Essays in Economic History:

Demographic, Political, and Methodological Changes

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

der Rechts- und Wirtschaftswissenschaftlichen Fakultät

der Universität Bayreuth

Vorgelegt

von

Volkan Timur Öztürk

aus

Karabük

Dekan/-in:	Prof. Dr. Claas Christian Germelmann
Erstberichterstatter/-in:	Prof. Dr. Sebastian T. Braun
Zweitberichterstatter/-in:	Prof. Dr. Mario Larch
Tag der mündlichen Prüfung:	18.09.2026

Mustafa Kemal'e

Acknowledgments

The number of words I will write on this page will not be enough. Nevertheless, I would like to thank Prof. Dr. Sebastian Braun. It is difficult to articulate the relationship between us but his warm presence, wisdom, and much needed passion for Baden-Württemberg made this thesis (and my intellectual growth) possible. He somehow knew how to preserve my childlike curiosity, endless ideas, and sanity, while guiding me through this chapter of my life. It may not have seemed like it, but during the toughest times I looked up to him, and will continue to do so.

There have been many mentors who guided this intellectual journey, but some were central to it. I would like to thank Prof. Dr. Jan-Otmar Hesse, Prof. Dr. Mario Larch and Prof. Dr. Hartmut Egger for their support, time and unlimited encouragement. I also thank my co-author Dr. Richard Franke, and my colleague Dr. Bang Nguyen.

I have called Bayreuth home for eight years now, and have spent almost all of my twenties here. I have met so many people, and I am very lucky to call many of them good friends. Listing them here would make this thesis twice the size it is now, not because I am very popular, but because each of them holds a special place in my heart that I could write a book about. I still would like to thank some of them. Philipp has become one of my oldest friends, and without our friendship Bayreuth would have been a hollow place. The 11:30 group, as it is unofficially named: Jonathan, Alessandro, Maximilian, Tim Röthel, Tim Baule, Tim Lüdiger, Michael and Joshua made the daily struggles of being a PhD student trivial. The 12:30 group, with many of its members, but especially Yasemin and Leandro, energized me every time we exchanged ideas. A special thanks to Henrik and Christoph. Maybe I saw myself in them, or wanted to, and their presence made me a better person; I learned so much from them.

Finally, I would like to thank my parents and my brother. They let me go to another country, knowing that we would not be able to see each other as much as we wanted, never at the times we wanted, and confined to seeing each other through a small screen for months on end. Perhaps the most bittersweet part of it all is that I missed so many years of bonding with my brother, only to observe how different we have become. This dissertation is dedicated to him, since I gain so many fresh perspectives and ideas just by talking to him.

Summary

This thesis studies how persistent social and institutional structures shape long-run historical change, and how new advances in artificial intelligence methods can help build datasets from historical sources to examine such change. It focuses on the confessional divide in southwestern Germany, especially the role of Protestant and Catholic communities in demographic and political transformation. Its methodological contribution is a reproducible workflow for creating structured historical data at scale.

The first chapter examines the fertility decline in Württemberg between 1871 and 1933. Using novel parish-level data, it provides evidence that Protestant communities entered the fertility transition earlier than Catholic parishes. The results suggest that this difference was not primarily driven by economic modernization, but was instead linked to religious norms, secularization, and changing attitudes toward deliberate fertility control.

The second chapter studies how confessional cleavages shaped political competition from the Weimar Republic to postwar West Germany. It shows that Catholic areas were strongly anchored by the Catholic party *Zentrum*, while Protestant areas were more politically fragmented and later more exposed to Nazi electoral consolidation. After 1945, the CDU weakened the older confessional divide, although Catholic areas remained less prone to far-right voting.

The third chapter turns to the production of historical data itself. It develops a reproducible LLM-based workflow for extracting structured information from historical biographical sources. Applied to *Degener's Wer Ist's?* from 1911, the chapter shows how artificial intelligence supports the creation of large-scale historical datasets while preserving transparency and reproducibility.

Contents

List of Figures

List of Tables

Introduction

The questions that matter most in economics often unfold too slowly to be understood from the present alone: Why some societies grow rich, why institutions endure or fail, and how demographic change begins and reshapes societies. Even when the mechanisms behind such processes are partly understood, their long-run consequences are difficult to observe while they are still in motion. This is one reason why economists repeatedly turn to history when contemporary questions demand longer horizons. Recent recognition within economics reflects this turn toward historical evidence. In 2023, Claudia Goldin was awarded the economics Nobel for work reconstructing the long-run evolution of women's labor-market outcomes (Goldin, 1990); in 2024, Acemoglu, Johnson, and Robinson were recognized for research linking institutions and prosperity across centuries (Acemoglu et al., 2001); and in 2025, Joel Mokyr shared the prize for identifying the historical conditions under which technological progress became self-sustaining (Mokyr, 2016). Together, these contributions underline the value of history which lies not only in looking further back, but in showing when change began, how it unfolded, and why its effects persisted.

Many of the transformations unfolding today are of exactly this kind. Fertility is falling and populations are aging across much of the developed world (United Nations Department of Economic and Social Affairs, Population Division, 2024). At the same time, many democracies face rising polarization, institutional strain, and new forms of party-system fragmentation and radicalization (Guriev and Papaioannou, 2022; Nord et al., 2026). These developments are often studied separately, but both raise a similar question: How do long-standing social divisions shape the way societies adapt to major economic, cultural, and institutional change? Fertility decline depends not only on income and technology, but also on norms about family, gender, and religion (Doepke et al., 2023). Political realignment depends not only on parties and institutions, but also on the social identities through which voters interpret change (Lipset and Rokkan, 1967). This dissertation turns to earlier episodes of both. Its first two chapters use the confessional geography of southwestern Germany to answer how a single social cleavage, religious affiliation, shaped the fertility decline and the reordering of political competition and fragmentation.

The first chapter, co-authored with Sebastian Braun and Richard Franke, studies the role of religious affiliation in the fertility decline, one of the central demographic transformations of industrializing Europe. Using newly assembled parish-level data for 1,764 localities in Württemberg between 1871 and 1933, it exploits the region's historically determined Catholic-Protestant mosaic to compare nearby parishes that differed sharply in confession but were otherwise similar in geography and economic conditions. The chapter shows that Protestant parishes experienced the onset of fertility decline substantially earlier than Catholic parishes. Since these differences are not explained by local economic modernization or gender wage gaps, the results point to religious norms, secularization, and responses to falling infant mortality as important forces behind the timing of demographic change.

The second chapter, co-authored with Sebastian Braun, examines how the same confessional cleavage shaped the political landscape from the Weimar Republic to postwar West Germany. Using novel, parish-level election data for more than 3,000 localities in Baden and Württemberg between 1928 and 1969, it studies how religious composition affected party-system fragmentation and far-right voting. The argument centers on an asymmetry in Weimar's center-right: Catholic voters were organized around the Zentrum, while Protestant voters lacked an equivalent political anchor and were dispersed across liberal, conservative, and nationalist parties. As a result, fragmentation peaked where a cohesive Catholic bloc coexisted with a fragmented Protestant electorate. During the 1930s, the NSDAP consolidated much of this previously fragmented Protestant vote, while Catholic areas remained anchored by the Zentrum.

After 1945, the CDU as a cross-confessional party changed this relationship. By integrating Catholic and Protestant voters into one political unit, the CDU weakened the link between religious composition and party-system fragmentation. Yet the confessional divide did not disappear entirely: Catholic areas remained less prone to far-right voting when the NPD emerged in the late 1960s. The chapter combines parish-level election results with individual-level survey evidence from the 1968 Baden-Württemberg state election to examine this persistence. The evidence suggests that church attachment and political cues from religious authorities, rather than ideological differences, explain this persistence.

While the first two chapters use fine-grained historical data to study demographic and political change, the third chapter turns to the production of such data itself. Much of economic history depends on archival and printed sources that are rich in information but costly to transform into structured data. Artificial intelligence is changing how researchers can process large bodies of evidence from historical sources (Brynjolfsson et al., 2025; Messeri and Crockett, 2024), but these tools raise new problems of transparency, reproducibility, and verification. The chapter addresses this problem by developing an agentic LLM workflow that separates text recognition, variable identification, extraction, validation, and normalization into auditable steps. Applied to Herrmann A. L. Degener's *Wer Ist's?* (1911), a dense German biographical directory, the workflow extracts over 16 thousand biographies into a structured dataset containing demographic information, education, career histories, publications, memberships, occupations, addresses, and related variables. The contribution is therefore not only a new dataset, but a transparent and reproducible workflow for building historical datasets.

Taken together, the dissertation shows how fine-grained historical evidence can clarify both substantive and methodological questions in economics. The first two chapters use the confessional geography of southwestern Germany to show that religious cleavages shaped demographic and political change through social norms, party organization, church attachment, and institutional context. The third chapter turns to the production of such evidence itself, showing how AI-assisted workflows can make complex historical sources more accessible while preserving transparency, reproducibility, and human oversight.

The dissertation proceeds as follows. Chapters 1 to 3 present the three papers. The conclusion summarizes the main findings. The bibliography lists the references cited across all chapters, and the appendices provide supplementary material for the papers.

