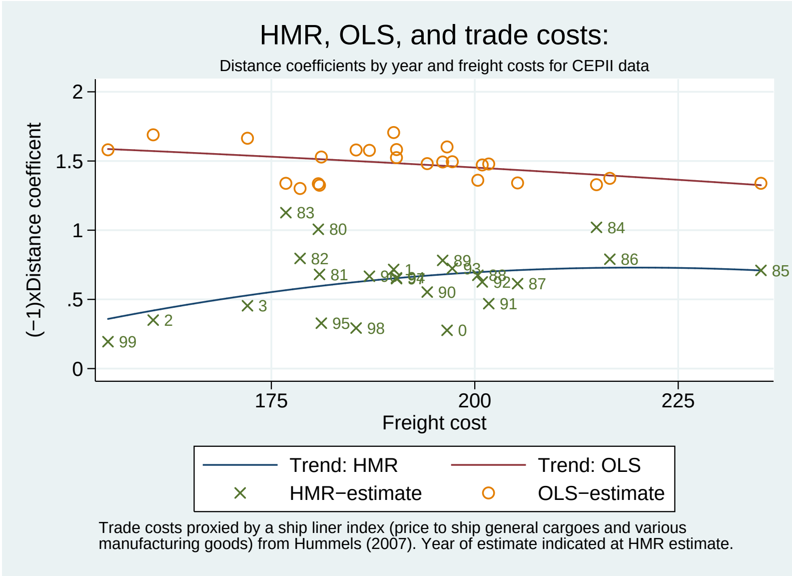


Figure 4.15.: HMR, OLS and Oil Prices for CEPII Data.

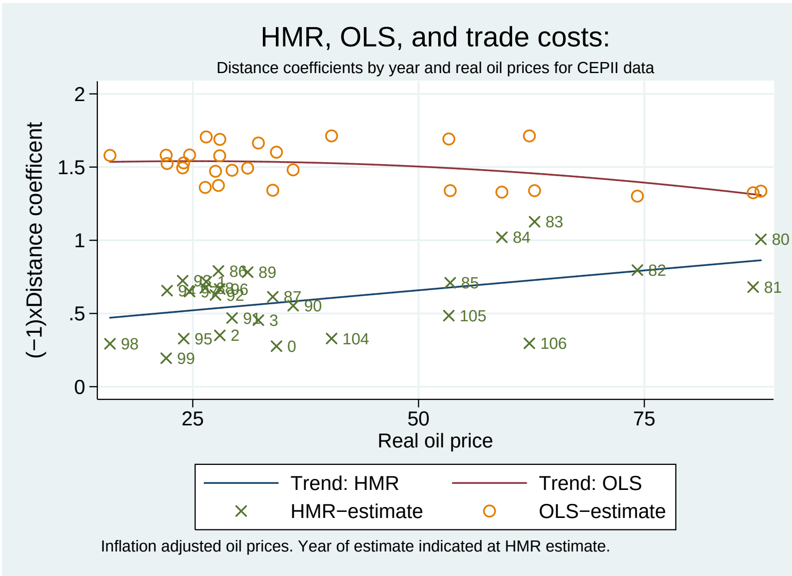


Figure 4.16.: HMR, OLS and Freight Costs for Industry-Level Data (Averaged).

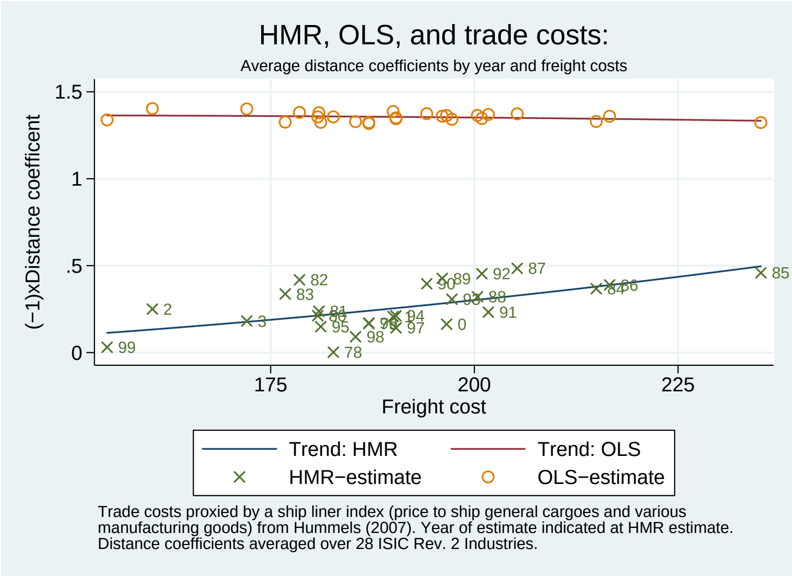


Figure 4.17.: HMR, OLS and Oil Prices for Industry-Level Data (Averaged).

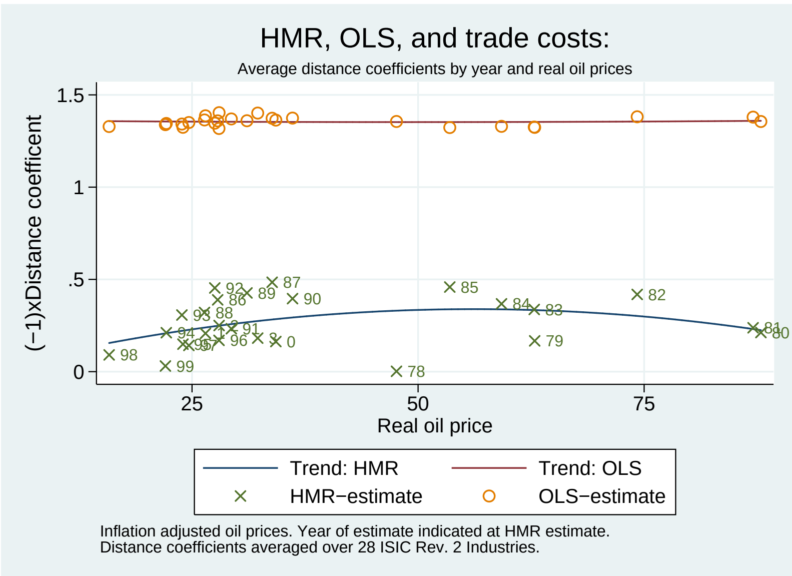


Figure 4.18.: Bias Terms Over Time HMR Data.

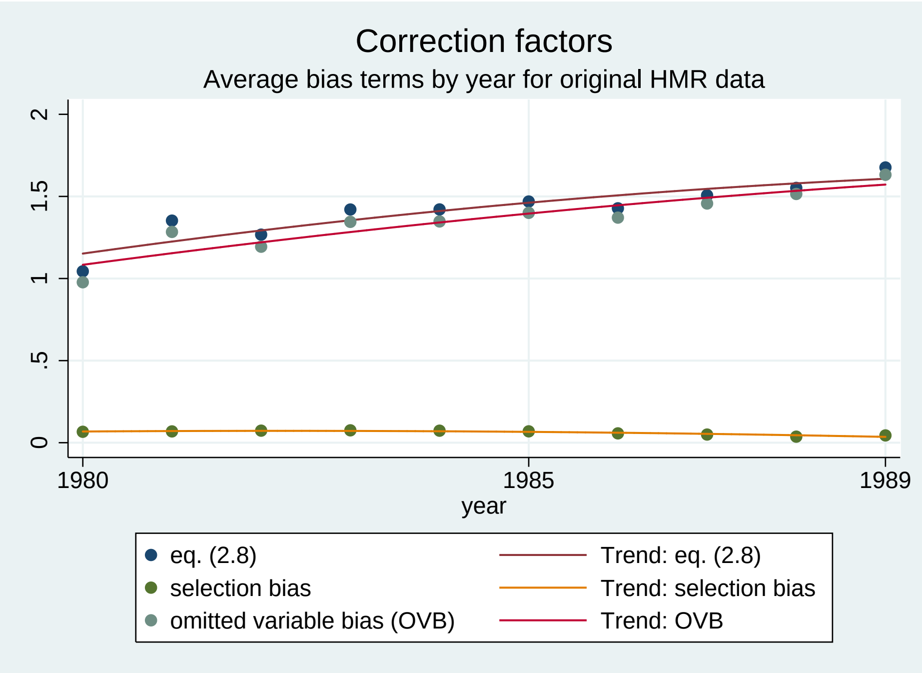


Figure 4.19.: Bias Terms Over Time CEPII Data.

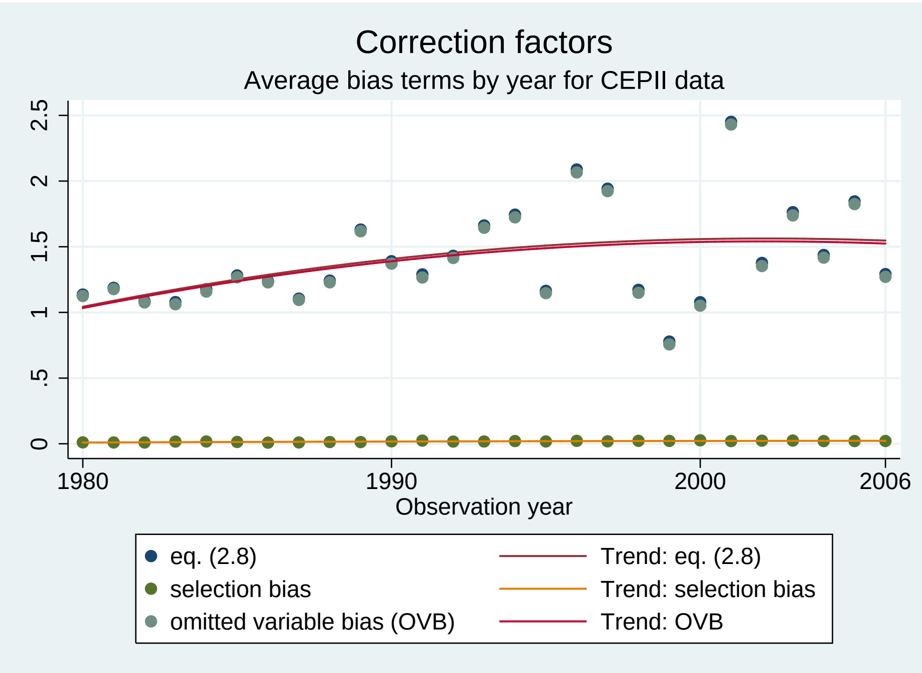
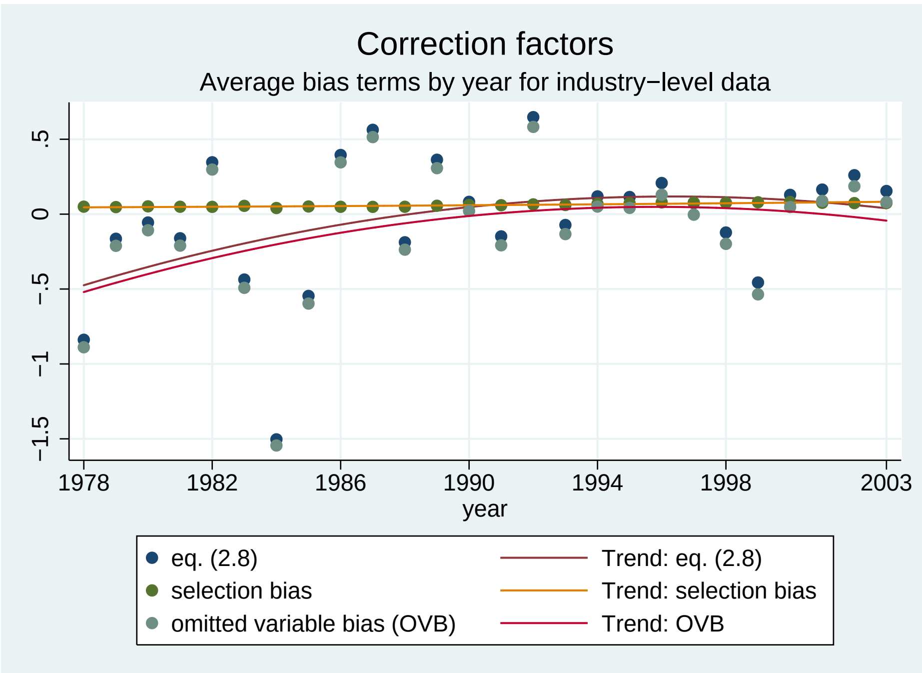


Figure 4.20.: Bias Terms Over Time Industry Data.



5. Employment Effects of Offshoring and FDI – Disentangling Modes and Measures[1]

5.1. Introduction

The recent two decades of globalization have been vividly debated in the media and in politics (Mankiw and Swagel, 2006), fearing the exodus of production jobs in highly developed countries. But was the economic impact really as large as the intensity of the debate on it? At the very heart of the recent wave of globalization was the unbundling of the production process, a shift of production steps to locations with lower costs (Baldwin, 2006).

Contrary to the strong media perception, academic research on employment effects of offshoring or correlated measures of FDI are ambiguous.[2] A significant number of studies find positive employment effects contrary to the perception of the public and partly contrary to economic theory. A potential explanation is that offshoring has not been systematically measured by statistical offices and researchers have to refer to proxy variables, which may erroneously include many events other than offshoring.

Using two different micro-data sets, this study investigates whether ambiguous employment effects of international sourcing arise from using different proxy variables of offshoring and FDI, or whether different estimation methods, different samples, or different control/selection variables are responsible for the ambiguity.

Considering only studies which use micro-data and firm-level measures of offshoring or FDI, studies still differ by their choice of the employed measures of FDI and offshoring: new investments abroad (Barba Navaretti and Castellani, 2004; Barba Navaretti et al.,

[1] This chapter bases on joint work with Dieter Urban. All remaining errors in this volume are mine.
[2] See Crinó (2009) for a survey.

2010), expansion of employment in foreign affiliates (Becker and Muendler, 2008), increase in intermediate input purchases from abroad (Biscourp and Kramarz, 2007; Moser et al., 2015), increase in usage of intermediate inputs interacted with contemporaneous domestic establishment restructuring (Moser et al., 2015) or relocation (Wagner, 2011).

These studies differ also by countries on which data were drawn: Italy (Barba Navaretti et al., 2010), France (Biscourp and Kramarz, 2007), and Germany (all other above mentioned studies). They differ further by the estimation technique: OLS, dynamic panel data, or, in most studies, matching estimators, where control or selection variables again differ across studies.

As large as the range of choices in study design are, as large is the range of results with strong positive employment effects from foreign employment expansion (Becker and Muendler, 2008) on one end, and (slightly) negative employment effects in some sample subgroup (Biscourp and Kramarz, 2007), or when not excluding outliers (Wagner, 2011), or when interacting offshoring treatment with contemporaneous establishment closure events (Moser et al., 2015) on the other end.[3]

We investigate a unique and discrete offshoring measure of German establishments that experienced offshoring during the time period 2004-2006 and compare several internationalization measures. On the one hand, we apply FDI, market seeking FDI and cost saving FDI measures using different control variables and different estimation methods. On the other hand, we apply a measure of offshoring, which is similar to a measure used by Wagner (2011), albeit the data period is different and the data differ by their coverage and their quality (response rates and missing values). To compare methods, we apply both OLS estimators and matching techniques. To keep results comparable, we use three different sets of selection variables to determine the probability of FDI or relocation abroad of an establishment for the matching methods. Note that for the latter, Smith and Todd (2005a) find that an analysis of the prominent LaLonde data (LaLonde, 1986) is quite sensitive to different samples and selection variables.

Our analysis is closest to Moser et al. (2015) and Wagner (2011) with respect to the study design and empirical results. When focusing on different channels through which international sourcing might affect domestic employment, Moser et al. (2015) also find negative employment effects of one offshoring measure. Overall they do find positive employment effects of internationalization on establishment employment including offshoring measured via an increase in intermediate input purchases from abroad. They do find a negative employment effect, whenever they restrict their sample to establishments

[3]See Table D.1 for an even broader picture of measures, methods, data, and results of other studies.

which undergo a general domestic restructuring of a part of the establishment in the same period. This result for a refined measure of offshoring already indicates that potential negative employment effects occur if domestic layoffs are temporarily linked; Moser et al. (2015) conclude that the downsizing channel dominates potential productivity effects of offshoring in these cases. Wagner (2011) coins the term relocation for a firm-level measure of internationalization where a domestic part of the firm is replaced by a foreign one. Note that for this measure a domestic restructuring is causally, not just temporarily, aligned compared to the measure for which Moser et al. (2015) find negative effects. Wagner (2011) finds comparably small negative employment effects and concludes that the economic impact is not as large as feared.

Even within a unified data framework, we find significant positive employment effects from different FDI measures and strongly negative and significant employment effects from relocation abroad. The latter result is robust evidence for significant employment losses from relocation abroad as one mode of foreign activity measured at the micro-level. Moreover, the disparity of results on the two types of measures of offshoring or FDI does neither hinge on differences in estimation methods (OLS vs. matching), nor on the choices of selection or control variables. We explain this disparity of results by the variety of activities that are captured by these different measures. None of the measures captures only one single type of FDI. For example, the FDI measure may comprise horizontal FDI, vertical FDI, export platform FDI, etc.; relocation abroad may also consist of horizontal or vertical FDI. Some of these activities may occur in the vein of a general expansion of a firm both abroad, but also at home. This may explain why most FDI measures, even those of cost-saving FDI may go hand in hand with domestic employment expansion. Only in some cases, an expansion abroad substitutes for domestic production, and sheds off domestic labor. Most FDI activities abroad either stimulate domestic activities, or are concomitant to a general expansion of a multinational firm. This result is also in line with the study of Moser et al. (2015) which also finds positive employment effects from offshoring on the one hand, and negative ones from offshoring if accompanied by partial establishment closures on the other hand. To sum up, even when we have a close look at different micro-level measures of foreign activity within one data set, we find tremendously different employment effects.

Since strongly negative employment effects from relocation abroad are, to best of our knowledge, documented the first time in this study,[4] we confirm this new result by a quasi natural experiment which is unique to our data. As a second robustness check we

[4] Only Wagner (2011) provides for some sample evidence of a negative employment effect from relocation. These effects are much smaller and seem to be more sensitive to an outlier correction.

employ a second administrative data set which offers a very similar micro-level measure of relocation. We can confirm negative employment effects of a foreign activity when a restructuring at home is causally coherent.

The paper is organized as follows. The next section gives a framework for a establishment-level analysis of offshoring, including a comparison of different empirical measures linked to theoretical concepts. Section 5.3 outlines the empirical method and section 5.4 discusses briefly the data we use. In section 5.5 we provide the estimation of the propensity score of offshoring and FDI measures and various auxiliary tests. Section 5.6 presents the results of the main estimations of the average treatment effects of offshoring or FDI on employment for both micro-data sets and additional results of the quasi natural experiment. The last section concludes.

5.2. Employment Effects of Offshoring and FDI

To explore why employment effects differ across various studies on FDI/offshoring, we need to understand first how these studies differ in data, measurement, and methodology. We focus in this section on a comparison of measures of FDI/offshoring and ask what types of FDI or outsourcing are captured by each of those measures and which employment effects are expected from each type of FDI.

For example, Becker and Muendler (2008) use expansion of employment in foreign affiliates, which may capture both an incremental increase in horizontal and vertical FDI. If foreign markets grow fast and FDI is of the horizontal type, then foreign affiliates increase and employment at home will not be affected if horizontal FDI is literally replicating the domestic production process abroad. If it is instead of the horizontal type according to Venables (1999), the first stage of the production process may take place at home, and a second one abroad, while the product is always sold abroad in equilibrium. An expansion abroad will then go along with a positive employment effect at home. If the investment is of the vertical type according to Venables (1999), the foreign affiliate produces intermediate inputs for assembly and sales at home. Expansion abroad will occur, because there is increased demand at home, increasing the demand for intermediate inputs from the foreign affiliates. Again, a positive employment effect at home is expected. A negative employment effect may arise, instead, if some production steps, undertaken previously at home, are shifted abroad. But even then, a firm's relocation of domestic production steps abroad may help to save costs, increase its competitiveness, and subsequently augment its world market share, which in turn may stimulate the activities related to production steps that remain at home.

A similar argumentation may result if the measure is a dummy variable for a domestic establishment having a new foreign affiliate (for example Barba Navaretti and Castellani (2004), Buch and Lipponer (2010)). Again this may be horizontal or vertical in nature, yielding ambiguous effects on employment in dependence on which of the above mentioned cases is taking place. On top of the previous cases, some of the new investments may even be mergers & acquisitions which may be completely detached from the domestic production process and domestic employment effects are absent.[5]

FDI measures may be further specified by the motivation for the investment. Mattes (2010) for example distinguishes FDI that is undertaken for the purpose to seek new markets and FDI that is seeking to reduce costs according to self-assessments of firms. While the market seeking motive is rather associated with horizontal FDI, cost reduction is typically associated with vertical FDI. Still, also horizontal FDI can be driven by cost savings (Markusen, 2002). Also Hering et al. (2010) distinguish different motives for FDI according to the location of the sales of the foreign affiliate. Again, both positive or negative employment effects may arise for these specifications for reasons outlined above.

A fourth measure of FDI or offshoring is imported intermediate input demand (Biscourp and Kramarz, 2007; Moser et al., 2015). While this measure is excluding horizontal FDI, but focuses on vertical FDI and international outsourcing instead, a domestic plant may substitute domestic suppliers for foreign suppliers, leaving employment in its own domestic plant possibly unaffected. Alternatively, cost savings through offshoring render the firm more competitive on world markets and stimulate domestic employment. Should the increase in intermediate inputs, instead, go along with a substitution of domestic in-house production, then there may be an employment decline in the domestic plant.

A fifth measure is relocation of domestic production to a plant abroad, for example in (Wagner, 2011). This measure may again capture both, FDI (either horizontal or vertical) and international outsourcing. However, it excludes foreign expansions of operations, which are detached from domestic operations and excludes substitution of domestic for foreign suppliers, too. Still, the closure of a part of a plant may go along with a change in the specialization pattern, giving up some tasks, but expanding others instead. For example, certain low-skilled production activities may be shifted outside of the home country (and probably causing domestic dismissals), while high-skilled intensive headquarter services are extended at home.

[5] See Stiebale and Trax (2011) focusing on mergers & acquisitions.

In summary, all types of FDI or offshoring have ambiguous employment effects in theory, and all measures available in existing data capture several types of FDI. If one wants to pin down employment effects unambiguously, then positive employment effects at home will arise if a firm is expanding both at home and abroad. Instead, a negative employment effect at home is to be expected if domestic production is substituted for production abroad, keeping the overall level of activity constant. As the literature, so far, has used mostly one of these measures at a time, and research designs have been different with respect to the data, the estimation method, and the control or selection variables, the previous results are hard to compare.[6] We investigate the above mentioned measures in a unified estimation design on the same data set to investigate systematically why studies differ in their empirical results so strongly.

5.3. Empirical Method

We closely follow Moser et al. (2015), Wagner (2011), and combine a difference-in-differences estimator with a propensity score matching technique to investigate the relationship between FDI or offshoring and establishment-level employment. The econometric problem is one of the missing counterfactual, i.e. what would have happened if plants had not undergone treatment (i.e. offshoring/FDI). Matching techniques address this problem by statistically designing a counterfactual, while controlling for self selection on observables. To do this in the simplest possible way, a non-treated observation is assigned to each treated one that had ex ante the same probability of obtaining treatment than its treated twin. Treatment is then purely random conditional on the selection variables x, which determine the probability of treatment, $P(D = 1 \mid x)$, where D is a binary variable with value 1 if an observation obtained treatment.

The coefficient of interest is the average treatment effect on the treated (ATT) on establishment-level employment. The ATT measures the average difference between the outcome of the treated observations and the hypothetical outcome without treatment.

To apply matching methods, three core-assumptions of matching must be fulfilled:

1. *Conditional-Mean-Independence assumption* (CMIA):

$$E[y_1 \mid D = 0, x] = E[y_1 \mid D = 1, x] = E[y_1 \mid x],$$

$$E[y_0 \mid D = 0, x] = E[y_0 \mid D = 1, x] = E[y_0 \mid x],$$

[6] See Table D.1.

where y_1 is the employment outcome of an average establishment under treatment and y_0 is the outcome if the same establishment does not experience treatment. This assumption ensures that the assignment to the treatment group is random conditional on observable characteristics, i.e. self-selection into treatment is allowed, conditional on observable characteristics of the establishment. This implies that the mean of observations' outcomes with the same observable characteristics without treatment would be the same.

2. *Overlap Assumption*:

$$0 < P(D = 1 \mid x) < 1.$$

   This assumption ensures that observations with probability zero or one are excluded from the matching process because their assignment is not random by definition.

3. *Stable Unit Treatment Value assumption* (SUTVA):

   SUTVA means there exist no inter-dependencies between the two matching groups. Under this assumption the treatment only affects the treated observation itself. Thus, the effects on the treated have no impacts on the non-treated observations (Rosenbaum and Rubin, 1983).

The combination of the matching estimator with the difference-in-difference approach somewhat relaxes a part of the CMIA. The measurement of the outcome variables in differences eliminates constant time trends based on unobservables like a first difference estimator or a fixed effect model. Indeed, a varying different time trend between treated and non-treated observations might remain and is excluded by assumption.

Observations that are off the overlapping support region are not a problem in our analysis. Non-overlapping observations would pose a problem if many observations would be lost by controlling for this assumption. However, in our specifications we exclude the observations of the treatment group that have a lower propensity score than the lowest of the non-treatment group and non-treated observations that have a higher propensity score as the highest of the treatment observations vice versa. The estimations are constrained to this sample.

Program evaluation methods are typically used to investigate the effects of *small* treatments that have no general equilibrium effects.[7] Consider a job-training program only

[7] An introduction to matching methods is given in Imbens and Wooldridge (2009). For a useful textbook section see Cameron and Trivedi (2005). Angrist and Pischke (2009) follow a new approach to teach these methods and compare them to standard econometrics expediently. A general implementation guide is Caliendo and Kopeinig (2008) and specific problems are discussed for instance in Abadie (2005), Abadie and Imbens (2006), Angrist and Hahn (2004), Dehejia and Wahba (2002), Dehejia

for a small number of unemployed people that does not change the overall skill of all unemployed people and thus does not change the labor demand at all. Since the amount of offshoring units is quite small we follow the literature (for example Wagner (2011)) and exclude general equilibrium or spillover effects which potentially would hurt the SUTVA by assumption.[8]

Consider the following two data generating processes:

$$y_{it}^{T} = g(x_{i0})t + f^{T}(x_{i0})t + \delta_{it}^{T}t + \gamma_i + U_{it}t + \varepsilon_{it}, \tag{5.1}$$

$$y_{it}^{NT} = g(x_{i0})t + f^{NT}(x_{i0})t + \delta_{it}^{NT}t + \gamma_i + U_{it}t + \varepsilon_{it}. \tag{5.2}$$

y_{it} is the total employment of an establishment i at time $t = \{0, 1\}$, where 0 denotes the period before and 1 denotes the period after FDI/offshoring. Equation (5.1) describes the data generating process for the offshoring establishments and Equation (5.2) describes it for the non-treated establishments. $g(x_{i0})t$ is the function of the growth trend depending on observables x_{i0} before treatment which is independent of the treatment. $f^{NT}(x_{i0})$ captures the causal impact of offshoring also depending on the observable characteristics x_{i0}; this is allowed to be heterogeneous across establishments. The unobservable heterogeneous causal impact of the treatment is δ_{it}^{NT} which the establishments also include in their decision. γ_i are time invariant attributes that affect the outcome, both observable and/or unobservable. $U_{it}t$ varies over time and is not observable but affects the outcome, too.

Assuming, we could observe the same establishment's outcome first in the offshoring situation and then in the non-offshoring situation, $g(x_{i0})t$, γ_i and $U_{it}t$ cancel out, and we would end up with

$$f^{T}(x_{i0}) + \delta_{i1}^{T} - f^{NT}(x_{i0}) - \delta_{i1}^{NT}.$$

(2005), Heckman et al. (1998a), Heckman et al. (1998b) or Smith and Todd (2005b). Holland (1986) discusses general causal inference based on the potential outcome model and Rosenbaum and Rubin (1983) concentrate on the propensity score.

[8] See Ferracci et al. (2014) for a solution of this problem if segmentation of markets is reasonable. Moser et al. (2015) cope with the same problem by modifying their econometric model and do not exclude these effects by assumption. By conditioning on time effects, they can allow for a special case of a spillover effects. Supposing that the observations belong to the same competitive price-market, only the aggregate share of firms that decide in the period before treatment to offshore is relevant for the equilibrium employment. So they include time dummies in their selection regression to capture the amount of offshoring firms. Importantly, the ATT cannot be interpreted as usual. The resulting ATTs of this approach must then be interpreted as relative effects instead of absolute causal effects. This is sufficient for their purpose, because their main interest is to identify different channels through which activities abroad influence performance at home. As we do not want to segment markets by industries to capture integrated production strategies, and because we are interested in the causal effect, both alternatives do not seem to be appropriate here.

This difference is hypothetical. We cannot observe the counterfactual of a establishment's outcome. Therefore we have to design a counterfactual outcome conditional on the observables for every establishment and estimate the average difference in these outcomes over all observations. As mentioned above we concentrate on the ATT which can be formalized as

$$E[y_{i1}^{T} - y_{i1}^{NT} \mid D_{i1} = 1] = E[f^{T}(x_{i0}) + \delta_{i1}^{T} - f^{NT}(x_{i0}) - \delta_{i1}^{NT} \mid D_{i1} = 1],$$

where D_{it} is an indicator variable with value of one for the treatment group in period one and zero if there is no offshoring event. $E[f^{NT}(x_{i0}) + \delta_{i1}^{NT} \mid D_{i1} = 1]$ is the part we have to construct where the matching algorithms select a most similar control group on observables.