Chapter 1

Protestantism, Industrialization, and the Fertility Decline: Evidence from Religious Enclaves in Württemberg, 1871-1933[1]

1.1 Introduction

One of the most critical events in human history is the demographic transition and the escape from the Malthusian trap. The European Core experienced this transition from high to low birth and death rates during the 19th century, which coincided with the first industrialization. Notably, these transition processes occurred unevenly across Europe. There is still considerable debate about the drivers of this demographic transition and their relative importance (e.g., Galor, 2012; Chapman,

[1]This paper is joint work with Sebastian Braun and Richard Franke. All authors contributed equally. Sarah Schaller, Charlotte Rehling, and Bastian Prinz provided excellent research assistance. We also thank seminar and conference participants at EHES 2023, Bayreuth Graduate Seminar (2022, 2023), CEPH 2024, and EHS 2024. Funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–project number 471335227 is gratefully acknowledged. All remaining errors are my own.

2019; Doepke et al., 2023). A critical challenge in the literature is that proposed drivers of the fertility decline, such as the increase in women's relative wages during industrialization (Galor and Weil, 1996), are often endogenous to fertility decisions (Cinnirella, 2019). Similar endogeneity problems arise in empirical analyses of the mortality decline (Alsan and Goldin, 2019). On top of that, due to the scarcity of granular historical data, the literature relied on relatively highly aggregated data, which masks a substantial heterogeneity within regions. Building upon this literature, this paper provides new evidence on the timing and spread of the demographic transition at the micro-regional level within Württemberg between 1871 and 1933.

We study the descriptive trends of the demographic transition and the impact of religion on fertility for the universe of localities in Württemberg. We then examine whether the earlier Protestant fertility decline can be explained by economic modernization, a central mechanism in the literature, or by other demographic and institutional channels. The evidence does not support these alternatives: localized comparisons show no consistent Protestant advantage in industrialization, non-agricultural employment, firm size, steam power, or the gender wage gap, and we find no evidence that marriage rates, out-of-wedlock births, infant mortality, border exposure, monastery proximity, and inheritance regimes account for the timing gap. This leaves religious denomination as the most plausible driver: It shaped the acceptability and timing of deliberate fertility control. This interpretation is consistent with contemporary observers, who linked fertility decline to secularization and the weakening of religious constraints on sexual and family behavior.

First, we contribute to the discussion of dating the demographic transition (Coale and Watkins, 1986; Guinnane et al., 1994; Brown and Guinnane, 2007). In particular, we focus on the fertility decline which we measure as a permanent drop of the 5-year moving average crude birth rate by at least 10% below its pre-transition level. Parishes in Württemberg experienced the fertility decline at the beginning of the 20th century. However, heterogeneity of the timing across the 1,764 parishes in

Württemberg matches the heterogeneity in the timing of the fertility decline across European regions, a fact that would be masked in more aggregated data.

Second, we investigate the impact of religion on the fertility decline. In 1871, the population of Württemberg consisted largely of Protestants (69%) and Catholics (30%). Most parishes had a clear Catholic or Protestant majority in 1871. Only in 152 of the 1,764 parishes were less than 90% of the population members of the majority denomination. This clear divide was predetermined centuries ago by territorial fragmentation and the "cuius regio, eius religio" principle in the Peace of Augsburg 1555. At that time, local rulers decided about their religious affiliation and their subjects had to follow or emigrate.

For a causal interpretation of the effect of religion on fertility, it must hold that, conditional on control variables, Protestant and Catholic parishes do not differ in unobserved predetermined characteristics affecting fertility. At the time of the fertility decline all parishes were under the same administrative rule and did not experience discriminatory rule changes. A potential concern is that predominantly Protestant and Catholic areas differ in their geographic location, potentially affecting access to markets and knowledge. To address these concerns, we add suitable control variables, and restrict the analysis to localized comparisons, i.e. using only within-county variation, demeaning by a local average, and comparing religious enclaves with their neighbors. Neighboring parishes should share similar pre-Reformation characteristics. Indeed, balancing tests show that these localized comparisons address observable differences between Protestant and Catholic parishes.

Regression results indicate that Protestant parishes experienced the onset of the fertility transition between 6.3 and 10.3 years (or between 0.55 and 0.88 standard deviations) earlier than Catholic parishes. Using the same localized settings, we find no large scale differences in economic modernization and the gender wage gap between Protestant and Catholic parishes. Economic factors, however, have been widely discussed as a trigger for the fertility decline (Guinnane, 2011). In the absence of strong economic differences, our discussion focuses on social norms associated with religion as drivers for the fertility decline. Contemporaneous observers like Offenbacher (1900) and Wolf (1912a) already discussed this potential mechanism and provide qualitative evidence.

The remainder of the paper is organized as follows. Section 1.2 reviews the related literature on the demographic transition, fertility decline, and religion. Section 1.3 describes the historical background of Württemberg, including its denominational geography, demographic transition, and industrialization. Section 1.4 introduces the parish-level data and the construction of our fertility transition measures. Section 1.5 outlines the empirical strategy, and the one after that presents the main results on Protestantism and the timing of the fertility decline. Section 1.7 explores whether these differences are better explained by economic modernization or by secularization and changing social norms. Section 1.8 concludes.

1.2 Related Literature

The demographic transition is the shift from high to low birth and death rates, and it is among the most studied phenomena in economics and economic history (Guinnane, 2011; Doepke et al., 2023). Economic theories of the fertility decline emphasize the changing costs and benefits of children. In the quantity-quality framework of Becker (1960), rising returns to human capital lead parents to substitute fewer, better-educated children for larger families. This logic lies at the heart of unified growth theory (Galor and Weil, 1996, 2000; Galor, 2011). Related accounts stress the rising opportunity cost of women's time as female wages increased during industrialization (Galor and Weil, 1996), as well as the decline in child mortality, which reduced the number of births needed to reach a target number of surviving children (Doepke et al., 2023). A recurring empirical difficulty is that these drivers are themselves outcomes of the development process. Women's relative wages, child mortality, and the returns to schooling are all endogenous to fertility (Cinnirella, 2019; Alsan and Goldin, 2019). Micro-level evidence from before the transition nonetheless supports a tight link between higher education and lower fertility (Becker et al., 2013).

A second tradition locates the origins of the decline less in economic incentives than in cultural change. Analyzing the Princeton European Fertility Project, Coale and Watkins (1986) found that the transition often spread along linguistic and regional lines rather than tracking socioeconomic conditions. This pattern suggested that the diffusion of new attitudes toward family limitation mattered independently of economic demand (Cleland and Wilson, 1987; Lesthaeghe and Wilson, 1986). Recent work has revived and sharpened this view. Spolaore and Wacziarg (2022) show that the decline diffused more readily between culturally and ancestrally related populations, and Beach and Hanlon (2023b) document a direct role for culture in the timing of the transition. The French case is especially instructive. Blanc (2023) argues that secularization, understood as the erosion of religious constraints on behavior, can account for the exceptionally early French fertility decline, drawing on variation across French departments and crowd-sourced genealogies. Cummins (2013), by contrast, attributes the same decline to falling inequality and social mobility after the Revolution.

Religion occupies a central but contested place in this debate, especially for the European continent. The association between Protestantism and lower fertility was already debated by contemporaries such as Offenbacher (1900) and Wolf (1912a), the latter linking the decline to secularization and to social-democratic milieus. Subsequent quantitative work reached mixed conclusions and was hampered by aggregation. Studying 176 Bavarian districts, Brown and Guinnane (2002) find higher fertility in Catholic areas but argue that economic factors, obscured by aggregate data, were decisive. Galloway et al. (1994) find religion to be the single most important predictor of fertility levels in Prussia between 1875 and 1910, but they do not find it to be a significant determinant of the timing of the decline. Becker and Cinnirella (2020), using more than 50,000 Prussian localities in 1871, document lower Protestant fertility that becomes insignificant once literacy is controlled for. This body of work leaves open a basic question. The difference between Protestant and Catholic parishes may reflect denomination itself, or it may reflect correlated factors such as education, economic structure, and geography. Aggregated, cross-sectional data cannot tell these explanations apart.

Whether religion shaped fertility through the economy or through norms connects directly to the long-standing debate over the economic consequences of Protestantism. Weber (1904) famously tied Protestantism to modern economic development. Becker and Woessmann (2009) reinterpret this association as operating through human capital rather than a work ethic, whereas Cantoni (2015) finds no effect of Protestantism on city growth and so casts doubt on a purely economic channel. The implication for fertility is sharp. If Protestantism advanced economic modernization, it could have hastened the fertility decline indirectly. If it did not, then any denominational difference points instead toward norms and the legitimacy of deliberate family limitation (McQuillan, 2004). Distinguishing these channels requires data fine enough to hold economic conditions fixed while comparing otherwise similar Protestant and Catholic communities.

We contribute in three ways. First, we provide the first parish-level evidence on the timing of the fertility decline for Germany, covering the universe of 1,764 localities. We show that the heterogeneity in transition dates within Württemberg is as large as the heterogeneity observed across European regions, which aggregated data conceal. Second, we estimate the effect of religion on the timing of the transition within a common institutional setting. We exploit the predetermined confessional geography created by the Peace of Augsburg together with a set of religious enclaves, and we rely on localized comparisons within counties, within 10 km neighborhoods, and between enclaves and their neighbors, each of which holds geography and pre-Reformation development fixed. This design confronts the confounding that has clouded earlier studies of religion and fertility, and it distinguishes our paper from cross-sectional analyses of fertility levels such as Becker and Cinnirella (2020). Third, using the same design, we adjudicate between economic and normative mechanisms. We find little evidence that economic modernization differed across denominations, and the results point instead toward secularization and changing social norms.

1.3 Historical Background

On the eve of the French Revolution, the German Southwest was composed of hundreds of independent territories, exemplifying the fragmented nature of the German lands during the Holy Roman Empire. With the end of the Old Empire, as result of the Napoleonic Wars, ended also the territorial fragmentation in the German Southwest. The main beneficiaries were the Grand Duchy of Baden and the Kingdom of Württemberg. The latter emerged from the Duchy of Württemberg in 1806 and is the focus of our analysis.

As a close ally of Napoleon, Württemberg acquired several formerly independent territories during the Napoleonic era, particularly to the east and south of the old duchy. As a result, Württemberg doubled its size and population between 1803 and 1810 (see Appendix Figure A.1 for the territorial division of the future Kingdom of Württemberg in 1789). The new state joined the German Confederation (*Deutscher Bund*) formed at the Congress of Vienna in 1815, and later became part of the newly established German Empire in 1871. It was one of the four kingdoms within the Empire and the third largest state after Prussia and Bavaria (see Appendix Figure A.2 for a map of Württemberg within the German Empire).

Protestants and Catholics in Württemberg. The lands of the old Duchy of Württemberg had been Protestant. In contrast, the territories acquired by Württemberg during the Napoleonic era were predominantly Catholic. This included, for example, the numerous monastic, noble, and imperial city dominions in the south, between the Swabian Alb and Lake Constance, as well as parts of eastern Württemberg, such as the imperial city of Schwäbisch Gmünd or the prince-province of Ellwangen.

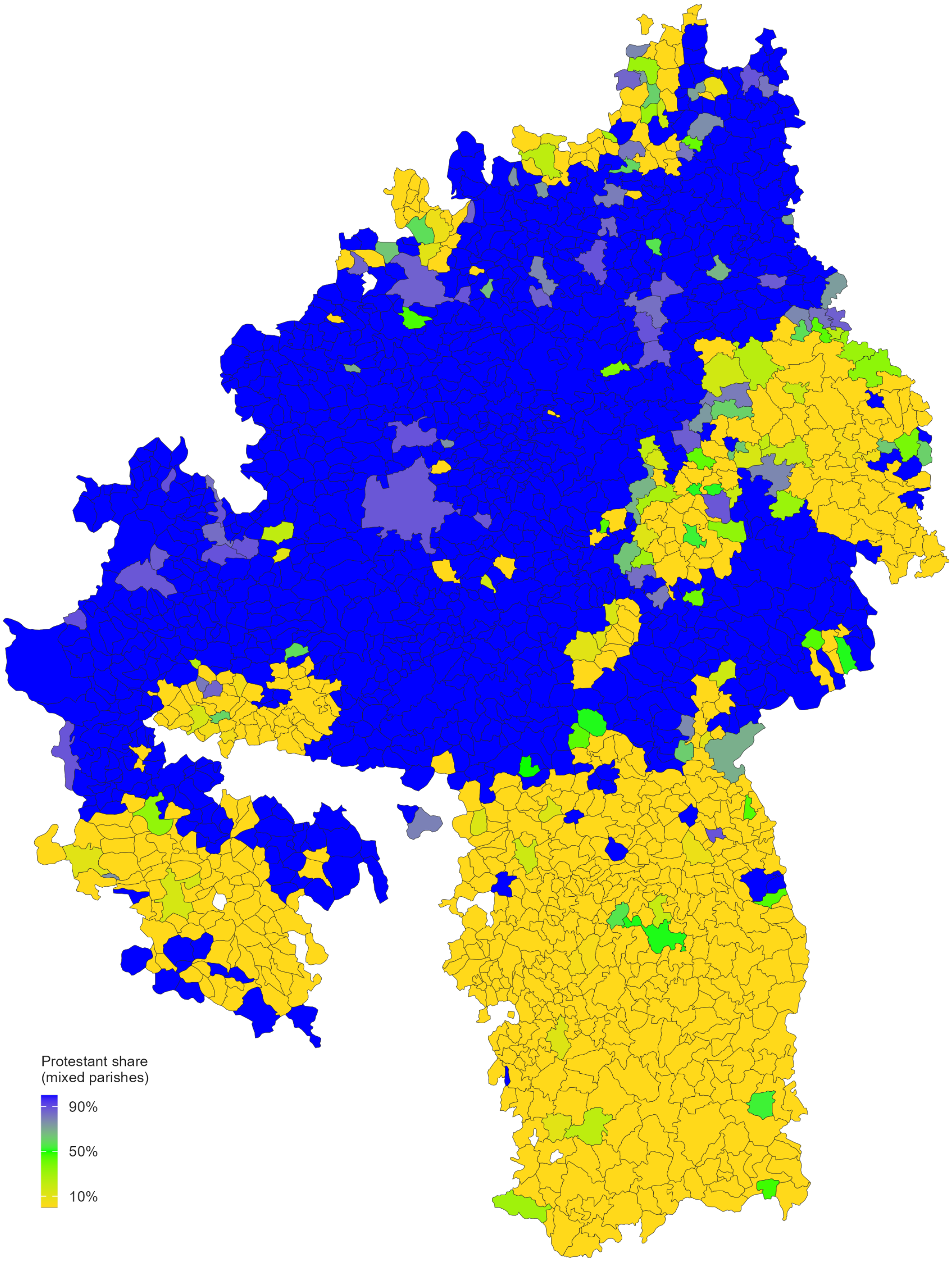
Overall, the 1871 census counted 1.249 million Protestants and 0.554 million Catholics in the Kingdom of Württemberg (Statistisches Landesamt Baden-Württemberg, 2008). Only 16,000 people belonged to other denominations, most of them Jews. Figure 1.1 shows the denominational structure of Württemberg at the parish level at the time. The different confessional areas reflect the territorial divisions of the Old Empire. Since the Peace of Augsburg in 1555, the local sovereign determined the confession in his territory according to the principle of "cuius regio eius religio" ("whose realm, his religion").[2]

For common subjects who refused to convert to the Lord's faith, emigration was the only option. Consequently, the vast majority of parishes still had a clear Catholic or Protestant majority in 1871. Only in 152 of the 1,764 parishes were less than 90% of the population members of the majority denomination, and in more than 40% of the parishes were all residents of the same denomination.

The principle of "cuius regio eius religio", together with the territorial fragmentation of the German Southwest before 1803, created several religious enclaves, i.e. Catholic parishes surrounded by Protestant areas and vice versa. It is these religious enclaves that our empirical approach exploits. To give an example, *Neuhausen auf den Fildern* was a Catholic village in the middle of Württemberg surrounded by Protestant parishes. The parish became part of *Vorderösterreich* in the 14th century after Rudolf III of Hohenberg, heavily in debt and without male heirs, sold his property to the Habsburgs. As part of Vorderösterreich, Neuhausen remained Catholic while all the surrounding villages became Protestant during the Reformation. In 1769 the town was purchased by the Archbishopric of Speyer, and in 1806, under pressure from

[2]Only ecclesiastical rulers were not covered, and would lose their office and associated possessions if they converted to another faith.

Figure 1.1: Württemberg's denominational structure, 1871



Notes: Protestant parishes are defined as those with a Protestant population of 90% or more. Similarly, Catholic parishes are those with a Catholic population of 90% or more. The remaining parishes are classified as mixed and shaded by their Protestant population share. Our identification strategy in this paper relies on having the religious enclaves that are over 90% one religion surrounded by the other.

Source: Statistisches Landesamt Baden-Württemberg (2008). Own illustration.

Napoleon, Neuhausen became part of the newly founded Kingdom of Württemberg.