The difference-in-difference estimator is given by

$$\Delta y_{i1} = \beta_0 + \beta_1 x_{i0} + \beta_2 D_{i1} + \varepsilon_i,$$

with D_{it} as treatment indicator. This estimator needs four assumptions to estimate the ATT consistently: (i) no heterogeneous treatment effects based on observables, (ii)x_{i0} are exogenous time trend determinants, (iii) a linear functional form for the time trend and (iv) the time trend on observables x_{i0} has a common average for treated and non-treated. The last assumption implies that for a consistent estimator there are no self-selection effects into offshoring for the establishments on unobservables (i.e. $E[U_{i1} \mid D_{i1} = 1, x_{i0}] = 0$) and no heterogeneous causal effects on unobservables (i.e. $E[\delta_{i1}^{T} - \delta_{it}^{NT} \mid D_{i1} = 1, x_{i0}] = 0$).[9]

In combination with the matching estimator the difference-in-difference-matching approach relaxes the first three assumptions as described above (see for this approach Heckman et al. (1997)). The ATT under the remaining assumption of conditional mean independence is given as

$$E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 0] = E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 1] = E[\Delta y_{i1} \mid x_{i0}],$$

and the ATT in the population is

$$ATT = E[\delta_x \mid D_{i1} = 1],$$

[9]We use this estimator as a robustness check for our results by estimating a twofold differentiated equation via OLS, see Angrist and Pischke (2009).

with

$$\delta_x \equiv E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 1] - E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 0].$$

Obviously, if we try to match the observations by x_{i0} or if we try to condition on x_{i0} respectively, there is a problem of dimensionality. Consider the case of some continuous variables or a large set of categorical variables or any combination of these two as determinants of the treatment. Hence, exact matching is not useful or practicable. We prefer to match on the propensity score. The propensity score is the conditional probability of getting treated of an establishment i, $P(D_{i1} = 1) = P(x_{i0}) \equiv p_i$. Rosenbaum and Rubin (1983) show that conditioning on the propensity score p_i instead of conditioning on x_{i0} is a consistent. The central idea is that if the outcome is independent of the selection into treatment D_{i1} conditional on x_{i0}, the same is valid conditional on $P(x_{i0})$:

$$y_{it}^T, y_{it}^{NT} \perp D_{it} \mid x_{i0} \Rightarrow y_{it}^T, y_{it}^{NT} \perp D_{it} \mid P(x_{i0}).$$

The propensity score has to be estimated. Typically, a binary outcome model is used for that purpose. We choose a multinomial-logit model to estimate the propensity score of establishments to offshore (McFadden, 1974).

The idea of the propensity score matching estimator is to find for any treated observation another non-treated observation with the same estimated probability of treatment, $\hat{p}_i$, as for the treated one and compare their outcomes. But the propensity score is also a continuous variable and to find a matching partner with the same estimated $\hat{p}_i$ has zero probability in a random sample. We have to include similar observations instead of (non-existent) identical one to compare the outcomes. Various matching algorithms exist to tackle this problem. They vary in their idea of defining the "right" set of matching partners or control observations, their measurement of the distance or in weighting issues. Note that every deviation from the identical propensity score matching makes the estimated coefficient potentially biased.

In this study we employ two different, but intuitive matching strategies. We use a kernel and a k-nearest neighbor approach.[10] They differ in the number of observations included and in their underlying non-parametric weighting function $g(.)$ of the included control observations. To formalize this we follow the difference-in-difference matching ATT for-

[10] Caliendo and Kopeinig (2008) present other matching algorithms.

mulation of Heckman et al. (1997),

$$\hat{\delta} = \sum_i D_{i1} \left[\Delta y_{i1} - \sum_j \Big(((1 - D_{j1}) g(p_i, p_j) \Delta y_{j1}) \Big) \right],$$

where for this estimated $\widehat{ATT}$ the expected value is replaced by the sample mean. The weighting function for the kernel-estimator can then be formalized as

$$g(p_i, p_j) = \frac{K((p_j - p_i)/h)}{\sum_{j \in A(i)} K((p_j - p_i)/h)}.$$

$A(i) = (j \mid\mid p_i - p_j \mid < h)$ is the set of control group observations and $K(.)$ is the Epanechnikov Kernel function which defines the weights in particular.[11] h is a parameter that defines the bandwidth around the treated observation where the potential control observations are located. The bandwidth allows one to vary the number of control observations that are included for calculating the $\widehat{ATT}$ and the Epanechnikov kernel function allows one to weigh the more distant observation less in the calculation. Heckman et al. (1998b) have shown that this approach generates consistent estimates of the ATT under common assumptions.

The second estimator in this analysis is the k-nearest neighbor estimator. We use it for some variation and robustness checks and to employ the necessary balancing tests. It substitutes the function $g(.)$ of $\hat{\delta}$ with:

$$g(p_i, p_j) = \begin{cases} 1, & \text{if } j = \arg\min \mid p_i - p_j \mid \\ 0, & \text{else }. \end{cases}$$

This function uses the k-nearest non-treated neighbor observations of the treatment observation by the propensity score and weighs them with factor one. If there is only one neighbor the outcome of this one non-treated observation is compared to one treatment observation.

The choice of the bandwidth h for the Kernel approach or the number of neighbors for the k-nearest neighbors approach is a trade-off. On the one hand a bigger set of neighbors or a bigger parameter h for the bandwidth go along with a bias in the estimator; every match which is not a perfect match biases the estimator and if the bandwidth of this potential matching partners increases – i.e. h increases – also the bias increases poten-

[11] There are several other kernel functions available aside from the Epanechnikov function. For example we also use a Gaussian kernel, but it does not matter for any qualitative result.

tially.[12] The same is intuitive for the k-nearest neighbor approach. The more neighbors are included the lower is the quality of the matches by propensity score. Put differently, a distant neighbor is distant because its observable characteristics, x_{i0}, are different from the treatment observation at hand. On the other hand, every single observation added increases the efficiency of the estimator as usual. This trade-off applies to both approaches. Therefore we use the variation of the parameter h or k to check the sensitivity of our results subsequently.

One remaining problem of matching is to size the standard errors to enable inference. A general approach to get such missing standard errors is bootstrapping. This seems to be useful for matching estimators, too (Caliendo and Kopeinig, 2008). But Abadie and Imbens (2008) proof formally that bootstrapping is not valid for the nearest neighbor approach with replacement that we employ. On the other hand they suggest that bootstrapping is valid for the kernel matching estimator. Hence, for the kernel estimator we provide the bootstrapped standard errors and for the the nearest neighbor results we provide the analytical but only asymptotically valid standard errors of Abadie and Imbens (2006).[13]

As mentioned above one robustness check we perform is to vary the two different matching estimators by their parameters for the bandwidth h and different numbers of neighbors. As a second robustness check we use different logit specifications, stemming from Wagner (2011) and Moser et al. (2015).

The crucial assumption of the matching approach is the CMIA. The selection into treatment has to be exhaustively determined by observables to get consistent ATTs (Becker and Muendler, 2008). Regrettably and logically, there is no formal test of this assumption. One way to indicate validity of the CMIA is a pre-test, following Heckman and Hotz (1989), Imbens (2004), and Smith and Todd (2005a). The idea is to perform the matching estimator for the same observations but before the treatment period. If there would be a significant difference of the ATTs without treatment, conditional on the same x_{i0}, the CMIA does not seem to hold. If there is no difference a self-selection effect into treatment is less plausible.

[12]Except for the case where there are no other observations within the bandwidth.

[13]The practical implementation is done in STATA version 10.1. The point estimates and standard errors for the kernel matching stem from the PSMATCH2 package by Leuven and Sinaesi (2003) with 500 bootstrap iterations, where the propensity score estimation is repeated in every iteration. The standard errors for the nearest neighbor approach stem from the NNMATCH package (Abadie et al., 2004) which uses the calculation of Abadie and Imbens (2006) for valid standard errors. A practical guide to implement these matching estimators is given by Abadie et al. (2004) and by the help-file of the PSMATCH2 package.

After running the matching algorithms, testing whether covariates or selection variables are balanced between the matching partners is at the core of the matching estimator and indicates the quality of the matching procedure itself. Balancing in the population is not a problem (Rosenbaum and Rubin, 1983). There are three possible reasons why the balancing is not fulfilled in the sample. First, the estimated propensity score is different from the real propensity due to a misspecification of the binary model. Second, as mentioned, the matching is not an exact procedure, and third, even if the propensity score estimation is correct and the matching is exact – i.e. identical propensity scores for treated and matched-control can be found – the balancing property could be invalid due to an "unlucky" sample draw (Rosenbaum and Rubin, 1985). Hence, we have to test the balancing of the covariates between the two matching groups after the matching on the propensity score. The literature offers a set of balancing tests. We decide to perform three typical balancing tests in our analysis namely the standardized-difference test between the treatment group and the matched-control group according to Rosenbaum and Rubin (1985), a t-test of mean-difference between these groups and a t-squared Hotelling test by propensity score quantiles. The first two tests check the balancing of the covariates separately. The big advantage of the Hotelling test is that the selection variables of every matching group are tested jointly. All balancing tests are provided for the simplest case of nearest neighbor matching with one neighbor, because there are no statistical problems stemming from the weighting function in this case.

5.4. Data

As our main data source, we compile a data set of recent waves of the so-called IAB Establishment Panel. This is a stratified, annual survey on behalf of the Institute for Employment Research (IAB) from 1993 onwards on West German establishments and from 1996 onwards additionally on East German establishments. The sample is drawn from a nationwide population consisting of about two million establishments. There is no size cut-off in the panel, thus, every establishment with at least one employee who is liable to the German social security system is included. Such are all sectors, subdivided into 17 industries. The stratification occurs along the dimensions of 16 federal states (Bundesländer), establishment size class in terms of employees, and industry. Thereby, establishments of large size located in small regions, and belonging to industries with few establishments are oversampled. Within the 170 cells of the stratification matrix, the sampling is random. Establishments that refuse to answer are replaced by randomly drawn establishments of the same strata.[14]

The high quality of the data and high response rates are ensured by attributes of the survey like professional face-to-face interviews (response rates up to 84%), elaborated questionnaire designs with pre-tests and a complex editing process after the field phase with comprehensive plausibility and consistency checks.[15]

The questionnaire consists of several topic blocks like employment, business policy, investments, wages and salaries, and so on. The main interest of the survey is to collect labor market related information. Furthermore, it consists of regular and irregular questions. The former are asked every year. The latter are dependent on actual developments or policy interests and on experiences of previous questionnaires and therefore asked only once or a few times. Unfortunately, our treatment variables of offshoring and FDI belong to the irregular questions, constraining our analysis to various treatment periods in between the years 2004 and 2006 (see details below). Subsequently, we describe the variables we use in detail according to their function within the matching approach: treatment variables, outcome variable, and selection variables.

5.4.1. Treatment Variables

We are interested in the establishment-level employment effects of different modes of internationalization. We analyze five different treatments: *FDI*, *market-seeking FDI*, *cost-saving FDI*, *low-wage-region FDI* and *Relocation*. In the following, we outline the details of measurement for these treatment variables.

FDI

According to the 2006 IAB Establishment Panel questionnaire, we call an international sourcing mode *FDI* if an establishment has invested abroad in the two previous business years, i.e. usually in legal years 2004 and/or 2005. If an establishment answered to this question with "yes", it belongs to the treatment group. If it answered with "no" it potentially belongs to the control group. Altogether, 170 out of 5759 establishments

[14]See for example Fischer et al. (2008) for details on the data set. See also `http://www.iab.de/en/erhebungen/iab-betriebspanel.aspx`.

[15]For instance, implausibilities in the data are cleared up with individual telephone calls with the interviewee. Highly erroneous or implausible questionnaires are excluded from the data.

engaged in new FDI during the years 2004 or 2005 within our estimation sample.[16] For Germany this measure has also been used by Mattes (2010), albeit applying a different estimation technique. A very similar treatment is used by Barba Navaretti and Castellani (2004) (and others, see Table D.1) on Italian firm data using different selection variables.

Market-Seeking FDI

FDI can be refined further by the main objective or motive according to which an establishment made its decision on its most important foreign investment. The questionnaire of 2006 offers seven motives: penetrate new markets/protect market share, procurement options for intermediate inputs, lower costs, taxes and contributions, lower labor costs, fewer administrative regulations, option of public funding, and other motive. Multiple answers are possible.

If one motive for the most important new foreign investment was to penetrate new markets or to protect foreign market shares, but not labor cost savings at the same time, we call this mode *market-seeking FDI* which may capture horizontal or export platform FDI. There are 84 such modes of FDI in our estimation sample during the years 2004 or 2005.

Cost-Saving FDI

Likewise, we call a mode *cost-saving FDI* if one motive for the most important new foreign investment was to save labor cost, but not to penetrate new markets or to protect foreign market shares at the same time as before. While labor cost savings are associated foremost with vertical FDI, they may also be relevant for horizontal FDI (Markusen, 2002; Braconier et al., 2005). Within the estimation sample there are 25 such cases of FDI during the years 2004 or 2005.

Low-Wage-Region FDI

A different refinement of the FDI variable tracks its geographic destination. The 2006's survey offers 5 destination regions for the most important foreign investment of the establishments. First, the Euro-area, second, the new European Union members since May 2004, a third region which includes south-east Europe with Russia and Turkey, fourth,

[16] The numbers of establishments counted as treated in the estimation sample may differ from the overall number of treated observation in the whole sample due to missing values of some covariates/selection variables.

Asia and at last the rest of the world. From the point of view of German establishments, we define an investment to the second, third and/or fourth region as *low-wage-region FDI*. We find 99 establishments in the estimation sample of this mode of FDI. For example Debaere et al. (2010) find FDI to less developed countries to have negative effects on the change in employment at home for middle income country's multinationals. Although, at a glance, labor cost savings might dominate the decision to invest in such countries for German establishments, we expect market-seeking interests to be important as well, especially since the regions include big emerging markets as China and India.

Relocation

In the wave of 2007 the establishments were asked whether they closed down a domestic in-house activity in the period from July 1st, 2006 to June 30th, 2007, and whether they re-opened this particular division abroad. We count these cases as *Relocation* treatment. Note that we do not further distinguish between a cross-border spin-off or a cross-border spun-off, although possible.[17] Much more important for our purpose is that we know that for these events a domestic restructuring or downsizing is causally aligned to the internationalization of the establishment. Relocation is the only empirical measure where we can be sure that going abroad involves domestic organizational changes at the establishment-level which plausibly lead to a domestic downsizing. Establishments that did not close down any division or closed down a division but displaced it only domestically belong potentially to the control group for this treatment. Altogether there are 43 relocation cases among 6496 establishments. Note also that this measure is most likely to exclude an expansion of the firm and differs in that vein to all other employed measures.

5.4.2. Outcome Variables

Our aim is to estimate the offshoring effect on a German establishment's total employment. To capture the impact of treatment, difference-in-differences estimation compares employment before and after treatment of treated establishments with appropriately chosen establishments which are not treated during the same time period. To allow for some adjustment period, we take the difference in log employment before treatment with log

[17] The 2008th survey of the German Federal Statistical Office (Statistisches Bundesamt, 2008) supports spin-off and spun-offs to be typical offshoring events. The biggest part of German cross-national displacements are represented by foundations of new establishments within the business-network of the firm (spun-off: 50,6%), or by displacing the domestic activity to an organizationally aligned firm that already exists (spin-off: 38%).

employment up to one year after the treatment period. The treatment period of the FDI variables covers the years 2004 and 2005. Total establishment-level employment is counted at June 30th in every year. Hence, we take the difference in log employment from June 30th, 2003 to June 30th, 2006, if treatment is one of the FDI variables. Likewise, relocation abroad occurs during July 1st, 2006 and June 30th, 2007. Hence, we take the difference in log employment during the period June 30th, 2006, and June 30th, 2008.

Total employment is the most reliable variable of the IAB Establishment Panel. First, it stems from the social security register, the reporting of which is obligatory by law for the establishments. Second, it is checked before the current interview in the establishment starts. Before the interview of the following year starts, the last year's employment is checked again. In a last step it is additionally checked during the editing process (Fischer et al., 2008).