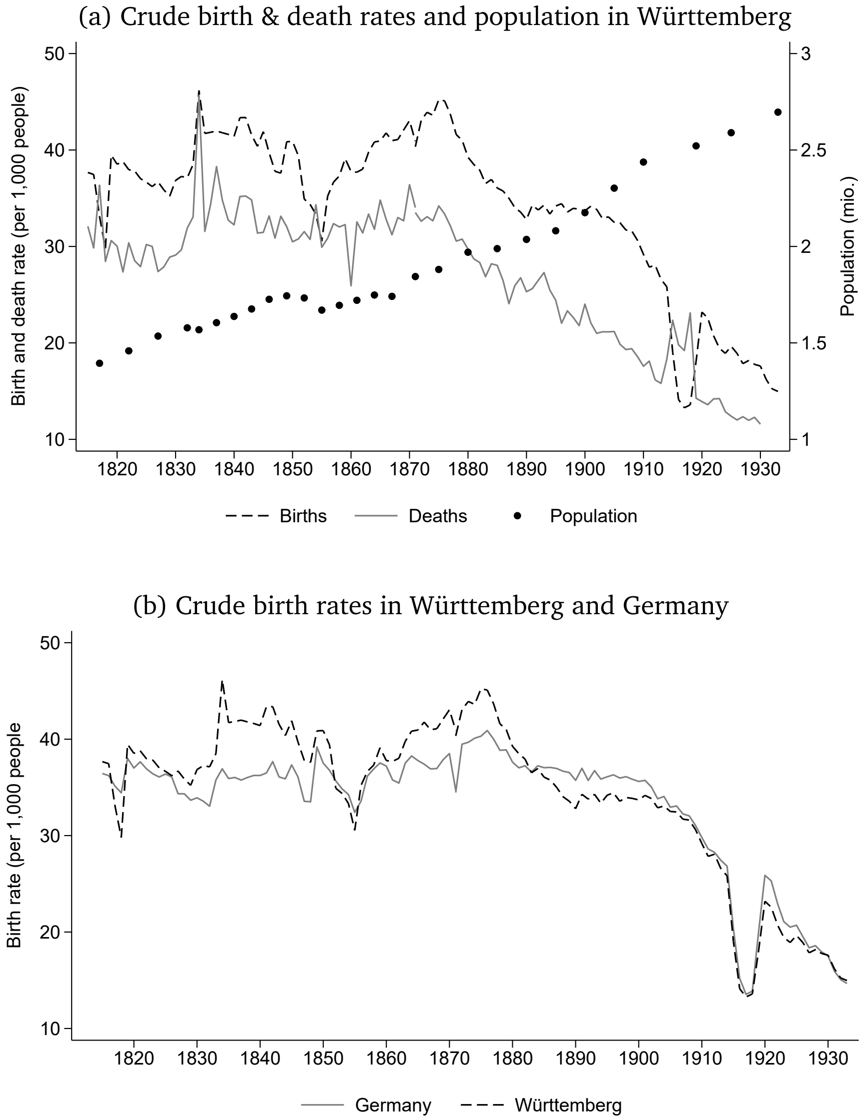
Births, deaths, and population in Württemberg, 1815–1933. After the Napoleonic era, the Kingdom of Württemberg counted 1.397 million inhabitants in 1815. Panel (a) of Figure 1.2 shows the subsequent population movements in Württemberg until 1933. It depicts crude birth and death rates, defined as the number of births and deaths per 1,000 persons, along with the state's population.[3] Several facts stand out.

First, fertility exceeded mortality in almost every year from 1815 to 1933 (except, of course, World War I). As a result, the population grew from 1.568 in 1834 to 1.844 million in 1871, 2.175 million in 1900, and 2.696 million in 1933. Especially until the turn of the century, Württemberg's negative migration balance partly offset the birth surplus and limited population growth (Griesmeier, 1954). More than 337,000 people left Württemberg in 1834-1871 alone (von Hippel, 1984), a result of the state's relatively poor economic conditions (see the paragraph on industrialization below).

Second, population development was initially highly dependent on fluctuating crop yields. The poor harvests and hunger crisis of 1816/17, the 1830s, and the mid-19th century are clearly visible in rising mortality and falling fertility rates. At the middle of the century, the absolute number of inhabitants even declined (from 1.755 million in 1849 to 1.670 million in 1855), as mortality briefly exceeded fertility and a new wave of emigration set in.

[3]In Württemberg, the origins of state statistics go back to the 16th century (Schaab, 1967). The upheavals of the Napoleonic era greatly increased the importance of collecting statistical data, especially on the newly acquired territories. In 1822, the *Statistisch-Topographisches Bureau* was established to provide an accurate and comprehensive description of Württemberg (see Statistisches Landesamt Baden-Württemberg, 1970, for an overview of the administrative statistic in Württemberg since 1820). Historically, censuses were conducted to assess the local population based on family registers kept by the clergy. A royal decree issued in 1823 required that the "population trend" be determined annually. To meet this requirement, births and deaths were to be recorded each year in every parish. From 1834 to 1867, triennial censuses conducted under the regulations of the German Customs Union (*Zollverein*) counted all persons present in a parish on the day of the census (*ortsanwesende Bevölkerung*). The first joint census in all states of the newly founded German Empire took place on December 1, 1871. With the general introduction of the civil registry system, births, marriages, and deaths were recorded and analyzed according to the regulations established for the German Empire beginning in 1875.

Figure 1.2: The demographic transition in Württemberg, 1815-1933



Notes: Panel (a) shows crude birth and death rates, defined as the number of births and deaths per 1,000 persons, in Württemberg, together with the population of the country. Panel (b) contrasts crude birth rates in Württemberg and the (future territory of the) German Empire. The population refers to the population living permanently in a locality (*ortsangehörige Bevölkerung*) until 1832, the population present in a locality area (*ortsanwesende Bevölkerung*) from 1834 to 1925, and the residential population (*Wohnbevölkerung*) in 1933.

Source: Data on birth and death rates for 1815-1871 come from Fertig et al. (2018). For Germany, data on birth and death rates for 1871-1933 are from Rothenbacher and Fertig (2022). For Württemberg, we aggregate our own data from the Staatsarchiv Ludwigsburg (see Section 1.4) for details. The population figures in panel (a) are from Statistisch-Topographisches Bureau (1875) for 1817-1832, from Zimmermann et al. (2023) for 1834-1925 and from Statistisches Reichsamt (1934) for 1933.

Third, the fertility rate began its steady decline in the late 1870s. The decline began from an unusually high level: after the mid-century crisis, the fertility rate rose again in the second half of the 1850s and reached a new peak in the 1870s. As shown in Figure 1.2 (b), the peak of the fertility rate in the 1870s (and its subsequent decline) is also visible for the German Empire as a whole. The high birth rates of the 1870s are usually attributed to the high number of marriages following the Franco-Prussian War of 1870/71, liberalized marriage laws and freedom of movement, and the (short-lived) economic boom after the founding of the German Empire (Knodel, 1967, 1974). Importantly for our purposes, the fertility decline in Württemberg (and elsewhere in Germany, see Knodel, 1974) did not begin before 1871, the starting point of our sample.[4]

Industrialization. The newly founded Kingdom was a relatively poor German state in the 19th century and a latecomer to industrialization (Marquardt, 1985; Seybold, 1974). Its lack of raw materials such as coal and ore hindered the development of heavy industry, leaving manufacturing dependent on water and animal power. The state's hilly terrain and poor transportation infrastructure further slowed growth by limiting access to raw materials and markets (Braun and Franke, 2022). Württemberg's poor economic conditions made the kingdom a major source of emigration from Germany in the 19th century. Industrialization did not pick up until the end of the century. In 1895, 44.4 percent of Württemberg's workforce was still employed in agriculture, compared to 36.2 percent in the broader German Empire. It was not until 1933 that Württemberg's share of the population in industry surpassed the national level, and the state became an immigration country for the first time during the Nazi era.

[4]In his landmark study, Knodel (1974) dates the beginning of the fertility decline in the four Württemberg counties to between 1883 and 1890, when he identifies the beginning of the transition as the first year in which fertility declined by 10 percent. If he instead uses a target fertility rate that "is clearly within the range of transition fertility" as a measure, he dates the fertility decline in Württemberg only to the early 20th century.

1.4 Data

Our dataset covers all civil parishes (*Gemeinden*) in the Kingdom of Württemberg. These parishes were the smallest administrative units in Württemberg. We use parish boundaries as of 1939 and merge parishes that were merged in earlier years.[5] This leaves 1764 parishes. They have a median area of 8.8 square kilometers and their median population is 653 inhabitants in 1871. Our dataset starts with the foundation of the German Empire in 1871 and ends with the Nazis' seizure of power in 1933. More details about the sources and the definition of all variables can be found in Appendix Section A.1.

Demography. We digitized annual data on births and deaths for all parishes in Württemberg for the years 1871 to 1933 from archival records. Birth records include live births, stillbirths (until 1930), and out-of-wedlock births. Death records include information on infants who died in their first year of life in addition to the total number of deaths. We add parish-level population data from the 12 censuses conducted in the Kingdom of Württemberg between 1871 and 1933 (Statistisches Landesamt Baden-Württemberg, 2008; Württembergisches Statistische Landesamt, 1935) and linearly interpolate population for the years between censuses.

We calculate crude birth rates as the total number of births (including stillbirths) per 1,000 inhabitants.[6] We also use crude birth rates to determine the onset of fertility

[5]We digitized parish boundaries from Kommission für geschichtliche Landeskunde in Baden-Württemberg (1972) and use information on boundary changes from Statistisches Landesamt Baden-Württemberg (2008). A total of 130 parishes are affected by mergers.

[6]Because we observe parish-level births and population but not the underlying age, sex, and marital-status structure, we cannot construct Coale's fertility indices (marital fertility index I_g, the marriage index I_m, the overall fertility index I_f) or age-standardized rates. The crude birth rate therefore also reflects nuptiality, age structure, and migration. Nuptiality is similar across denominations and, if anything, slightly higher in Protestant parishes (Appendix Figure A.5), which works against the earlier Protestant decline. Age structure and migration are held roughly fixed by our localized designs: neighboring parishes, and especially enclaves and their immediate neighbors, shared the same labor markets and migration environment, so differences in age composition between them are small. We also exclude the 24 towns above 5,000 inhabitants, which removes the parishes most affected by selective in-migration during industrialization. Our within-county, neighborhood, and enclave comparisons further difference out the broad compositional differences that would most plausibly contaminate a crude birth rate comparison.

decline. We define the fertility transition date FTD_i of a locality i as the first date for which the 5-year moving average of the crude birth rate falls at least 10% below its pre-transition level and never returns to the pre-transition level.[7] At each time t, the fertility transition status T_{it} is then defined as 1 if $t \geq FTD_i$ and 0 otherwise. The main difficulty of this approach is to determine the pre-transition fertility (Knodel, 1974).[8] This is particularly true in our context, where fertility levels were higher in the 1870s, our first years of data, than either before or after (see Section 1.3).

For robustness, we use two measures of fertility before the transition. The baseline measure is the average crude birth rate of the county in which a parish is located from 1812 to 1866 (Kull, 1875). The main advantage of this measure is that it is a long-term average from before the fertility decline, unaffected by the unusually high fertility of the 1870s (and other short-term fluctuations). The disadvantage is that the threshold is the same for all parishes in the same county (of which there are 64), and thus it ignores within-county differences in the level from which the decline began. Our second measure takes such differences into account. It is the average crude birth rate at the parish level in the first five years of our parish data, but excluding the years of unusually high fertility from 1872 to 1877. In other words, this second measure of pre-transition fertility is the average crude birth rate of a parish in 1871, 1878, 1879, 1880, and 1881.[9]

We acknowledge that the estimated transition years depend strongly on the (arbitrary) choice of the required percentage decline and the measure of pre-transition fertility. However, our interest is less in precisely dating the onset of the decline than in examining the *differences* across parishes.

[7]We use 5-year moving averages to smooth out the large year-to-year fluctuations that are particularly visible in small villages. The threshold of 10% is arbitrary but has been used in previous literature Knodel (1974); Spolaore and Wacziarg (2022).

[8]In his data for 1871-1939 at district level, of which there are four in Württemberg, Knodel (1974) defines the pre-transition level as the maximum fertility level, excluding the years 1874-77 from consideration.

[9]Ideally, we would like to have the long-term average at the parish level from before 1871. However, the archival records only contain fertility data beginning in 1871.

Industrialization. We measure early industrialization using an indicator for the local presence of a steam engine and total steam capacity per 1,000 persons in 1869, both from archival records. For later years, we use parish-level information on the number of full-time industrial workers per 100 persons from the 1895 and 1907 occupational censuses (Königliches Statistisches Landesamt, 1900, 1910).

Religion. Data on the number of Protestants, Catholics, and other denominations are taken from Statistisches Landesamt Baden-Württemberg (2008) and refer to the year 1843. Unless otherwise noted, we define Protestant parishes as those with a Protestant population of 90% or more. Similarly, Catholic parishes are those with a Catholic population of 90% or more. The remaining parishes are classified as mixed. We also check our classification against census data on religious groups for 1871 (Statistisches Landesamt Baden-Württemberg, 2008), but the religious composition of a parish changes very little before the industrialization process begins (see Table A.3).[10]

Control variables. Our control variables include average altitude and ruggedness from Bundesamt für Kartographie und Geodäsie (2017) and soil quality from Bundesanstalt für Geowissenschaften und Rohstoffe (2013), and a dummy variable for parishes located on a river navigable in 1845 (Kunz and Zipf, 2008). As a measure of pre-Reformation development, we use indicator variables for parishes with city and market rights before the Reformation (Kommission für geschichtliche Landeskunde in Baden-Württemberg, 1973, 1982).

[10]Only then does internal migration make the religious composition more mixed, but these changes apply mostly to larger cities. Only 52 parishes change classification during these years, mostly going from high 80s to 90s. Only 12 out of 1764 observations start with over 90% one denomination and then drop below this number. The appendix shows that the choice of year does not change the results. 1843 is chosen to obtain binary religion identification for all parishes before any major events in 19th century.

1.5 Empirical Strategy

This section outlines our empirical strategy for estimating the causal effect of a parish population being Protestant rather than Catholic on crude birth rates, the onset of the fertility decline, and industrialization.

Baseline specification. Our baseline regression for examining the impact of Protestantism is:

$$y_{it} = \alpha + \beta_t Protestant_i + f(lat_i, lon_i) + \boldsymbol{X}_i\boldsymbol{\gamma} + \epsilon_{it}, \tag{1.1}$$

where y_{it} is a particular outcome of parish i measured at year t, $Protestant_i$ is an indicator variable for Protestant parishes, $f(lat_i, lon_i)$ is a spatial trend based on longitude and latitude, $\boldsymbol{X}_i$ is a (row) vector of control variables, and ϵ_{it} is an error term. The parameter of interest, β_t, measures the mean outcome difference between Protestant and Catholic parishes in year t. For period-specific outcomes, we estimate equation (1.1) separately for each five-year period, so that β_t traces the Protestant-Catholic difference over time.

Identification requires that, conditional on the control variables, Protestant parishes do not differ from Catholic parishes in unobserved predetermined characteristics that affect their fertility behavior or industrial development in our sample period. For example, Protestant and Catholic areas were characterized by different geographies, as most Catholics and Protestants lived far apart in different parts of Württemberg (see Section 1.3). These geographic characteristics may have had an independent effect on local fertility patterns, for example by influencing the process of industrialization (Wanamaker, 2012) or the spread of knowledge about new contraceptive techniques (Klüsener et al., 2019).[11]

[11] In fact, pre-industrial population densities were largely determined by natural conditions, so that in 1834, for example, the sunny Neckar basin around Protestant Stuttgart had a significantly higher population density than largely Catholic Upper Swabia (Kommission für geschichtliche Landeskunde in Baden-Württemberg, 1988). Access to navigable rivers was also unequally distributed, as the Danube, for example, flowed predominantly through Catholic areas.

We address these concerns in four main ways: by adding suitable controls, by using only within-county variation, by demeaning the data with respect to a neighborhood of parishes, and by comparing religious enclaves with their neighboring communities.

Control variables. All our regressions include a quadratic in longitude and latitude as a simple way of capturing spatial trends (as in Conley and Kelly, 2025). We also include controls for geographical characteristics and indicators of pre-Reformation development. We describe these variables in more detail in our balancing tests below.

Within-county variation. Our second identification strategy exploits only within-county variation by adding county-fixed effects to equation (1.1). Of the 64 counties, 40 have both Catholic and Protestant parishes. The other 24 do not contribute to the identification of our parameters of interest, β_t. Counties have an average size of 310 square kilometers, and are thus comparatively small administrative units. By comparing only parishes within counties, we are comparing parishes that are located in a geographically relatively homogeneous area. Therefore, Protestant and Catholic parishes within counties should differ less on unobserved characteristics (and we provide evidence of this below).

Within-neighborhood variation. Our third identification strategy differences equation (1.1) by subtracting from the dependent and independent variables the respective mean of all observations within a 10 km radius of the centroid of each parish.[12] The idea is very similar to our second identification strategy, but instead of comparing parishes within administrative boundaries, we compare parishes within geographic neighborhoods. Note that both our second and third identification strategies have been proposed to eliminate unit roots in spatial processes (Becker et al., 2025; Müller and Watson, 2024).

Religious enclaves. Our fourth strategy relies on "historical accidents" that created several religious enclaves in Württemberg (see Section 1.3 for details). We define

[12] We choose the radius to ensure that each parish has at least one neighbor within the radius.

religious enclaves as parishes that have been completely surrounded by parishes of the other Christian denomination since the 16th century (excluding "mixed parishes" with less than 90% adherence to the majority denomination). In total, there are 11 Catholic and 11 Protestant such enclaves in Württemberg. Again, the idea is that the surrounding communities provide a suitable control group, as they are likely to differ from the enclaves only in their Christian denomination (and we provide evidence for this below). We always include an indicator variable for enclaves to account for possible effects of being an enclave (for the effect, see A.6).

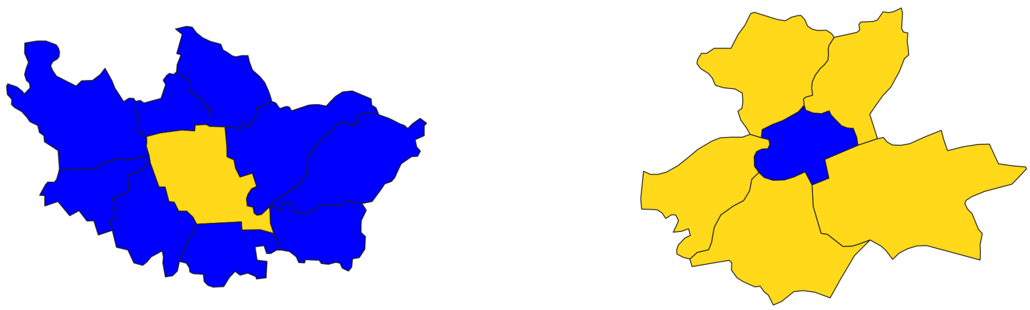
Figure 1.3 provides two examples of religious enclaves. The left panel shows *Neuhausen auf den Fildern*, a Catholic village in the middle of Württemberg surrounded by Protestant parishes. The parish became part of *Vorderösterreich* in the 14th century after Rudolf III von Hohenberg, heavily in debt and without male heirs, sold his property to the Habsburgs. As part of Vorderösterreich, Neuhausen remained Catholic while all of the surrounding villages became Protestant during the Reformation. Neuhausen became part of the Kingdom of Württemberg in 1806. Our empirical strategy compares Neuhausen with the surrounding villages of Unterensingen, Wolfschlugen, Denkendorf, Köngen, Nellingen auf den Fildern, Bernhausen, Plieningen, Scharnhausen, and Sielmingen. The right panel of Figure 1.3 shows *Mundingen*, a Protestant parish near Ehingen, surrounded by Catholic parishes. The parish was founded during the 9th century and belonged to Württemberg since the 14th century. This parish was one of the first ones to become Protestant in the 16th century. The surrounding villages, however, stayed Catholic. These parishes are Dächingen, Granheim, Kirchen, Lauterach, and Erbstetten. Appendix Table A.1 lists all 22 enclaves, along with their surrounding villages that we use as control group.