5.4.3. Selection Variables

The last group of variables are the covariates that are necessary to estimate the propensity score for every establishment. Selection variables serve as the decision criteria according to which management may have decided upon FDI or relocation. Hence, we include the lags of the time varying variables in the propensity score estimation and only the time invariant or persistent selection variables are included with their value contemporaneous to treatment, in order to loose as few observations as possible. Concerning the FDI treatments, selection variables date back to the period of the year 2003, still before treatment starts potentially on January 1st, 2004. Concerning the relocation treatment, selection variables date from the period of the year 2005 or June 30th, 2006 – still before treatment starts potentially on July 1st, 2006.

To take into account sample stratification, we always include the stratification variables among the selection variables, i.e. 16 regional dummies and 17 industry dummies, and firm size in terms of employment. In this way we take into account that relocation activity varies by industries, federal state, and firm size (see for instance the descriptive statistics in Statistisches Bundesamt (2008)).

Further, we investigate whether different choices of selection variables matter for the results. In particular, we choose the selection variables previously used in the studies of Moser et al. (2015), selection variables MUW henceforth, and of Wagner (2011), selection variables Wagner, henceforth. We use an additional specification that explains the probability of the relocation treatment better than the previous two specifications. This specification we call SU henceforth.

Selection Variables MUW

For the FDI variables we use the same selection variables as Moser et al. (2015) do. Their logit estimates show that offshoring is significantly more likely the larger an establishment in terms of full-time employees is, the more advanced its technology, the higher average wage costs, and the larger the share of high-skilled workers. If we assume these variables to proxy productivity of a firm, their choice is perfectly in line with heterogeneous firms literature following the seminal theoretical contribution of Melitz (2003). Also foreign-owned establishments seem to have a higher probability of offshoring or FDI.[18] These selection variables are measured as follows:

- *log total employment*: logarithm of total employment at establishment i at time t-1, i.e. before treatment;
- *log wage per employee*: logarithm of total wage cost per employee at establishment i at time t-1;
- *high technology*: dummy variable taking value of one if an establishment self reports to employ a technology which is above average or state-of-the-art at time t-1;
- *high-skilled*: percentage share of high-skilled employees at establishment i at time t-1;
- *foreign ownership*: dummy variable taking value of one if majority of the establishment is held by a foreign investor.

Selection Variables Wagner

Additionally we provide the same variables as Wagner (2011) for all offshoring measures as a robustness check for our results. These selection variables are measured as follows:

- *employment*: total employment at establishment i at time t-1;

[18] Similar sets of selection variables are applied for instance by Becker and Muendler (2008) or Barba Navaretti and Castellani (2004). According to the former study the firms that displace their activities internationally, stem from the high technology (manufacturing) sectors and are larger in terms of employment. Additionally, Barba Navaretti and Castellani (2004) find the size of a firm and its productivity and profitability to be relevant covariates for the treatment of international investments. Becker and Muendler (2008) also identify the domestic employment and the establishment's average wage costs per employee to be significant selection variables for their measure of foreign employment expansion. Furthermore, they employ variables that describe the skill composition of the workforce at the establishment.

- *employment squared*: total employment squared;
- *employment cubic*: cubic term of total employment;
- *sales per employee*: sales per employee at establishment i at time t-1;
- *wage per employee*: wage per employee at establishment i at time t-1;
- *export share*: percentage share of total exports of total sales at establishment i at time t-1;
- *employment change*: change of total employment at establishment i from time t-2 to t-1.[19]

Selection Variables SU

This specification adjusts the estimation of the propensity score to fit it better to the relocation case. Apart from the industry and region dummies and the firm size variable in terms of number of employees, we additionally include the export share as in the specification ofWagner (2011) and the technology variable as in the MUW specification. As a new variable we include an indicator for an establishment that belongs to a corporate group, and an indicator whether an establishment has a works council.

Affiliates of a corporate group may be more likely to be relocated, because these often are purely production units which are intensive in production workers and therefore may be relatively cheap elsewhere. Instead, headquarters are intensive in high-skilled labor which is relatively cheap in Germany. Moreover, single establishment corporations are often too small to finance foreign investments, or lack the managerial experience of supervising affiliates.

establishments with more than five employees are eligible in Germany to have a works council if there are employees who desire to have one. In fact, many, even large firms do not have a works council. The decision to close an in-house activity and to dismiss employees is a prototypical situation where a works council takes part in the decision. Because it is in the interest of the works council to secure domestic employment and works councils can increase the cost of relocation (if not block it at all), its presence is likely to reduce the probability of relocation.

[19] Note that this selection variable partly accounts for different growth paths of the treated and non-treated observations.

Quasi Natural Control Group

In addition to the designed control groups through the matching method we offer a unique quasi natural control group for the relocation variable. Within the 2006's survey (one period before treatment) the establishments were asked if an agreement for employment and location assurance with their workforce or its representation exists and of what content it is. The establishments were asked what promises they make within this agreement and which promise the bargaining workforce makes in turn. One promise of the establishments is disclaiming to outsource/relocate any activity of the establishment and possible counterparts of the agreement are typically lower wages or increased hours of work. We assume establishments that disclaim to relocate most likely to be potential offshoring units, because their workforce would not bargain about this probably expensive promise if it is unlikely to happen. We present results for the relocation variable as described above, but with a restricted control group consisting only of such disclaimers.

DESTATIS Data

The second data set we employ is a special purpose survey on relocation in 2006 on behalf of the German federal statistical office (DESTATIS). The DESTATIS data provide a comparable relocation measure on employment for German firms. These data have also been used in Wagner (2011).[20] Here a representative sample of about 20000 German establishments is interviewed about their relocation activities before 2001, between 2001 and 2003 and between 2004 and 2006. Especially they are asked for relocation which implicitly includes a restructuring at home. We merge this information to characteristics of regular reports on manufacturing establishment activities of DESTATIS via a unique establishment identifier. We end up with a second data set for a micro-level relocation measure. Again we need three types of variables: treatment, outcome, and selection variables. The outcome variable is equal to the outcome variable before, the difference in log employment from the period before relocation to the period after relocation. The relocation treatment variable is differentiated for three time periods: relocation to a foreign country in years (i) 2001-2006, (ii) 2001-2003 and (ii) 2004-2006. Unfortunately, the set of available selection characteristics is limited for this data set. Thus, we include *log employment*, *log sales per employee*, and *log wage per employee* beside 2-digit industry- and 16 regional dummies.

[20]See Wagner (2011) for a comprehensive description of the data.

5.5. Propensity Score Estimation and Matching

The first estimates we depict stem from the binary model which predicts the conditional probability for every establishment to be an offshoring establishment. We split up the auxiliary estimates into two tables. Table 5.1 presents the results from our logit specification for the different FDI measures. Column (1) presents the MUW selection variable specification for the FDI variable. The same specification is used in columns (3), (4) and (5) for the market-seeking FDI treatment (column 3), for the cost-saving FDI variable (column 4) and for the low wage region FDI (column 5). Column (2) provides the estimations from the Wagner specification as a robustness check.

As expected we find in our baseline the logarithm of the number of employees at a firm as measure for firm size to have a positive sign and to be highly significant. The same holds for the logarithm of wage per employee, the high technology measure and the skill composition of the establishment. All these coefficients have the expected signs and are highly significant. The foreign ownership dummy is significant as well, but shows a counter intuitive sign at first glance. We have expected a positive sign for foreign owned firms. To explain the negative sign, we have to keep in mind that we observe single establishments instead of whole firms or headquarters. If we observe an establishment that is foreign owned it is likely that this establishment is part of a multinational. Hence, it might be just a subsidiary. If we look at a foreign direct investment decision, as we do here, it is fair to say that this decision is undertaken more likely by the (foreign) headquarter. Hence, it might not be surprising that we find a negative sign.

If we compare the coefficients of the covariates of FDI in general to our market-seeking FDI or low wage region FDI measure we find no major differences. All point estimates stay at similar values and stay significant, too.

If we look at the covariates Wagner (2011) uses, we find no counter intuitive results. Moreover we find the same signs for every covariate as Wagner (2011) does and mostly no differences in the significance level to his trimmed baseline specification. Additionally, we find no important differences in the explanatory power across all specifications presented in Table 5.1, except for the cost saving FDI treatment. Here just the size measure turns out to be significant. All other loose their significance. We suspect the minimal number of treatment cases of 25 as the reason for this.

Table 5.2 presents the effects of covariates on the relocation decision. We provide four specifications. First, column (1) shows the coefficients of our baseline selection variable specification. Column (3) and (4) serve as robustness checks as before; therefore we use

Table 5.1.: Propensity Score Estimation – FDI

	MUW (1)	Wagner (2)	Market seeking (3)	Cost saving (4)	Low wage region (5)
ln employment	0.724***		0.713***	0.692***	0.715***
(t-1)	(0.065)		(0.086)	(0.150)	(0.081)
ln wage per employee	0.682***		0.927**	-0.132	0.613*
(t-1)	(0.266)		(0.368)	(0.559)	(0.350)
high technology	0.797***		0.807**	0.632	1.073***
(t-1)	(0.253)		(0.355)	(0.567)	(0.351)
high-skilled	1.918***		2.479***	0.657	1.954***
(t-1)	(0.406)		(0.544)	(0.958)	(0.523)
foreign ownership	-1.268***		-1.216**	-1.379	-1.028**
	(0.40)		(0.523)	(1.070)	(0.459)
employment		7.66e-04***			
(t-1)		(1.46e-04)			
employment squared		7.42e-08**			
(t-1)		(2.99e-08)			
employment cubic		1.73e-12			
(t-1)		1.27e-12			
sales per employee		-2.61e-08			
(t-1)		(8.14e-08)			
wage per employee		2.30e-04***			
(t-1)		(4.89e-05)			
export share		0.015***			
(t-1)		(0.002)			
employment change		-0.811***			
((t-2) - (t-1))		(0.299)			
17 industry dummies	yes	yes	yes	yes	yes
16 regional dummies	yes	yes	yes	yes	yes
Pseudo R^2	0.3322	0.3261	0.2791	0.1851	0.3136
Number of Obs.	5759	4972	4364	3018	5121

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level. (1) MUW: dependent variable investment abroad in business years 2004/2005; selection variables as in Moser et al. (2015). (2) Wagner: dependent variable investment abroad in business yeas 2004/2005; selection variables as in Wagner (2011). (3) Market seeking: dependent variable investment abroad in business years 2004/2005 if motivation is market seeking but not labor cost savings; selection variables as in Moser et al. (2015). (4) Cost saving: dependent variable investment abroad in business years 2004/2005 if motivation is labor cost saving but not market seeking; selection variables as in Moser et al. (2015). (5) Low wage region: dependent variable investment abroad in business years 2004/2005 to Asia, new EU members or Russia and south-east Europe; selection variables as in Moser et al. (2015).

the MUW and the Wagner (2011) selection variables. Column (2) presents the results for the quasi natural control group.

The baseline shows the expected positive and significant signs for size, export share and the affiliate dummy. Works councils are found to have a significant negative impact on the probability to offshore, too. Contrary to the FDI cases, establishments that relocate abroad self-assess to be further away from their technology frontier than establishments that do not relocate. In specifications (3) and (4), only the size and the export share variables remain significant with the expected signs. For the quasi natural control group specification we find the export share, the affiliate and the works council dummy to be significant determinants with the expected signs. Here the size and the high technology status loose their explanatory power.

Table 5.3 reports the results of the propensity score estimation for the DESTATIS data. Again, as expected we find a significant effect of the size measure with respect to *log employment*. The two selection variables left, *log sales per employee* and *log wage per employee*, remain insignificant for this data set.

Tables 5.4 and 5.5 provide the balancing tests for the general FDI indicator between the treatment and matched-control observations. Tables 5.7 and 5.8 do so for the relocation variable. Unfortunately there is no analytical measure for the standardized difference test but a percent bias below 20 is mentioned by Rosenbaum and Rubin (1985) to be sufficient to state balanced covariates. None of the remaining percent biases after the matching process reaches this critical value. Also the mean difference t-test in column five does not reject the null hypothesis. All p-values are far away from indicating an unbalanced variable. The last balancing test of Hotelling is performed over three quantiles and the hypothesis of an unbalanced composition in treatment and matched-control group is clearly rejected.

Tables 5.6 and 5.9 provide the pre-test test for FDI variables and for relocation. The first column compares only the baseline estimates of the matching procedure with a standard difference-in-differences approach which employs the OLS estimator on a differentiated estimation equation.[21] According to the idea of the test, all outcomes stem from the last and the second last period before treatment, respectively. None of the ATTs show a significant difference before treatment for the same matching partners as in the actual matching period with treatment. Hence, we do not find an indication of a violation of the CMIA assumption.

[21] This standard approach is reported in all ATT output tables in the following.

Table 5.2.: Propensity Score Logit Estimation – Relocation

	SU	Quasi natural control group	MUW	Wagner
	(1)	(2)	(3)	(4)
ln employment	0.396***	-0.084	0.228**	
(t-1)	(0.121)	(0.210)	(0.101)	
high technology	-0.570*	0.333	-0.419	
(t-1)	(0.330)	(0.683)	(0.309)	
export share	0.023***	0.028**		0.009***
(t-1)	(0.006)	(0.012)		(0.003)
affiliate	0.782***	1.522**		
	(0.365)	(0.701)		
works council	-1.049***	-5.299***		
	(0.460)	(1.120)		
log wage per employee			-0.086	
(t-1)			(0.335)	
high-skilled			0.147	
(t-1)			(0.592)	
foreign ownership			0.783	
			(0.415)	
employment				3.87e-04***
(t-1)				(1.47e-04)
employment squared				-3.19e-08*
(t-1)				1.81e-08
employment cubic				4.81e-13
(t-1)				3.98e-13
sales per employee				-1.13e-07
(t-1)				(3.79e-07)
wage per employee				2.69e-05
(t-1)				(7.96e-05)
employment change				-0.175
((t-2) - (t-1))				(0.373)
17 industry dummies	yes	yes	yes	yes
16 regional dummies	yes	yes	yes	yes
Pseudo R^2	0.1259	0.4159	0.0819	0.1262
Number of Obs.	6496	214	7347	5271

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level. (1) SU: dependent variable displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; selection variables as described in text. (2) Quasi natural control group: dependent variable displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; selection variables as described in text. (3) MUW: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; selection variables as in Moser et al. (2015). (4) Cost saving: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; selection variables as in Wagner (2011).

Table 5.3.: Propensity Score Logit Estimation – Relocation DESTATIS

	relocation 04-06 (1)	relocation 01-03 (2)	relocation 01-06 (3)
log employment	0.446***	0.429***	0.358***
(t-1)	(0.059)	(0.085)	(0.075)
log sales per employee	0.0834	0.000	0.008
(t-1)	(0.125)	(0.197)	(0.162)
log wage per employee	0.203	-0.083	0.135
(t-1)	(0.286)	(0.497)	(0.411)
2-digit industry dummies	yes	yes	yes
16 regional dummies	yes	yes	yes
Pseudo R^2	0.1056	0.0934	0.1008
Number of Obs.	2674	1259	1283

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level. (1) relocation 04-06: dependent variable displacement of an in-house activity to a foreign country in period 2004 to 2006; selection variables as described in text. (2) relocation 01-03: dependent variable displacement of an in-house activity to a foreign country in period 2001 to 2003; selection variables as described in text. (3) relocation 01-06: displacement of an in-house activity to a foreign country in period 2001 to 2006; selection variables as described in text.