Standard errors. We report standard errors clustered at the county or enclave level to account for spatial correlation within broader geographic areas. In robustness checks, we also construct confidence intervals based on the spatial correlation principal components (SCPC) methods suggested by Müller and Watson (2023), which

Figure 1.3: Examples of religious enclaves and their surrounding parishes

(a) Neuhausen auf den Fildern

(b) Mundingen



Notes: The plot shows two examples of religious enclaves and their surrounding parishes (of the other Christian denomination). *Neuhausen auf den Fildern* in Panel (a) is a Catholic parish surrounded by Protestant parishes. *Mundingen* in Panel (b) is a Protestant parish surrounded by Catholic parishes.

accounts for many forms of spatial correlation.

Balancing test. Our main identification strategies to deal with unobserved heterogeneity compare Protestant parishes to nearby Catholic parishes (and vice versa). We define "nearby" either as being in the same county, within a 10 km radius, or as the parishes surrounding a religious enclave. The idea is simple: Neighboring parishes should share similar pre-Reformation characteristics. Thus, post-Reformation differences in fertility should be due to religious denomination rather than some third factor that correlates with religion and determines fertility.

Table 1.1 supports our empirical strategy. The table first documents marked differences between Catholic and Protestant parishes in predetermined characteristics (Columns (1)-(3)). For example, Catholic parishes are located on average 118 m higher than Protestant parishes and are more than twice as likely to be located on a navigable river. At the same time, only 6.2% of Catholic parishes had city status before the Reformation, 4 percentage points less than later Protestant parishes. These differences illustrate the problems of comparing Protestant and Catholic parishes even within relatively small states such as Württemberg: because they are typically located in different parts of a state, they often differed considerably already before the Reformation.

Columns (4) to (6) show that our identification strategies level out differences in (observable) characteristics. Column (4) reports within-county differences between Catholic and Protestant parishes (from regressions of each characteristic on a Protestant parish indicator and county fixed effects). The differences shrink markedly for all characteristics, although a few differences remain statistically significant. Column (5) regresses the deviation of each characteristic from the mean of the characteristics on the Protestant indicator transformed by the same spatial difference transformation. Again, differences shrink markedly and mostly turn statistically insignificant. Finally, Column (6) restricts the sample to religious enclaves and their surrounding parishes, comparing only neighboring parishes in regressions of pre-Reformation characteristics on a Protestant indicator and case-fixed effects (where a case refers to the enclave and its surrounding "control parishes"). None of the differences remain statistically significant (although the small sample size may contribute to this result).

1.6 Protestantism and the Fertility Decline

This section reports our results on the impact of Protestantism on the (timing of the) fertility decline. We begin with a descriptive analysis of Protestant-Catholic differences in fertility before presenting (causal) regression estimates. Our analysis focuses on the years 1871 to 1933. We exclude denominationally mixed parishes where less than 90% of the population belong to the majority denomination.

Table 1.1: Balancing Test for Pre-Reformation Characteristics

	Averages:		Protestant coefficient:			
				Conditional on		
	Protestant	Catholic	Unconditional	Counties	Neighborhoods	Enclaves
	(1)	(2)	(3)	(4)	(5)	(6)
Area (*sq. m*)	622.471	685.431	-62.960	71.010	41.965	133.770
	(530.927)	(468.362)	(61.780)	(46.124)	(32.706)	(119.654)
Average elevation (*in m*)	454.665	573.403	-118.738***	24.179***	1.022	3.995
	(156.554)	(126.481)	(24.219)	(8.366)	(5.321)	(5.355)
Ruggedness	48.952	36.960	11.992***	-1.369	-3.447*	-5.312
	(25.232)	(22.923)	(3.645)	(2.883)	(1.833)	(4.673)
Navigable river (0/1)	0.047	0.117	-0.069**	-0.049*	-0.016	0.005
	(0.213)	(0.321)	(0.030)	(0.029)	(0.028)	(0.074)
Soil type:						
Coastal area and bog soils	0.002	0.012	-0.010**	0.007	0.007	0.026
	(0.033)	(0.070)	(0.005)	(0.008)	(0.007)	(0.029)
Broad river valleys	0.004	0.122	-0.118***	-0.026*	-0.006	0.020
	(0.053)	(0.237)	(0.024)	(0.015)	(0.014)	(0.026)
Undulating lowlands and hilly areas	0.001	0.294	-0.292***	-0.012	-0.009	-0.033
	(0.029)	(0.403)	(0.070)	(0.009)	(0.008)	(0.032)
Loess area	0.208	0.071	0.136***	0.042	0.044*	-0.014
	(0.325)	(0.198)	(0.037)	(0.029)	(0.023)	(0.048)
Mountain and hill soils	0.781	0.501	0.281***	-0.010	-0.038*	-0.004
	(0.331)	(0.454)	(0.088)	(0.031)	(0.020)	(0.053)
Alpine soils	0.004	0.001	0.003	-0.000	0.001*	0.005
	(0.039)	(0.006)	(0.002)	(0.000)	(0.001)	(0.004)
Pre-reformation city (0/1)	0.102	0.062	0.040***	0.001	0.014	0.032
	(0.303)	(0.241)	(0.013)	(0.027)	(0.023)	(0.083)
Pre-reformation market (including cities) (0/1)	0.123	0.080	0.043***	0.006	0.007	0.027
	(0.328)	(0.272)	(0.014)	(0.027)	(0.021)	(0.086)

Notes: The table shows the mean values of the pre-treatment characteristics for Protestant parishes (Column (1)) and Catholic parishes (Column (2)). Column (3) shows the mean difference in pre-treatment characteristics between Protestant and Catholic parishes. Column (4) reports mean differences between Protestant and Catholic parishes conditional on county fixed effects. Column (5) reports mean differences from a regression of transformed characteristics on a transformed indicator for Protestant parishes, where the transformation subtracts the respective mean of all observations within a 10 km radius of the centroid of each parish. Column (6) restricts the sample to religious enclaves and presents mean differences conditional on fixed effects for each enclave case (consisting of an enclave and its surrounding parishes). See main text and appendix for data sources and details. Standard deviations are in parentheses (Columns (1) and (2)). Robust standard errors clustered at the county level (Columns (3)-(5)) or at the enclave case level (column (6)) are in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

1.6.1 Descriptive Analysis

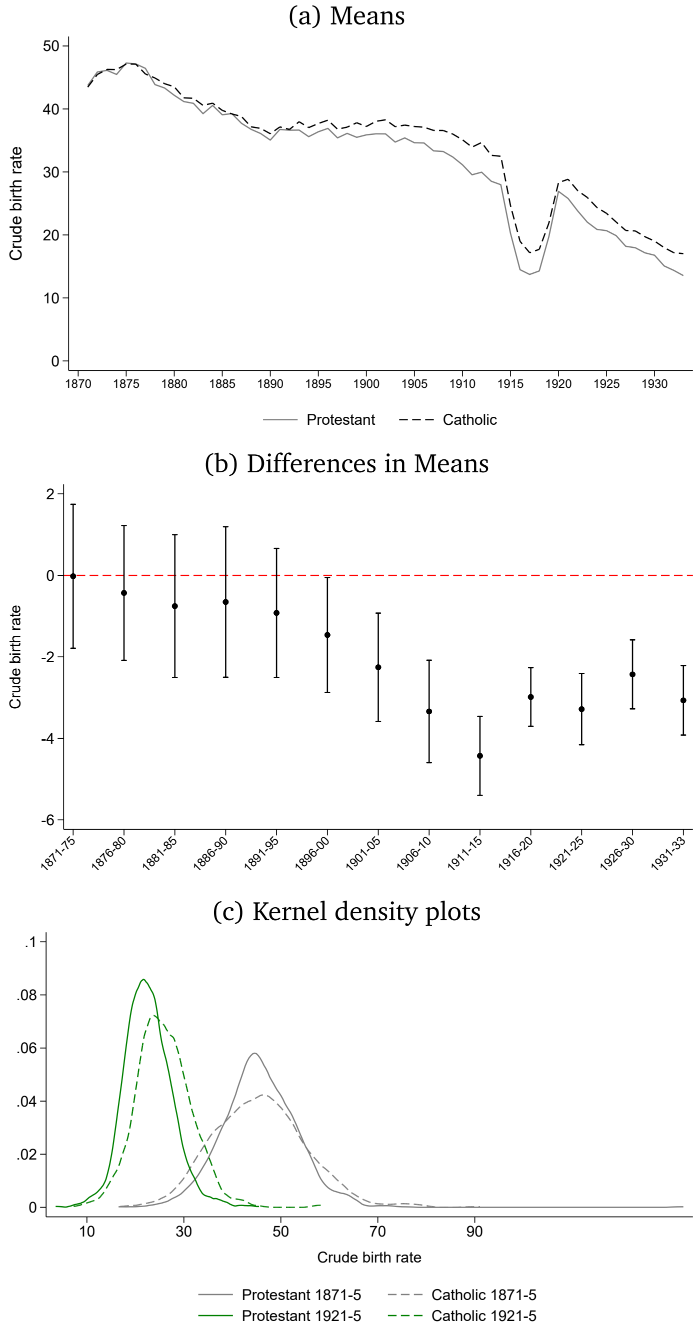
Crude birth rates. Figure 1.4 presents descriptive statistics on crude birth rates of Catholic and Protestant parishes over time. Panel (a) shows annual means separately for Protestant and Catholic parishes, whereas panel (b) depicts unconditional differences. Perhaps surprisingly, mean birth rates of Protestant and Catholic parishes evolve in tandem until 1895, falling from a peak of about 45 in the mid-1870s to about 35-36 in 1890. At least part of this decline reflects a normalization from the relatively high birth rates of the early 1870s rather than the beginning of the fertility decline.

It is only after 1890 that Protestant parishes begin to have systematically lower fertility rates than Catholic parishes. In fact, average birth rates in Catholic parishes do not decline between 1890 and 1910, while Protestant fertility gradually declines during this period. In 1910, the average crude birth rate was 31 in Protestant parishes and 35 in Catholic parishes. Over the next 23 years, until the Nazis come to power in 1933, fertility rates again move largely in parallel, with the differences between Catholic and Protestant parishes remaining relatively stable at 3 to 4 births per 1,000 people.

Panel (c) of figure 1.4 shows the full distribution of crude birth rates by denomination for two five-year averages, 1871 to 1875 and 1921 to 1925. Similar means at the beginning of the sample period hide differences at the top and bottom of the distribution. In particular, Catholic parishes have significantly more weight in the tails (relative to the center) of the distribution than Protestant parishes.[13] Forty years later, the distribution of both Protestant and Catholic parishes has shifted to the left, toward lower fertility rates. However, this shift is much more pronounced for Protestant parishes than for Catholic parishes. We observe considerable overlap in the

[13] Appendix Figure A.4 shows that in the early 1870s, Protestant parishes have higher fertility rates than their Catholic counterparts at the 10th percentile, while the opposite is true at the 90th percentile. This pattern does not disappear until the late 1890s, after which Protestant parishes also have lower fertility rates at the lower end of the distribution.

Figure 1.4: Crude birth rates by religious denomination



Notes: The plot shows crude birth rates at the parish level by religious denomination. Panel (a) shows annual means separately for Protestant and Catholic parishes. Panel (b) depicts unconditional mean differences between Protestant and Catholic parishes for crude birth rates averaged over five years, along with 95% confidence intervals based on robust standard errors clustered at the county level. Panel (c) shows kernel density estimate for average crude birth rates in 1871-75 and 1921-25 for Protestant and Catholic parishes.

parish-level distribution of 1871-75 and 1921-25 for Catholic parishes, while there is very little overlap for Protestant parishes.

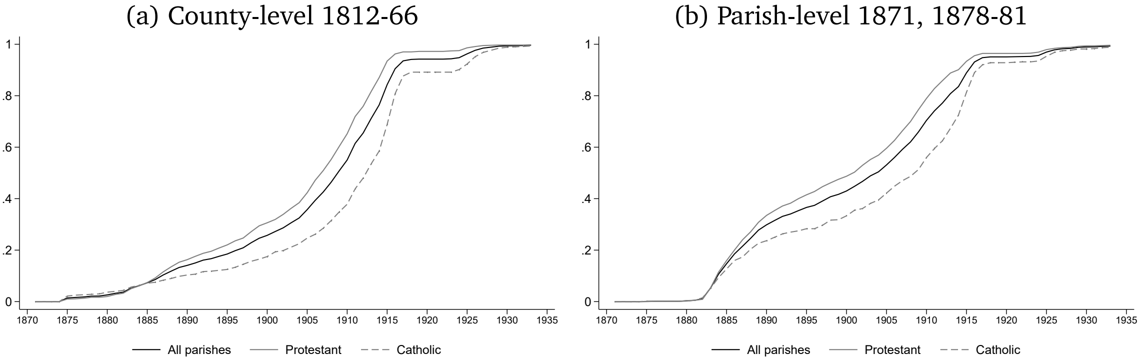
Transition dates. Next, we use crude birth rates to describe the differences in the onset of the fertility decline between Protestant and Catholic parishes. Figure 1.5 shows the cumulative distribution of fertility transition dates for all parishes and separately for Protestant and Catholic parishes. Panel (a) uses our baseline measure of pre-transition fertility measured at the county level from 1812-66, panel (b) our alternative measure measured at the parish level for 1871 and 1878-81 (see Section 1.4). Both indicators show significant heterogeneity in transition dates at the parish level. Under our baseline indicator, the mean fertility transition date is 1906, with a standard deviation of 11.5 years. The interquartile range is from 1900 to 1914. The alternative indicator dates the fertility transition earlier, with the mean in 1901 (and an even higher standard deviation of 12.6).

Both indicators show that Protestant parishes experienced the onset of fertility decline much earlier. For example, it took until 1898 for a quarter and until 1907 for half of the Protestant parishes to experience the onset of the fertility decline (baseline indicator). Catholic parishes did not reach these points until 1905 and 1913. On the eve of the First World War, in 1913, almost one third of Catholic parishes had not yet experienced a decline in fertility, while this was the case for only 11% of Protestant parishes. The differences tend to be even greater under the alternative indicator of pre-transition fertility.

1.6.2 Regression results.

We now assess whether raw differences in crude birth rates and transition dates can be given a causal interpretation or, alternatively, are due to omitted variables. From now on we drop towns with at least 5,000 inhabitants in 1871, of which there were 24 in Württemberg at that time. This is because our identification strategy of finding suitable control parishes among nearby parishes is unlikely to work for the few large towns. However, our results remain similar when we keep large towns in the sample (as Appendix Table A.3 shows).

Figure 1.5: Cumulative distribution of fertility transition dates by pre-transition measure

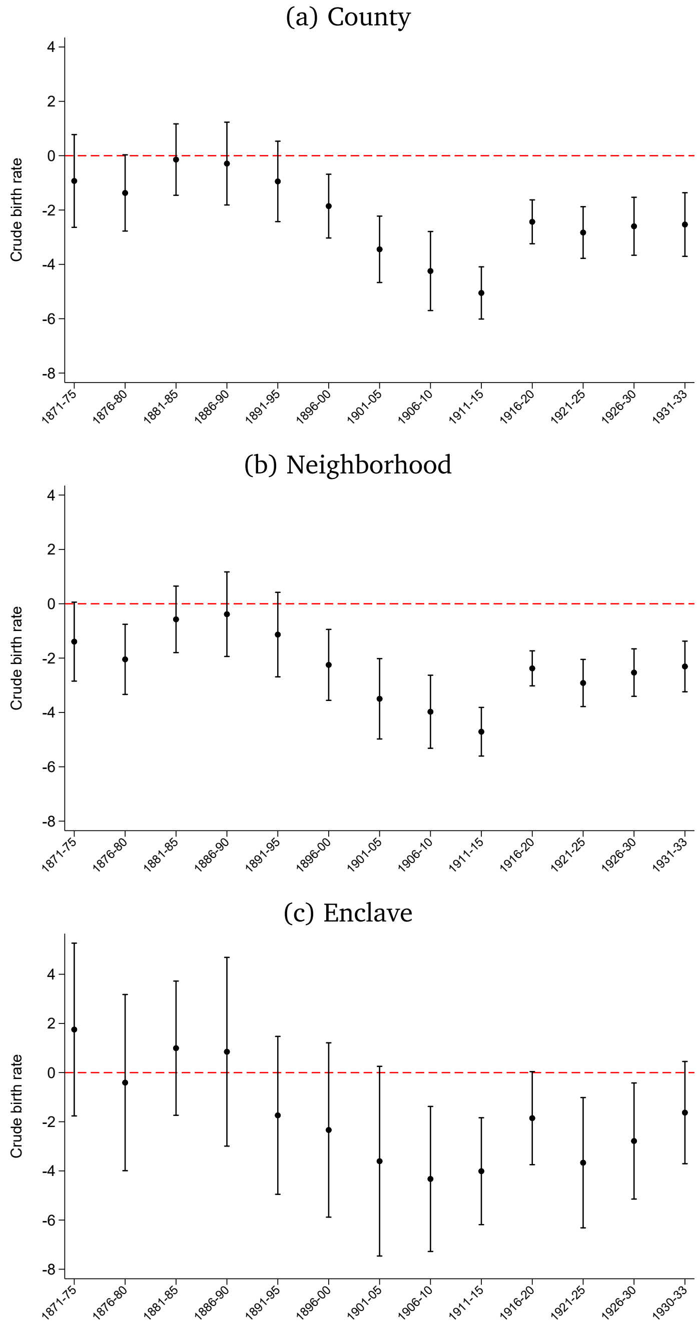


Notes: The figures show the cumulative distribution of fertility transition dates for all parishes and separately for Protestant and Catholic parishes. We define the fertility transition date of a locality i, FTD_i, as the first date for which the 5-year moving average of the crude birth rate falls permanently by at least 10% below the pre-transition fertility. For each date t, the fertility transition status T_{it} is defined as one if $t \geq FTD_i$ and zero otherwise. The plot shows the average T_{it} over time. Panel (a) defines pre-transition fertility as the average crude birth rate of the county (*Oberamt*) in which a parish is located from 1812 to 1866. Panel (b) measures pre-transition fertility as the average crude birth rate at the parish level in the years 1871, 1878, 1879, 1880, and 1881.