5.6. Results

We present our results of the ATTs of FDI and relocation on employment again in separated tables. Table 5.10 covers the specifications of FDI treatment variables. We present the ATTs for different bandwidths of kernel matching and different number of neighbors for k-nearest-neighbor matching. Table 5.11 covers the relocation treatment but reports for the baseline specification of covariates an additional column where the change in employment is measured one period later. Table 5.12 presents the results for the DESTATIS data.

For our FDI measures we find a robust positive treatment effect on the employment. Additionally, we cannot state a significant difference in the point estimates between the different measures of FDI. Hence, employment effects of FDI in general, market-seeking FDI, or low wage region FDI seem to have similar effects on employment. The cost saving FDI treatment does not yield significant effects at all, most possibly driven by the low number of observations. This result is in line with most findings of previous studies. In contrast to Debaere et al. (2010) we find positive employment effects, if the destination is a low wage country, indicating that the market seeking motive might dominate for German establishments in these countries.

Table 5.4.: Balancing Tests from Nearest-Neighbor-Matching – FDI

Covariate	Mean treatment group	Mean matched control group	% bias	Percent bias reduction	Mean difference test
ln employment	5.3857	5.4260	-2.5	98.3	-0.22(0.83)
ln wage per employee	7.8654	7.8651	0.1	99.9	0.01(0.99)
high technology	0.8765	0.9000	-5.8	88.9	-0.69(0.49)
high-skilled	0.5034	0.5021	0.4	95.6	0.04(0.96)
foreign ownership	0.0529	0.0529	0.0	100.0	0.00(1.00)

Notes: p-values in parenthesis; matching method: NN-matching; number of neighbors: one; caliper: no; treatment variable: investment abroad in the business years 2004 and/or 2005.

Table 5.5.: Hotelling's T-squared Test by Propensity Score 3-Quantile – FDI

Quantile	Frequency treatments	Frequency matched controls	T-squared statistics	F-Test statistics	p-value
First	52	48	38.825	0.7924	0.7654
Second	52	48	26.216	0.7511	0.7908
Third	66	33	21.143	0.7530	0.7700

Notes: Hotelling's T-squared Test for 3 Quantile for all covariates jointly; matching method: NN-matching; number of neighbors: one; no caliper; treatment variable: investment abroad in the business years 2004 and/or 2005.

Table 5.6.: Heckman and Hotz Pre-Test – FDI

Time	OLS for FDI	ATT for FDI
t-1	0.029**	0.013
	(0.012)	(0.019)

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors (White, 1980); matching method: kernel matching; weighting: epanechnikov; bandwidth: 0.01; standard errors are generated via bootstrapping with 500 replications; treatment variable: investment abroad in the business years 2004 and/or 2005.

The picture looks quite different if we look at the results for relocation (Table 5.11). Here, all point estimates are negative and mostly significant at the common levels. For the quasi natural control group we find very similar results to the estimated ATTs. Additionally, we do not find qualitatively different results for the OLS difference-in-difference estimates. The point estimates somehow differ in size – what is expected through a self selection of establishments into internationalization – but not by their sign. Finally, in Table 5.12 we provide the results from relocation for the DESTATIS data. Again we find negative point estimates for all three measures or time periods as described above. The quantitative difference might be due to different sizes of the observational units. For the IAB data we do not have a size cutoff, while the DESTATIS questionnaire only includes firms with at least 100 employees. These effects are qualitatively comparable to the relocation effects Wagner (2011) finds in some samples, but differ quantitatively with much bigger negative effects on employment for our results.

FDI expansion – independently of the type of FDI – seems to create jobs at domestic establishments or occurs in firms that expand both at home and abroad. Only in cases, when domestic production is substituted for foreign production while the firm stagnates, negative employment effects show up. This result is in line with Moser et al. (2015) who also use data on German establishments, but covering a different time period. They find positive employment effects from the increase in intermediate input purchases, but negative employment effects from the treatment where intermediate input purchases rise simultaneous to partial establishment closure. This suggests that potential negative effects on employment at the establishment-level due to restructuring dominate potential

Table 5.7.: Balancing Tests from Nearest-Neighbor-Matching – Relocation

Covariate	Mean treatment group	Mean matched control group	Percent bias	Percent bias reduction	Mean difference test
ln employment	4.4883	4.4352	2.7	96.0	0.11(0.92)
exports	30.721	32.349	6.1	92.6	-0.22(0.82)
affiliate	0.3256	0.3721	-11.0	73.0	-0.45(0.66)
works council	0.4419	0.4651	-4.9	86.8	-0.21(0.83)
high technology	0.6047	0.6744	-14.5	8.0	-0.67(0.51)

Notes: p-values in parenthesis; matching method: NN-matching; number of neighbors: one; caliper: no; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.

Table 5.8.: Hotelling's T-squared Test by Propensity Score 3-Quantile – Relocation

Quantile	Frequency treatments	Frequency matched controls	T-squared statistics	F-Test statistics	p-value
First	12	23	26.368	0.7990	0.6756
Second	15	15	60.285	0.4485	0.9157
Third	16	16	66.911	0.6505	0.7975

Notes: Hotelling's T-squared Test for 3 quantiles for all covariates jointly; matching method: NN-matching; number of neighbors: one; no caliper; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.

Table 5.9.: Heckman and Hotz Pre-Test – Relocation

Time	OLS relocation	ATT Relocation
t-1	-0.042 (0.027)	-0.038 (0.058)

Notes: standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors (White, 1980); matching method: kernel matching; weighting: epanechnikov; bandwidth: 0.01; standard errors are generated via bootstrapping with 500 replications; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.

productivity effects in cases where we observe a closure of an in-house activity. Our negative employment effects at the establishment-level are even bigger which might be driven by the fact that we can causally link a domestic restructuring to our relocation cases.

Our results seem to be sensitive to the mode of internationalization rather than to the estimation method, the choice of control or selection variables, or the employed data set. Qualitative differences in micro-level employment effects in the literature may be explained by differences in the actual proxy variables which are used to measure different theoretical concepts.

5.7. Conclusion

Empirical studies on employment effects of offshoring or FDI obtain opposing results. To understand why results differ so much, we have been investigating how different measures of offshoring or FDI impact on domestic employment in German establishments using different estimation techniques, and control or selection variables. While neither estimation techniques, nor the choice of variables is decisive for opposing employment effects, positive employment effects arise from FDI, market-seeking FDI, and even cost-saving FDI. Instead, negative employment effects derive from relocation abroad. We explain this disparity of results by the different types of FDI that are captured with the various measures. In most cases, FDI expansion may occur in the vein of a general expansion of a multinational firm, creating jobs both at home or abroad. In other cases, expansion abroad may even stimulate activities at home. Yet, in other cases, foreign activities may

Table 5.10.: ATTs – FDI

	MUW (1)	Wagner (2)	Market seeking (3)	Cost saving (4)	Low wage region (5)
OLS DiD	0.047	0.033	0.067**	0.062	0.047*
	(0.029)	(0.022)	(0.033)	(0.042)	(0.028)
kernel 0.01	0.087***	0.064*	0.103**	0.054	0.071*
	(0.028)	(0.033)	(0.045)	(0.047)	(0.036)
kernel 0.03	0.083***	0.047	0.111***	0.059	0.078**
	(0.027)	(0.031)	(0.040)	(0.043)	(0.034)
kernel 0.05	0.083***	0.047	0.112***	0.062	0.079**
	(0.026)	(0.029)	(0.039)	(0.043)	(0.032)
NN 1	0.095***	0.087***	0.092**	0.035	0.091**
	(0.035)	(0.042)	(0.039)	(0.061)	(0.041)
NN 2	0.081***	0.062*	0.109***	0.047	0.077**
	(0.028)	(0.034)	(0.038)	(0.053)	(0.034)
NN 3	0.074***	0.065*	0.114***	0.076	0.072**
	(0.025)	(0.034)	(0.038)	(0.049)	(0.030)
treated Obs.	170	148	84	25	99

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors (White, 1980); matching estimator PSMATCH2 (Leuven and Sinaesi, 2003); Kernel-matching: epanechnikov kernel; standard errors are generated via bootstrapping with 500 replications; NN-matching: no caliper; standard errors stem from Abadie and Imbens (2006) via NNMATCH (Abadie et al., 2004). (1) MUW: treatment investment abroad in business years 2004/2005; propensity score estimation Table 5.1, specification (1); control group establishments without treatment. (2) Wagner: treatment investment abroad in business yeas 2004/2005; propensity score estimation 5.1, specification (2); control group establishments without treatment. (3) Market seeking: treatment investment abroad in business years 2004/2005 if motivation is market seeking but not labor cost savings; propensity score estimation 5.1, specification (3); control group establishments without treatment. (4) Cost saving: treatment investment abroad in business years 2004/2005 if motivation is labor cost saving but not market seeking; propensity score estimation 5.1, specification (4); control group establishments without treatment. (5) Low wage region: treatment investment abroad in business years 2004/2005 to Asia, new EU members or Russia and south-east Europe; propensity score estimation 5.1, specification (5); control group establishments without treatment.

Table 5.11.: ATTs – Relocation

	SU (1)	Quasi natural control group (2)	MUW (3)	Wagner (4)
OLS DiD/ in (2)	-0.148*	-0.244***	-0.326*	-0.043**
mean comparison	(0.079)	(0.089)	(0.191)	(0.020)
kernel 0.01	-0.325*	-0.047	-0.310*	-0.356
	(0.170)	(0.416)	(0.180)	(0.221)
kernel 0.03	-0.328*	-0.263	-0.310*	-0.346
	(0.177)	(0.410)	(0.179)	(0.225)
kernel 0.05	-0.330*	-0.477	-0.310*	-0.344
	(0.178)	(0.352)	(0.179)	(0.223)
NN1	-0.365**	-0.459*	-0.287	-0.068
	(0.146)	(0.264)	(0.189)	(0.168)
NN2	-0.362***	-0.432*	-0.265*	-0.339
	(0.134)	(0.259)	(0.160)	(0.236)
NN3	-0.348	-0.462**	-0.307*	-0.361
	(0.188)	(0.232)	(0.163)	(0.288)
treated Obs.	43	40	48	37

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors (White, 1980); matching estimator PSMATCH2 (Leuven and Sinaesi, 2003); Kernel-matching: epanechnikov kernel; standard errors are generated via bootstrapping with 500 replications; NN-matching: no caliper; standard errors stem from Abadie and Imbens (2006) via NNMATCH (Abadie et al., 2004). (1) SU: treatment displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; propensity score estimation Table 5.2, specification (1); control group establishments without treatment. (2) Quasi natural control group: treatment displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; propensity score estimation Table 5.2, specification (2); control group establishments without treatment if they had disclaimed to relocate in an agreement with their workforce. (3) MUW: treatment displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; propensity score estimation Table 5.2, specification (3); control group establishments without treatment. (4) Wagner: treatment displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; propensity score estimation Table 1, specification (4); control group establishments without treatment.

Table 5.12.: ATTs – Relocation DESTATIS

	relocation 04-06 (1)	relocation 01-03 (2)	relocation 01-06 (3)
kernel 0.01	-0.026*	-0.008	-0.039
	(0.014)	(0.019)	(0.028)
kernel 0.03	-0.026*	-0.009	-0.044
	(0.014)	(0.018)	(0.027)
kernel 0.05	-0.027*	-0.008	-0.042
	(0.013)	(0.017)	(0.027)
NN1	-0.030*	-0.004	-0.048
	(0.016)	(0.023)	(0.032)
NN2	-0.026*	-0.001	-0.045*
	(0.014)	(0.020)	(0.026)
NN3	-0.020	-0.001	-0.057**
	(0.014)	(0.018)	(0.026)
treated Obs.	535	210	348

Notes: Standard errors in parenthesis; *** 1%, ** 5%, * 10% significance level; matching estimator PSMATCH2 (Leuven and Sinaesi, 2003); Kernel-matching: epanechnikov kernel; standard errors are generated via bootstrapping with 500 replications; NN-matching: no caliper; standard errors stem from Abadie and Imbens (2008) via NNMATCH (Abadie et al., 2004). (1) relocation 04-06: dependent variable displacement of an in-house activity to a foreign country in period 2004 to 2006; propensity score estimation Table 5.3, specification (1); control group establishments without treatment. (2) relocation 01-03: dependent variable displacement of an in-house activity to a foreign country in period 2001 to 2003; propensity score estimation Table 5.3, specification (2); (3) relocation 01-06: displacement of an in-house activity to a foreign country in period 2001 to 2006; selection variables as described in text; propensity score estimation Table 5.3, specification (3); control group establishments without treatment.

substitute for domestic activities while the firm as a whole stagnates. Different measures of offshoring or FDI capture those cases in different proportions.

Bibliography

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

Abadie, A., Drukker, D., Herr, J. L., and Imbens, G. W. (2004). Implementing Matching Estimators for Average Treatment Effects in Stata. *Stata Journal*, 4(3):290–311.

Abadie, A. and Imbens, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2008). On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, 76(6):1537–1557.

Adserà, A. and Pytlikova, M. (2015). The Role of Language in Shaping International Migration. *The Economic Journal*, 125(586):F49–F81.

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review*, 69(1):106–16.

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, 3:133–160.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Anderson, J. E. and Yotov, Y. V. (2010). The Changing Incidence of Geography. *American Economic Review*, 100(5):2157–86.

Angrist, J. D. and Hahn, J. (2004). When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects. *The Review of Economics and Statistics*, 86(1):58–72.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics - An Empiricist's Companion.* Princeton University Press, Princeton, 1 edition.

Antràs, P. and Yeaple, S. R. (2014). Chapter 2 - Multinational Firms and the Structure of International Trade. In *Handbook of International Economics*, volume 4, pages 55–130. Elsevier B.V.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New Trade Models, Same Old Gains? *American Economic Review*, 102(1):94–130.

Armington, P. S. (1969). A Theory of Demand for Products Distinguished by Place of Production. *Staff Papers - International Monetary Fund*, 16(1):159–178.

Baier, S. L. and Bergstrand, J. H. (2007). Do Free Trade Agreements Actually Increase Members' International Trade? *Journal of International Economics*, 71(1):72–95.

Baldwin, R. (2006). Globalisation: The Great Unbundling(s). *Paper prepared for Finnish Prime Minister's Office for EU Presidency.*

Baldwin, R. and Harrigan, J. (2011). Zeros, Quality, and Space: Trade Theory and Trade Evidence. *American Economic Journal: Microeconomics*, 3(2):60–88.

Barba Navaretti, G. and Castellani, D. (2004). Investments Abroad and Performance at Home Evidence from Italian Multinationals. *CEPR Discussion Papers*, 4284.

Barba Navaretti, G., Castellani, D., and Disdier, A.-C. (2010). How Does Investing in Cheap Labour Countries Affect Performance at Home? Firm-Level Evidence from France and Italy. *Oxford Economic Papers*, 62(2):234–260.

Becker, S. O., Ekholm, K., Jäckle, R., and Muendler, M.-A. (2005). Location Choice and Employment Decisions: A Comparison of German and Swedish Multinationals. *Review of World Economics*, 141(4):693–731.

Becker, S. O. and Muendler, M.-A. (2008). The Effect of FDI on Job Security. *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Beine, M., Bertoli, S., and Fernández-Huertas Moraga, J. (2015). A Practitioners' Guide to Gravity Models of International Migration. *The World Economy*, pages n/a–n/a.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1):30 – 41.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2006). Trade Costs, Firms and Productivity. *Journal of Monetary Economics*, 53(5):917–937.

Berthelon, M. and Freund, C. (2008). On the Conservation of Distance in International Trade. *Journal of International Economics*, 75(2):310–320.

Bertoli, S. and Fernández-Huertas Moraga, J. (2013). Multilateral Resistance to Migration. *Journal of Development Economics*, 102(C):79–100.

Biscourp, P. and Kramarz, F. (2007). Employment, Skill Structure and International Trade: Firm-level Evidence for France. *Journal of International Economics*, 72(1):22–51.

Braconier, H., Norbäck, P.-J., and Urban, D. M. (2005). Multinational Enterprises and Wage Costs: Vertical FDI Revisited. *Journal of International Economics*, 67(2):446–470.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains from Variety. *The Quarterly Journal of Economics*, 121(2):541–585.

Brun, J.-F., Carrère, C., Guillaumont, P., and de Melo, J. (2005). Has Distance Died? Evidence from a Panel Gravity Model. *World Bank Economic Review*, 19(1):99–120.

Buch, C. M., Kleinert, J., and Toubal, F. (2004). The Distance Puzzle: on the Interpretation of the Distance Coefficient in Gravity Equations. *Economics Letters*, 83(3):293–298.

Buch, C. M. and Lipponer, A. (2010). Volatile Multinationals? Evidence from the Labor Demand of German Firms. *Labour Economics*, 17(2):345–355.

Caliendo, M. and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1):31–72.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press, Cambridge, United Kingdom.

Carey, H. C. (1858). *Principles of Social Science.* J.B. Lippincott, Philadelphia.

Chiswick, C. U. (2015). Cameo 2 â€“ Immigrants and Religion. In *Handbook of the Economics of International Migration*, volume 1, pages 375–385.

Costinot, A. and Rodríguez-Clare, A. (2014). Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, volume 4, pages 197–261. Elsevier B.V.

Crinó, R. (2009). Offshoring, Multinationals and Labour Market: A Review of the Empirical Literature. *Journal of Economic Surveys*, 23(2):197–249.

Cuyvers, L., Dhyne, E., and Soeng, R. (2010). The Effects of Internationalisation on Domestic Labour Demand Skills: Firm-Level Evidence for Belgium. *National Bank of Belgium Working Paper*, 206.

Debaere, P., Lee, H., and Lee, J. (2010). It Matters Where You Go - Outward Foreign Direct Investment and Multinationals Employment Growth at Home. *Journal of Development Economics*, 91(2):301–309.

Dehejia, R. H. (2005). Practical Propensity Score Matching: a Reply to Smith and Todd. *Journal of Econometrics*, 125(1-2):355–364.

Dehejia, R. H. and Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84(1):151–161.

Desai, M. A., Foley, C. F., and Hines, J. R. (2009). Domestic Effects of the Foreign Activities of US Multinationals. *American Economic Journal: Economic Policy*, 1(1):181–203.

Dhyne, E. and Guerin, S. S. (2014). Outward Foreign Direct Investment and Domestic Performance: In Search of a Causal Link. *National Bank of Belgium Working Paper*, 272.

Disdier, A.-C. and Head, K. (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics*, 90(1):37–48.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297 – 308.

Duranton, G. and Storper, M. (2008). Rising Trade Costs? Agglomeration and Trade with Endogenous Transaction Costs. *Canadian Journal of Economics*, 41(1):292–319.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779.

Egger, P. H. and Larch, M. (2008). Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis. *Journal of International Economics*, 76(2):384–399.

Egger, P. H., Larch, M., Staub, K. E., and Winkelmann, R. (2011). The Trade Effects of Endogenous Preferential Trade Agreements. *American Economic Journal: Economic Policy*, 3(3):113–143.

Engel, D. and Procher, V. (2013). Home Firm Performance After Foreign Investments and Divestitures. *The World Economy*, 36(12):1478–1493.

Feenstra, R. C. (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press.

Felbermayr, G. J. and Kohler, W. (2006). Exploring the Intensive and Extensive Margins of World Trade. *Review of World Economics*, 142(4):642–674.

Ferracci, M., Jolivet, G., and van den Berg, G. J. (2014). Evidence of Treatment Spillovers Within Markets. *The Review of Economics and Statistics*, 96(5):812–823.

Figueiredo, E., Lima, L. R., and Orefice, G. (2016). Migration and Regional Trade Agreements: A (New) Gravity Estimation. *Review of International Economics*, 24(1):99–125.

Fischer, G., Janik, F., Müller, D., and Schmucker, A. (2008). The IAB Establishment Panel - from Sample to Survey to Prediction. *FDZ Methodenrteport*.

Flam, H. and Nordström, H. (2011). Gravity Estimation of the Intensive and Extensive Margins of Trade: An Alternative Procedure with Alternative Data. *CESifo Working Paper Series*, 3387.

Friedman, T. (2005). *The World is Flat: a Brief History of the Twenty-First Century*. Farrar Straus and Giroux, New York.

Glaeser, E. L. and Kohlhase, J. E. (2004). Cities, Regions and the Decline of Transport Costs. *Papers in Regional Science*, 83(1):197–228.

Greene, W. H. (2012). *Econometric Analysis*. Pearson, Upper Saddle River, United States, 7 edition.

Grogger, J. and Hanson, G. H. (2011). Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, 95(1):42–57.

Harrison, A. E. and Mcmillan, M. S. (2011). Offshoring jobs? Multinationals and U.S. Manufacturing Employment. *The Review of Economics and Statistics*, 93(3):857–875.

Head, K. and Mayer, T. (2014). Chapter 3 - Gravity Equations: Workhorse, Toolkit, and Cookbook. In *Handbook of International Economics*, volume 4, pages 131–195. Elsevier B.V.

Head, K., Mayer, T., and Ries, J. (2010). The Erosion of Colonial Trade Linkages after Independence. *Journal of International Economics*, 81(1):1 – 14.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–61.

Heckman, J. J. and Hotz, V. J. (1989). Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of the American Statistical Association*, 84(408):862– 874.

Heckman, J. J., Ichimura, H., Smith, J. A., and Todd, P. E. (1998a). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5):1017–1098.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4):605.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1998b). Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies*, 65(2):261–294.

Heid, B. S. and Larch, M. (2016). Gravity with Unemployment. *Journal of International Economics*, forthcomin.

Helpman, E., Melitz, M. J., and Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, 123(2):441–487.

Hering, L., Inui, T., and Poncet, S. (2010). The Elusive Impact of Investing Abroad for Japanese Parent Firms: Can Disaggregation According to FDI Motives Help? *CEPII Working Paper*, 2010-01.

Hijzen, A., Jean, S., and Mayer, T. (2011). The Effects at Home of Initiating Production Abroad: Evidence from Matched French Firms. *Review of World Economics*, 147(3):457–483.

Hillberry, R. H. and Hummels, D. (2008). Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data. *European Economic Review*, 52(3):527–550.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945– 960.

Hummels, D. (2007). Transportation Costs and International Trade in the Second Era of Globalization. *Journal of Economic Perspectives*, 21(3):131–154.

Hummels, D. and Klenow, P. J. (2005). The Variety and Quality of a Nation's Exports. *American Economic Review*, 95(3):704–723.

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *The Review of Economics and Statistics*, 86(1):4–29.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Imbriani, C., Pittiglio, R., and Reganati, F. (2011). Outward Foreign Direct Investment and Domestic Performance: the Italian Manufacturing and Services Sectors. *Atlantic Economic Journal*, 39(4):369–381.

Kleinert, J. and Toubal, F. (2008). The Impact of Locating Production Abroad on Activities at Home: Evidence from German Firm-Level Data. *Thübinger Diskussionsbeiträge*, 314.

Kramarz, F. (2008). Offshoring, Wages, and Employment: Evidence from Data Matching Imports, Firms, and Workers. *CREST Working Paper*.

Krautheim, S. (2012). Heterogeneous Firms, Exporter Networks and the Effect of Distance on International Trade. *Journal of International Economics*, 87(1):27–35.

LaLonde, R. J. (1986). Evaluating the Econometric Evaliuations of Training Programs with Experimental Data. *American Economic Review*, 76(4):604–620.

Larch, M., Norbäck, P.-J., Sirries, S., and Urban, D. M. (2015). Heterogeneous Firms, Globalisation and the Distance Puzzle. *The World Economy*, pages n/a–n/a.

Lawless, M. and Whelan, K. (2007). A Note on Trade Costs and Distance. *Central Bank of Ireland Research Technical Papers*, 7/RT/07.

Leuven, E. and Sinaesi, B. (2003). PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing, Software.

Liu, W.-H. and Nunnenkamp, P. (2011). Domestic Repercussions of Different Types of FDI: Firm-Level Evidence for Taiwanese Manufacturing. *World Development*, 39(5):808–823.

Mankiw, N. G. and Swagel, P. (2006). The Politics and Economics of Offshore Outsourcing. *Journal of Monetary Economics*, 53(5):1027–1056.

Markusen, J. R. (2002). *Multinational Firms and the Theory of International Trade*. MIT Press, Cambridge MA, London.

Marschak, J. (1959). Binary Choice Constraints on Random Utility Indicators. *Cowles Foundation Discussion Papers*, 74.

Mattes, A. (2010). The Impact of Horizontal and Vertical FDI on Labor Demand for Different Skill Groups. *IAW Discussion Papers*, 59.

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, pages 105–143.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M. J. and Redding, S. J. (2014). Chapter 1 - Heterogeneous Firms and Trade. In *Handbook of International Economics*, volume 4, pages 1–54. Elsevier B.V.

Moser, C., Urban, D., and Weder Di Mauro, B. (2015). On the Heterogeneous Employment Effects of Offshoring: Identifying Productivity and Downsizing Channels. *Economic Inquiry*, 53(1):220–239.

Nicita, A. and Olarreaga, M. (2001). Trade and Production, 1976-99. *Policy Research Working Paper Series*, 2701.

Orefice, G. (2015). International Migration and Trade Agreements: the New Role of PTAs. *Canadian Journal of Economics*, 48(1).

Ortega, F. and Peri, G. (2013). The Effect of Income and Immigration Policies on International Migration. *Migration Studies*, 1(1):47–74.

Poschke, M. (2014). The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology. *IZA Discussion Papers*, 7991.

Rauch, J. E. (1999). Networks versus Markets in International Trade. *Journal of International Economics*, 48(1):7–35.

Ravenstein, E. . G. . (1885). The Laws of Migration. *Journal of the Statistical Society of London*, 48(2):167–235.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Obsevational Studies for Causal Effects. *Biometrika*, 70(1):41–55.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, 39(1):33–38.

Sampford, M. (1953). Some Inequalities on Mill's Ratio and Related Functions. *The Annals of Mathematical Statistics*, 24(1):130–132.

Samuelson, P. A. (1952). The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent. *The Economic Journal*, 62(246):278–304.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Santos Silva, J. M. C. and Tenreyro, S. (2011). Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator. *Economics Letters*.

Santos Silva, J. M. C. and Tenreyro, S. (2015). Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically. *Oxford Bulletin of Economics and Statistics*, 77(1):93–105.

Smith, J. A. and Todd, P. E. (2005a). Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, 125(1-2):305–353.

Smith, J. A. and Todd, P. E. (2005b). Rejoinder. *Journal of Econometrics*, 125:365–375.

Statistisches Bundesamt (2008). Verlagerung Wirtschaftlicher Aktivitäten - Ergebnisse der Piloterhebung. Technical report, Statistisches-Bundesamt, Wiesbaden.

Stiebale, J. and Trax, M. (2011). The Effects of Cross-Border M&As on the Acquirers ' Domestic Performance: Firm-Level Evidence. *Canadian Journal of Economics*, 44(3):957–990.

Temouri, B. Y. and Driffield, N. L. (2009). Does German Foreign Direct Investment Lead to Job Losses at Home? *Applied Economics Quarterly*, 55(3):1–21.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, 1 edition.

Venables, A. J. (1999). Fragmentation and Multinational Production. *European Economic Review*, 43(4-6):935–945.

Wagner, J. (2011). Offshoring and Firm Performance: Self-Selection, Effects on Performance, or Both? *Review of World Economics*, 147(2):217–247.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838.

Yotov, Y. V. (2012). A Simple Solution to the Distance Puzzle in International Trade. *Economics Letters*, 117(3):794–798.

Zimmermann, K. F. (2014). Think in The Interest of the People. *The Atlantic Times*.

A. Appendix to Chapter 2

A.1. Derivation of Multinomial-Logit Probabilities

I follow a standard discrete choice setting à la McFadden (1974) to derive multinomial-logit probabilities. The presentation closely follows Train (2003). The probability of the decision of worker h to migrate from o to d, P_{odh}, is given by

$$\begin{aligned} P_{odh} &= \text{Prob}(V_{od} + \varepsilon_{odh} > V_{ok} + \varepsilon_{okh} \; \forall\, k \neq d) \\ P_{odh} &= \text{Prob}(\varepsilon_{okh} < \varepsilon_{odh} + V_{od} - V_{ok} \; \forall\, k \neq d), \end{aligned} \tag{A.1}$$

where V_{od} gives the observed part of the utility from immigration to country d for all workers from o, V_{ok} gives the observed part of the utility from immigrating to any other country k from o for all workers from o and ε_{odh} and ε_{okh} are the worker and country-pair specific utility components correspondingly. V's are known to the researcher and ε's are private information to the workers and assumed independently, identically distributed extreme value with density

$$f(\varepsilon_{odh}) = e^{-\varepsilon_{odh}} e^{-e^{-\varepsilon_{odh}}}, \tag{A.2}$$

and cumulative distribution

$$F(\varepsilon_{odh}) = e^{-e^{-\varepsilon_{odh}}}. \tag{A.3}$$

Conditional on ε_{odh} Equation (3.9) is the cumulative distribution for each ε_{okh} evaluated at $\varepsilon_{odh} + V_{od} - V_{ok}$ given by Equation (B.2). For independent εs we can write the cumulative distribution over all other alternatives as the product of the individual cumulative distributions. As usual we reach the unconditional P_{odh} by integrating $P_{odh} \mid \varepsilon_{odh}$ over all possible values of ε_{odh} for the given density (B.1):

$$P_{odh} = \int \left(\prod_{k \neq d} e^{-e^{-(\varepsilon_{odh} + V_{od} - V_{ok})}} \right) e^{-\varepsilon_{odh}} e^{-e^{-\varepsilon_{odh}}} \, d\varepsilon_{odh}. \tag{A.4}$$

Starting from this expression we can derive the standard multinomial-logit choice probability expression.

$$P_{odh} = \int_{\varepsilon_{odh}=-\infty}^{\infty} \left(\prod_k e^{-e^{-(\varepsilon_{odh}+V_{od}-V_{ok})}} \right) e^{-\varepsilon_{odh}} d\varepsilon_{odh} \tag{A.5}$$

$$= \int_{\varepsilon_{odh}=-\infty}^{\infty} \exp\left(-\sum_k -e^{-(\varepsilon_{odh}+V_{od}-V_{ok})} \right) e^{-\varepsilon_{odh}} d\varepsilon_{odh} \tag{A.6}$$

$$= \int_{\varepsilon_{odh}=-\infty}^{\infty} \exp\left(-e^{-\varepsilon_{odh}} \sum_k -e^{-(V_{od}-V_{ok})} \right) e^{-\varepsilon_{odh}} d\varepsilon_{odh}. \tag{A.7}$$

If we define $t = e^{(-\varepsilon_{odh})}$ such that $-e^{(-\varepsilon_{odh})} d\varepsilon_{odh} = dt$ and noting that t approaches 0 as ε_{odh} goes to infinity and t is infinite if ε_{odh} approaches negative infinity, we get

$$P_{odh} = \int_{\infty}^{0} \exp\left(-t \sum_k e^{-(V_{od}-V_{ok})} \right) (-dt) \tag{A.8}$$

$$= \int_{0}^{\infty} \exp\left(-t \sum_k e^{-(V_{od}-V_{ok})} \right) dt \tag{A.9}$$

$$= \left. \frac{\exp\left(-t \sum_k e^{-(V_{od}-V_{ok})}\right)}{-\sum_k e^{-(V_{od}-V_{ok})}} \right|_0^{\infty} \tag{A.10}$$

$$= \frac{1}{\sum_k e^{-(V_{od}-V_{ok})}} = \frac{e^{V_{od}}}{\sum_k e^{V_{ok}}}. \tag{A.11}$$

A.2. From Aggregate Migration Flow Equation to a Structural Migration Gravity System

Note that all natives from country o are split up over all n countries including the home country which leads to the accounting identity, $\sum_d M_{od} = N_o$. $\sum_o M_{od} = L_d$ is the number of all migrants coming to d, including natives that stay in d, M_{dd}. This is then the labor force available in country d. Following Anderson (2011), define $W_o \equiv \sum_k w_k/\delta_{ok}$ and note that the world labor supply is $N^w \equiv \sum_o N_o = \sum_d L_d$. So, assuming full employment in the world and using Equation (3.13) we can rewrite L_d as

$$L_d = w_d \sum_o ((1/\delta_{od})/W_o) N_o. \tag{A.12}$$

From this we can infer w_d as

$$w_d = \frac{L_d}{\Omega_d N^w}, \quad \text{with} \tag{A.13}$$

$$\Omega_d = \sum_o \frac{1/\delta_{od}}{W_o} \frac{N_o}{N^w}. \tag{A.14}$$

Using Equation (3.16) we can write W_o as

$$W_o = \sum_k \frac{w_k}{\delta_{ok}} = \sum_k \frac{L_k}{\Omega_k \delta_{ok} N^w}. \tag{A.15}$$

Substituting into M_{od} (Equation (2.3)) we can write

$$M_{od} = \frac{w_d/\delta_{od}}{\sum_k w_k/\delta_{ok}} N_o = \frac{L_d N_o 1/\delta_{od}}{N^w \Omega_d W_o}. \tag{A.16}$$

L_d is exogenous in this model by Anderson (2011). In Chapter 3 of this thesis, where wages of this model are determined in a linked trade system, we are able to endogenize L_d. The modularity of the gravity model still allows me here to calculate equilibrium changes of the multilateral resistance terms conditional on L_d.

B. Appendix to Chapter 3

B.1. Derivation of Multinomial-Logit Probabilities

We closely followed the standard logit setting à la McFadden (1974) and Train (2003) to derive these logit probabilities.

We start from Equation (3.9) with density

$$f(\varepsilon_{jih}) = e^{-\varepsilon_{jih}} e^{-e^{-\varepsilon_{jih}}}, \tag{B.1}$$

and cumulative distribution

$$F(\varepsilon_{jih}) = e^{-e^{-\varepsilon_{jih}}}. \tag{B.2}$$

Conditional on ε_{jih} Equation (3.9) is the cumulative distribution for each ε_{jkh} evaluated at $\varepsilon_{jih} + V_{ji} - V_{jk}$ given by equation (B.2)). For independent ε's we can write the cumulative distribution over all other alternatives as the product of the individual cumulative distributions. As usual we reach the unconditional P_{jih} by integrating $P_{jih} \mid \varepsilon_{jih}$ over all possible values of ε_{jih} for the given density B.1:

$$P_{jih} = \int \left(\prod_{k \neq i} e^{-e^{-(\varepsilon_{jih}+V_{ji}-V_{jk})}} \right) e^{-\varepsilon_{jih}} e^{-e^{-\varepsilon_{jih}}} d\varepsilon_{jih}.$$

Starting from this expression we can derive the standard multinomial-logit choice probability expression.

$$\begin{aligned} P_{jih} &= \int_{\varepsilon_{jih}=-\infty}^{\infty} \left(\prod_{k} e^{-e^{-(\varepsilon_{jih}+V_{ji}-V_{jk})}} \right) e^{-\varepsilon_{jih}} d\varepsilon_{jih} \\ &= \int_{\varepsilon_{jih}=-\infty}^{\infty} \exp\left(-\sum_{k} -e^{-(\varepsilon_{jih}+V_{ji}-V_{jk})} \right) e^{-\varepsilon_{jih}} d\varepsilon_{jih} \\ &= \int_{\varepsilon_{jih}=-\infty}^{\infty} \exp\left(-e^{-\varepsilon_{jih}} \sum_{k} -e^{-(V_{ji}-V_{jk})} \right) e^{-\varepsilon_{jih}} d\varepsilon_{jih} \end{aligned} \tag{B.3}$$

If we define $t = e^{(-\varepsilon_{jih})}$ such that $-e^{(-\varepsilon_{jih})} d\varepsilon_{jih} = dt$ and noting that t approaches 0 as ε_{jih} goes to infinity and t is infinite if ε_{jih} approaches negative infinity, we get

$$\begin{aligned} P_{jih} &= \int_{\infty}^{0} \exp\left(-t \sum_k e^{-(V_{ji}-V_{jk})}\right)(-dt) \\ &= \int_{0}^{\infty} \exp\left(-t \sum_k e^{-(V_{ji}-V_{jk})}\right) dt \\ &= \left. \frac{\exp\left(-t \sum_k e^{-(V_{ji}-V_{jk})}\right)}{-\sum_k e^{-(V_{ji}-V_{jk})}} \right|_0^{\infty} \\ &= \frac{1}{\sum_k e^{-(V_{ji}-V_{jk})}} = \frac{e^{V_{ji}}}{\sum_k e^{V_{jk}}}. \end{aligned} \tag{B.4}$$

B.2. Sufficient Statistics for Welfare with Bilateral Migration Based on GDP per Labor Force

We follow Arkolakis et al. (2012) in deriving a sufficient statistics for welfare when allowing for bilateral migration. Considered is a foreign trade shock that leaves the ability to serve the own market, t_{jj}, unchanged as in Arkolakis et al. (2012). Using $Y_j = w_j L_j$, we can write $d\ln Y_j = d\ln w_j + d\ln L_j$. Real consumer expenditure per labor force is given by $W_j \equiv Y_j/(P_j L_j)$ and taking logs, the total differential is given by $d\ln W_j = d\ln Y_j - d\ln P_j - d\ln L_j$.

We first take the total differential of $\ln P_j = \ln\left\{\left[\sum_{i=1}^{n} (\beta_i p_i t_{ij})^{1-\sigma}\right]^{\frac{1}{1-\sigma}}\right\}$:

$$d\ln P_j = \sum_{i=1}^{n} \left(\left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{1-\sigma} d\ln p_i + \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{1-\sigma} d\ln t_{ij}\right).$$

Using $X_{ij} = ((\beta_i p_i t_{ij})/P_j)^{1-\sigma} Y_j$ and defining $\lambda_{ij} = X_{ij}/Y_j = ((\beta_i p_i t_{ij})/P_j)^{1-\sigma}$, yields

$$d\ln P_j = \sum_{i=1}^{n} \lambda_{ij} (d\ln p_i + d\ln t_{ij}). \tag{B.5}$$

Noting that $d\ln p_i = d\ln w_i$ holds, we can also write $d\ln P_j = \sum_{i=1}^{n} \lambda_{ij} (d\ln w_i + d\ln t_{ij})$. Combining terms leads to $d\ln W_j = d\ln Y_j - d\ln P_j = d\ln w_j + d\ln L_j - \sum_{i=1}^{n} \lambda_{ij} (d\ln w_i + d\ln t_{ij})$. Taking the ratio of λ_{ij} and λ_{jj} we can write $\lambda_{ij}/\lambda_{jj} = [(\beta_i p_i t_{ij})/(\beta_j p_j t_{jj})]^{1-\sigma}$. Noting that $dt_{jj} = 0$ by assumption, the log-change of this ratio is given by $d\ln \lambda_{ij} - d\ln \lambda_{jj} = (1-\sigma)(d\ln p_i + d\ln t_{ij} - d\ln p_j)$. Combining this with

Equation (B.5) leads to:

$$d\ln P_j \quad = \quad \frac{1}{1-\sigma}\left(\sum_{i=1}^{n}\lambda_{ij}d\ln\lambda_{ij} - d\ln\lambda_{jj}\sum_{i=1}^{n}\lambda_{ij}\right) + d\ln p_j\sum_{i=1}^{n}\lambda_{ij}.$$

Noting that $Y_j = \sum_{i=1}^{n} X_{ij}$, it follows that $\sum_{i=1}^{n}\lambda_{ij} = 1$ and $d\sum_{i=1}^{n}\lambda_{ij} = \sum_{i=1}^{n} d\lambda_{ij} = 0$. Hence, $\sum_{i=1}^{n}\lambda_{ij}d\ln\lambda_{ij} = \sum_{i=1}^{n} d\lambda_{ij} = 0$. Using these facts, the above expression simplifies to $d\ln P_j = -\frac{1}{1-\sigma}d\ln\lambda_{jj} + d\ln p_j = -\frac{1}{1-\sigma}d\ln\lambda_{jj} + d\ln w_j$, using again $d\ln p_j = d\ln w_j$. Using these expression in $d\ln W_j$ leads to $d\ln W_j = d\ln w_j + d\ln L_j + \frac{1}{1-\sigma}d\ln\lambda_{jj} - d\ln w_j - d\ln L_j = \frac{1}{1-\sigma}d\ln\lambda_{jj}$. Integrating between the initial and the counterfactual situation we get $\ln\hat{W}_j = \frac{1}{1-\sigma}\ln\hat{\lambda}_{jj}$. Taking exponents leads to

$$\hat{W}_j = \hat{\lambda}_{jj}^{\frac{1}{1-\sigma}}. \tag{B.6}$$

This expression is identical to the sufficient statistics of Arkolakis et al. (2012). Hence, allowing for bilateral migration does not change the sufficient statistic for welfare when focusing on real expenditure per labor force.

C. Appendix to Chapter 4

C.1. Proof of Lemma 3

Proof: We first derive how the omitted variable ω_{ij} is correlated with distance d_{ij}. This can easily be seen by inserting (4.2) into (4.3) and taking the expected value conditional on distance d_{ij} and the other control variables $z_0 \equiv \gamma_0 + \zeta_i + \xi_j - \kappa\phi_{ij}$ to obtain

$$E[\omega_{ij} \, | d_{ij}, z_0] = \int f_{\omega_{ij}} (\omega_{ij}) \ln [exp [\delta (z_0 - \gamma d_{ij} + \eta_{ij})] - 1] \, d\omega_{ij} \equiv \Omega (z_0, d_{ij}) \, , \qquad \text{(C.1)}$$

where $f_{\omega_{ij}} (\omega_{ij})$ is the marginal distribution function of ω_{ij}, and we take ω_{ij} to be conditionally independent of d_{ij} and z_0, i.e. we investigate the omitted variable bias after having properly controlled for the selection bias (such as by the Heckman correction factor) or, equivalently, considering the case where no trade flows are missing.[1] This has the purpose of comparing the Heckman estimator, which controls the selection bias but suffers from the omitted variable bias with the HMR estimator which controls for both biases. Controlling conceptually for the selection bias while analyzing the omitted variable bias implies that $e^{\delta(z_o - \gamma d_{ij} + \eta_{ij})} > 1$ to ensure that there are no missing observations causing selection bias. Moreover, $\Omega (z_o, d_{ij})$ is the non-linear conditional expectation function, the shape of which is easy to analyze. Taking the derivative of (C.1) with respect to distance d_{ij}, we obtain

$$\frac{\partial E[\omega_{ij} \, | d_{ij}, z_0]}{\partial d_{ij}} = -\gamma\delta \int f_{\omega_{ij}} (\omega_{ij}) \frac{e^{\delta(z_0 - \gamma d_{ij} + \omega_{ij})}}{e^{\delta(z_0 - \gamma d_{ij} + \eta_{ij})} - 1} d\omega_{ij} < 0. \qquad \text{(C.2)}$$

Hence, there is a negative correlation between ω_{ij} and d_{ij}, because the share of exporting firms becomes smaller the larger is distance.

[1] To see how this equation is obtained, note that by definition of a conditional expected value $E[\omega_{ij} \, | d_{ij}, z_0] = \int f (\omega_{ij} \, | d_{ij}, z_0) \, \omega_{ij} d\omega_{ij}$, where $f (\omega_{ij} \, | d_{ij}, z_0)$ is the conditional distribution of ω_{ij} (see for example Greene (2012), (B-51)). If we then assume that ω_{ij} is conditionally independent of d_{ij} and z_0, we obtain from (B-60) $f (\omega_{ij} \, | d_{ij}, z_0) = f_{\omega_{ij}} (\omega_{ij})$, where $f_{\omega_{ij}} (\omega_{ij})$ is the marginal probability density (see B-45). Inserting this relation above, we obtain: $E[\omega_{ij} \, | d_{ij}, z_0] = \int f_{\omega_{ij}} (\omega_{ij}) \, \omega_{ij} d\omega_{ij}$. Inserting (4.2) and (4.3) into this relation yields (C.1).

■

C.2. Proof of Proposition 1

Proof: Rewrite (4.1) as $m_{ij} = \beta'\mathbf{X}_{ij} + \omega_{ij} + u_{ij}$ and z^*_{ij} in (4.4) as $z^*_{ij} = \varphi^{*\prime}\mathbf{X}_{ij} + \eta^*_{ij}$ where $\beta \equiv (\beta_0, \lambda_j, \chi_i, -\gamma)'$, and $\varphi^* = \left(\gamma^*_0, \xi^*_j, \zeta^*_i, -\gamma^*\right)', -\kappa^*$. Let, $\hat{\beta}^{OLS}$ denote the OLS estimator of β ignoring the sample selection and omitted variable corrections. We then obtain:

$$\mathrm{E}\left(\hat{\beta}^{OLS}\right) = \beta + \left[\mathbf{X}_{ij}\mathbf{X}'_{ij}\right]^{-1}\mathbf{X}_{ij}\mathrm{E}\left[\omega_{ij} + u_{ij}\left|z^*_{ij} > 0\right.\right], \tag{C.3}$$

where we have exploited that the $\mathbf{X}_{ij}$ variables contain only geography information and are therefore deterministic. To evaluate (C.3), examine the conditional expectations $\mathrm{E}\left[\omega_{ij}\left|z^*_{ij} > 0\right.\right]$ and $\mathrm{E}\left[u_{ij}\left|z^*_{ij} > 0\right.\right]$. Using formula (16.36) on p. 549 in Cameron and Trivedi (2005), we first obtain:

$$\begin{aligned} & E\left[u_{ij}\left|z^*_{ij} > 0\right.\right] = cov\left(u_{ij}, \eta^*_{ij}\right)\mathrm{E}\left[\eta^*_{ij}|\eta^*_{ij} > \varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right] & \text{(C.4)} \\ = \; & corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\frac{\sigma_u}{\sigma_\eta}\frac{\phi\left(\varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)}{\Phi\left(\varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right)} \\ \equiv \; & \beta_{u\eta}\bar{\eta}_{ij} > 0, \end{aligned}$$

where $\beta_{u\eta} = corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\sigma_u/\sigma_\eta$ and $\bar{\eta}_{ij} = \frac{\phi(\varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij})}{\Phi(\varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij})}$. Further, we have assumed that u_{ij} and η^*_{ij} are bivariate normally distributed. Note that this implies that $u_{ij} = cov(u_{ij}, \eta^*_{ij})\ \eta^*_{ij}/\sigma^2_\eta + \varrho_{ij}$, where ϱ_{ij} is independent of η^*_{ij} and has zero mean. Hence, $\mathrm{E}\left[u_{ij}\left|\eta^*_{ij} > \varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right.\right] = cov(u_{ij}, \eta^*_{ij})\ 1/\sigma^2_\eta\mathrm{E}\left[\eta^*_{ij}\left|\eta^*_{ij} > \varphi^{*\prime}\mathbf{X}_{ij} - \kappa^*\phi_{ij}\right.\right]$ and $cov\left(u_{ij}, \eta^*_{ij}\right) = corr\left(u_{ij}, u_{ij} + \nu_{ij}\right)\sigma_u\sigma_\eta$. To proceed, use a linear approximation of $\omega_{ij} = \ln\left[\left(Z^*_{ij}\right)^\delta - 1\right]$ for $z^*_{ij} > 0$. We can then write $\omega_{ij} = \ln\left[\left(Z^*_{ij}\right)^\delta - 1\right] = \ln\left[\exp\left(\delta z^*_{ijl}\right) - 1\right] \approx \delta z^*_{ij} > 0$, where $\delta = \sigma_\eta\frac{k-\varepsilon+1}{\varepsilon-1}$ is defined as above.[2] We then obtain:

$$\begin{aligned} & \mathrm{E}\left[\omega_{ij}\left|z^*_{ij} > 0\right.\right], & \text{(C.5)} \\ = \; & \mathrm{E}\left[\delta z^*_{ij}\left|z^*_{ij} > 0\right.\right] = \delta\mathrm{E}\left[\left\{\mathrm{E}\left[z^*_{ij}\left|\mathbf{X}_{ij}\right.\right] + \eta^*_{ij}\right\}\left|z^*_{ij} > 0\right.\right] \\ = \; & \delta\mathrm{E}\left[z^*_{ij}\left|\mathbf{X}_{ij}\right.\right] + \delta\mathrm{E}\left[\eta^*_{ij}\left|z^*_{ij} > 0\right.\right], \\ = \; & \delta\mathrm{E}\left[\gamma^*_0 + \xi^*_j + \zeta^*_i - \gamma^* d_{ij} - \kappa^*\phi_{ij}\left|\mathbf{X}_{ij}\right.\right] + \delta\mathrm{E}\left[\eta^*_{ij}\left|z^*_{ij} > 0\right.\right], \\ = \; & \delta\left[\gamma^*_0 + \xi^*_j + \zeta^*_i - \gamma d^*_{ij} - \kappa^*\phi_{ij} + \bar{\eta}^*_{ij}\right], \\ = \; & \delta\varphi^{*\prime}\mathbf{X}_{ij} + \delta\bar{\eta}^*_{ij}. \end{aligned}$$

[2]It can be shown that this approximation works very well in the range of ω_{ij} from $[0.5, \infty]$ and estimated values of δ around 1.

Noting that $[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{X}\varphi^*\delta = \varphi^*\delta$, we obtain:

$$\mathrm{E}\left(\hat{\beta}^{OLS}\right) = \beta + \varphi^*\delta + \left[\mathbf{X}_{ij}\mathbf{X}_{ij}'\right]^{-1}\mathbf{X}_{ij}\delta\bar{\eta}_{ij}^* + \left[\mathbf{X}_{ij}\mathbf{X}_{ij}'\right]^{-1}\mathbf{X}_{ij}\beta_{u\eta}\bar{\eta}_{ij}^* \gtreqless 0. \tag{C.6}$$

Since country dummies in $\mathbf{X}_{ij}$ are not correlated by construction and distance is hardly correlated with country dummies the matrix $\mathbf{X}'\mathbf{X}$ can be viewed as diagonal. But then:

$$\mathrm{E}\left(-\hat{\gamma}^{OLS}\right) = -\gamma - \gamma\delta + \frac{\sum_i \sum_j d_{ij}}{\sum_i \sum_j (d_{ij})^2}\left[\delta + \beta_{u\eta}\right]\bar{\eta}_{ij}^*, \tag{C.7}$$

and hence

$$\mathrm{Bias}(\hat{\gamma}^{OLS}) = \gamma\delta - \frac{\sum_i \sum_j d_{ij}}{\sum_i \sum_j (d_{ij})^2}\left[\delta + \beta_{u\eta}\right]\bar{\eta}_{ij}^*.$$

■

C.3. Proof of Proposition 2

Proof: From (4.8), we have $\mathrm{Bias}(\hat{\gamma}^{OLS}) = \gamma\delta - \Xi\left[\delta + \beta_{u\eta}\right]\bar{\eta}_{ij}^*$. Thus, it follows that $\frac{\partial \mathrm{Bias}(\hat{\gamma}^{OLS})}{\partial t} = \delta\frac{\partial\gamma}{\partial t} - \Xi\left[\delta + \beta_{u\eta}\right]\frac{\partial\bar{\eta}_{ij}^*}{\partial t}$. The change of the omitted variable bias over time is simply given by:

$$\frac{\partial(\delta\gamma)}{\partial t} = \delta\frac{\partial\gamma}{\partial t} < 0.$$

The sign of the change of the sample selection bias depends on the sign of

$$\begin{aligned}
\frac{\partial\bar{\eta}_{ij}^*}{\partial t} &= \frac{\partial\left(\frac{\phi(z_{ij}^*)}{\Phi(z_{ij}^*)}\right)}{\partial t} \\
&= \frac{1}{\Phi(z_{ij}^*)^2}\left[\left(\phi'\left(z_{ij}^*\right)\cdot\Phi\left(z_{ij}^*\right) - \phi\left(z_{ij}^*\right)^2\right)\right]\frac{\partial z_{ij}^*}{\partial t} \\
&= \left[\frac{-z_{ij}^*\phi\left(z_{ij}^*\right)}{\Phi\left(z_{ij}^*\right)} - \left(\frac{\phi\left(z_{ij}^*\right)}{\Phi\left(z_{ij}^*\right)}\right)^2\right]\frac{\partial z_{ij}^*}{\partial t} \\
&= \left[-z_{ij}^*\bar{\eta}_{ij}^* - \left(\bar{\eta}_{ij}^*\right)^2\right]\frac{\partial z_{ij}^*}{\partial t} \\
&= -\bar{\eta}_{ij}^*\left[z_{ij}^* + \bar{\eta}_{ij}^*\right]\frac{\partial z_{ij}^*}{\partial t}.
\end{aligned} \tag{C.8}$$

Note that

$$\frac{\partial z_{ij}^*}{\partial t} = -d_{ij}\frac{\partial\gamma(t)}{\partial t} > 0.$$

The derivative of the mills ratio $\frac{\partial \bar{\eta}_{ij}^*}{\partial z_{ij}^*} = -\bar{\eta}_{ij}^* \left[z_{ij}^* + \bar{\eta}_{ij}^* \right]$ is negative. This can be shown by noting that

$$E\left[\eta_{ij}^* \left| \eta_{ij}^* > -\varphi'\mathbf{X} \right.\right] = \frac{\phi\left(\varphi'\mathbf{X}\right)}{\Phi\left(\varphi'\mathbf{X}\right)} = \frac{\phi\left(-\varphi'\mathbf{X}\right)}{1 - \Phi\left(-\varphi'\mathbf{X}\right)}, \tag{C.9}$$

and using the result derived in Sampford (1953) and also given in Theorem 19.2 on page 876 in Greene (2012), that for $\phi(x) / (1 - \Phi(x))$ the derivative with respect to x is given by

$$\frac{\phi(x)}{1 - \Phi(x)} \left[\frac{\phi(x)}{1 - \Phi(x)} - x \right], \tag{C.10}$$

and bounded between zero and one. Using the equality given in Equation ((C.9)), we may write this as:

$$\frac{\phi\left(\varphi'\mathbf{X}\right)}{\Phi\left(\varphi'\mathbf{X}\right)} \left[\frac{\phi\left(\varphi'\mathbf{X}\right)}{\Phi\left(\varphi'\mathbf{X}\right)} + \varphi'\mathbf{X} \right] = \bar{\eta}_{ij} \left[z_{ij} + \bar{\eta}_{ij} \right]. \tag{C.11}$$

Hence, this expression differs from our derivative of $\bar{\eta}_{ij}^*$ only by the multiplication with -1. Hence, the derivative of $\bar{\eta}_{ij}^*$ with respect to z_{ij}^* is bounded between -1 and 0. But then $\frac{\partial \bar{\eta}_{ij}^*}{\partial t} = \partial \left[\phi\left(z_{ij}^*\right) / \Phi\left(z_{ij}^*\right) \right] / \partial t < 0$. The change in the bias for OLS is therefore ambiguous, depending on whether the change in the sample selection bias or the change in the omitted variable bias is larger:

$$\frac{\partial \text{Bias}(\hat{\gamma}^{OLS})}{\partial t} = \delta \frac{\partial \gamma}{\partial t} - \Xi \left[\delta + \beta_{u\eta} \right] \frac{\partial \bar{\eta}_{ij}^*}{\partial t} \gtreqless 0. \tag{C.12}$$

■

D. Appendix to Chapter 5

D.1. Synopsis of Studies Employing Micro-Level Measures of FDI or Offshoring

Table D.1.: Synopsis of Studies Employing Micro-Level Measures of FDI or Offshoring

Author(s)	**Main explanatory variable**	**Employed method**	**Main results**	**Number of treated observations**	**Country**
Barba Navaretti and Castellani (2004)	first time investment abroad	PSM	no significant employment effects	119 switching firms from national from 1995-1997	Italy
Barba Navaretti et al. (2010)	firms that switch from having zero to one foreign affiliate (first time investment)	PSM	no negative effect on employment at all, partly positive	171 switchers in France from 1993 to 2000, 269 in Italy from 1993-2001	France & Italy
Becker et al. (2005)	foreign affiliate wages	translog cost function regression	negative employment effect from decreased wages in affiliates	451 parent firms in 2000 for Germany and 92 parent firms in 1998 for Sweden	Germany & Sweden
Becker and Muendler (2008)	employment expansion in foreign affiliates	PSM	decreased probability of domestic worker separation	39,681 plants where firms expanded employment in foreign affiliates in year 2000	Germany
Biscourp and Kramarz (2007)	firm-level goods trade dummy and relative to sales, separable by finished goods and intermediates	descriptive statistics	evidence for negative employment effects of sourcing strategies	universe of manufacturing firms engaged in international trade from 1986 to 1992 (61540)	France
Buch and Lipponer (2010)	MNE status dummies	GMM	no higher employment volatility in MNEs versus domestic firms	500 MNEs	Germany

continued on next page

Table D.2.: Synopsis of Studies Employing Micro-Level Measures of FDI or Offshoring (Continued)

Author(s)	**Main explanatory variable**	**Employed method**	**Main results**	**Number of treated observations**	**Country**
Cuyvers et al. (2010)	new investment abroad	translog cost function regression and dynamic versions via GMM	mixed effects for blue- and white-collar workers	521 firms which are engaged in FDI from 1997-2007	Belgium
Debaere et al. (2010)	new investments abroad	PSM	slightly negative effects for investment in least developed countries, no effect to more-advanced countries	462 new MNEs from 1981-1995	South Korea
Desai et al. (2009)	foreign affiliate employment growth	panel regressions, panel IV	positive domestic employment growth	2834 manufacturing firms in 1982, 1989, 1994, 1999, and 2004	USA
Dhyne and Guerin (2014)	first time investment abroad	PSM	no employment effects	4699 firms from 1998 to 2008	Belgium
Engel and Procher (2013)	exporters investment abroad	PSM	positive employment effects	884 investments in 2001, 2003 and 2006	France

continued on next page

Table D.3.: Synopsis of Studies Employing Micro-Level Measures of FDI or Offshoring (Continued)

Author(s)	**Main explanatory variable**	**Employed method**	**Main results**	**Number of treated observations**	**Country**
Harrison and Mcmillan (2011)	changes in foreign affiliate wages for horizontally and vertically integrated firms and different locations	panel regressions	mixed results slightly negative impact from horizontal FDI, slightly positive for vertical FDI	2088 vertically and 1778 horizontally integrated firms from 1982-1999	USA
Hering et al. (2010)	first time investment abroad	PSM	limited effects (either positive or negative)	150 switching firms from 1995-2003	Japan
Hijzen et al. (2011)	firms that switch from having zero to one foreign affiliate (first time investment)	PSM	non-negative employment effects, partly positive	309 switchers in manufacturing and 185 in services from 1987 to 1999	France
Imbriani et al. (2011)	first time investment abroad	PSM	slightly negative for second year service sector MNEs in one specification	2002-2007	Italy
Mattes (2010)	imputed FDI dummies by region and motivation	GMM	no negative effect of FDI on firm-level employment	artificial	Germany
Moser et al. (2015)	qualitative increase in intermediate input purchases from abroad	PSM	positive in general, but negative for restructuring subsample (no causal link)	1084 manufacturing plants in 1998, 2000 and 2002	Germany

continued on next page

Table D.4.: Synopsis of Studies Employing Micro-Level Measures of FDI or Offshoring (Continued)

Author(s)	**Main explanatory variable**	**Employed method**	**Main results**	**Number of treated observations**	**Country**
Kleinert and Toubal (2008)	first time investment abroad	PSM	positive and non-negative effect on employment growth	936 switching firms from 1997-2003	Germany
Kramarz (2008)	increased imports of finished goods	panel regressions	partly negative impact on employment	112686 firm-person-year observations from 1986 to 1992	France
Liu and Nunnenkamp (2011)	new investments abroad	ordered probit	mixed results: also small negative self assessed effect for increasing volume of investment	1770 MNEs in 2006	Taiwan
Temouri and Driffield (2009)	average subsidiary wages	GMM	no significant effects	2129 MNEs and their subsidiaries from 1997-2008	Germany
Wagner (2011)	first time relocation	PSM	non-negative/slightly negative employment effects	160 firms from 2001-2003	Germany

Eidesstattliche Versicherung

(§8 S. 1 Nr. 5 PromO):
Hiermit erkläre ich, dass gegen mich wegen eines Verbrechens oder eines vorsätzlichen Vergehens weder ein Ermittlungs- noch ein Strafverfahren anhängig ist; ich bin wegen einer solchen Tat weder rechtskräftig verurteilt noch verbüÃŸe ich eine Freiheitsstrafe.

(§8 S. 1 Nr. 6 PromO):
Hiermit erkläre ich eidesstattlich, dass ich die Dissertation selbstständig verfasst und keine anderen als die von mir angegeben Quellen und Hilfsmittel benutzt habe. Ich habe die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht und habe auch nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden.

(§8 S. 1 Nr. 8 PromO):
Hiermit erkläre ich eidesstattlich, dass ich keine gewerbliche Promotionsvermittlung und -beratung in Anspruch genommen habe und auch nicht nehmen werde.

(§8 S. 1 Nr. 9 PromO):
Hiermit erkläre ich mich damit einverstanden, dass die elektronische Fassung meiner Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.

(§8 S. 1 Nr. 10 PromO):
Hiermit erkläre ich mich damit einverstanden, dass bei Verdacht wissenschaftlichen SFehlverhaltens Untersuchungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

Bayreuth, den 2.02.2016 Steffen Sirries