Crude birth rates. Figure 1.6 shows the impact of Protestantism on crude birth rates over time. It plots β_t from equation (1.1) for five-year periods t from 1871-75 to 1931-33. We focus on the results of regressions that add county fixed effects to the baseline regression (panel (a)), use only deviations from the mean in a circular neighborhood around each parish (panel (b)), or compare religious enclaves to their surrounding parishes (panel (c)). All regressions include a quadratic in longitude and latitude. Enclave regressions also include an indicator variable for enclaves. Appendix Figure A.7 shows very similar results for regressions that also include the control variables listed in Table 1.1.

The regression results from all three types of regressions confirm that in the 1870s and 1880s there was little difference in crude fertility rates between Protestant and Catholic parishes (only the neighborhood regressions show statistically significant differences in the 1870s, which disappear in the 1880s). It is only from the 1890s that we observe that Protestant parishes begin to have systematically lower fertility rates than Catholic parishes.

Figure 1.6: Impact of Protestantism on crude birth rates, 1871-1933



Notes: The figure shows the effect of Protestantism on crude birth rates in each five-year interval from 1871-75 to 1931-33, based on regression model 1.1. Panel (a) adds county fixed effects to the baseline regression. Panel (b) estimates the baseline regression on data transformed by taking deviations from the mean in a circular neighborhood around each parish. Panel (c) restricts the sample to enclaves and surrounding parishes, adding fixed effects for each case. All regressions include a quadratic in longitude and latitude. Enclave regressions also include an indicator variable for enclaves. Robust standard errors clustered at the county level (panels (a) and (b)) or at the enclave case level (panel (c)). Point estimates are indicated by a dot. Vertical bands indicate the 90 percent confidence interval around each estimate.

In the within-county regressions, the gap gradually widens until 1911-15, when it reaches -5.0 births per 1000 persons, a decline in fertility of 0.80 standard deviations. For comparison, the urban-rural fertility gap at that time was less than 4.3 births.[14] Compared to a control mean of 31.7 births in Catholic parishes at that time, Protestantism reduced birth rates by almost 16%. The gap then closed with World War I in 1914-18. By the end of our sample period, in 1931-33, the crude birth rate in Protestant parishes was 2.5 births lower than in Catholic parishes. The patterns are similar in the neighborhood and enclave regressions. In the former regression, the fertility gap peaks in 1911-15 at -4.7 births. In the latter regressions, the gap peaks at -4.3 births in 1906-10. In both regressions, the gap closes again with World War I but remains above 2 births in the 1920s.

Transition date. Table 1.2 reports our results of the effect of Protestantism on local fertility transition dates. We focus on our baseline measure of pre-transition fertility measured at the county level from 1812-66. Appendix Table A.2 reports the results for the alternative measure.[15] Columns (1) and (2) identify the effect by comparing Protestant and Catholic parishes within counties, with and without additional control variables. Columns (3) to (4) identify the effect from comparisons within a circular neighborhood, again with and without additional controls. The specification in Columns (5) and (6) compares religious enclaves to their neighboring parishes.

Overall, we confirm a large effect of Protestantism on the year of transition. Depending on the specification, Protestant parishes experience the onset of the fertility transition between 6.3 and 10.3 years (or between 0.55 and 0.88 standard deviations) earlier. These estimates are slightly larger than the unconditional difference between Protestants and Catholics–and underline the comparatively large influence of religious denomination on the timing of the fertility transition. Since the control

[14]For example, the 10% largest parishes in Württemberg recorded 2.2 fewer births than the 10% smallest parishes in 1911-15.

[15]The results confirm that the onset of the fertility transition occurred earlier in Protestant parishes. However, the magnitude of the effect tends to be smaller and varies more across samples.

Table 1.2: The effect of religion on fertility transition dates

	Within transformation					
	County		Neighborhood		Enclave	
	(1)	(2)	(3)	(4)	(5)	(6)
Protestant (0/1)	-7.009***	-7.261***	-6.603***	-6.350***	-10.011***	-10.316***
	(1.337)	(1.196)	(1.307)	(1.244)	(3.096)	(3.124)
	[-0.608]	[-0.630]	[-0.573]	[-0.551]	[-0.849]	[-0.875]
Additional controls	no	yes	no	yes	no	yes
Observations	1,619	1,619	1,619	1,619	122	122

Notes: The table reports estimates of the effect of Protestantism on fertility transition dates. Columns (1) and (2) compare Protestant and Catholic parishes within counties and include county fixed effects. Columns (3) and (4) use the 10 km neighborhood transformation, subtracting local neighborhood means from the outcome and covariates. Columns (5) and (6) restrict the sample to religious enclaves and their neighboring parishes and include enclave-case fixed effects and an enclave indicator. Regressions (2), (4), and (6) include geographic characteristics and indicators of pre-Reformation development, as listed in Table 1.1, as additional control variables. Robust standard errors clustered at the county level in Columns (1)–(4) and at the enclave-case level in Columns (5)–(6) are shown in parentheses. Standardized coefficients are in brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively.

variables are balanced between Protestant and Catholic parishes within counties and neighborhoods and between enclaves and their neighbors (see table 1.1), their addition has little effect on the estimates.

1.7 Discussing the Mechanism

We have established that religion is a key predictor of regional differences in the timing of the fertility transition: The onset of the fertility transition occurred between 0.55 and 0.88 standard deviations earlier in Protestant parishes than in Catholic parishes, depending on the specification. This difference is unlikely to be explained by differences between parishes that already existed before the Reformation–and thus can be attributed to denomination. What then explains that Protestantism brought the transition forward?

One of the main hypotheses in the literature is related to economic development. Economic modernization during industrialization is a potential trigger for the fertility decline (see, e.g., Guinnane, 2011, for an overview). At the same time, Weber (1904) famously suggested that the Protestant work ethic promoted modern economic development and his hypothesis is still debated today (see, e.g., Becker and Woessmann, 2009; Kersting et al., 2020, for evidence on Prussia in the late 19th century). Thus, Protestantism may have facilitated the fertility decline by promoting economic modernization.

In this section, we start with revisiting the impact of Protestantism on economic development. We focus on economic outcomes that have been suggested as potential triggers of the demographic transition, namely industrialization, organizational and technological change, income, and the gender wage gap (see, e.g., Galor, 2012, for an overview). The combination of fine-grained data at the parish level and a novel identification strategy makes our results valuable in their own right, not only for exploring the mechanism behind the accelerated fertility decline in Protestant parishes.

Table 1.3 summarizes our main results. Panel A presents OLS regressions that include a quadratic in longitude and latitude as well as geographic and pre-Reformation controls. With the exception of the gender wage gap, all of our indicators of economic modernization are positively correlated with Protestantism. Protestant parishes have, on average, higher shares of employment in industry and non-agriculture at the turn of the century, have larger firms in 1907, were more likely to adopt steam technology, and have higher taxable income per capita in 1907. Some of the differences between Protestant and Catholic parishes are sizeable: For instance, Protestant parishes have about 3.7 percentage points (or 19.0%) higher employment rates in industry than Catholic parishes.

Are these positive correlations driven by unobserved heterogeneity between Protestant and Catholic parishes? The remainder of Table 1.3 re-estimates the relationship using only variation within counties (Panel B), within the neighborhood in a 10 km radius (Panel C), and within religious enclaves and their neighbors (Panel D). In contrast to our fertility results, the correlations largely disappear once we compare only nearby

Table 1.3: The impact of Protestantism on economic modernization

	Employment (%)				Firm size (logs)	Steam engine (0/1)	Taxable income (*Mark*)	Gender wage gap (%)	
	Industry		Outside agriculture						
	1895	1907	1895	1907	1907	1869	1907	1884	1909
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Spatial trend and controls									
Protestant (0/1)	3.647**	3.591*	3.187	3.794	0.112***	0.029*	14.672	0.010	-0.004
	(1.734)	(2.126)	(1.936)	(2.334)	(0.039)	(0.017)	(8.848)	(0.016)	(0.009)
B. Within counties									
Protestant (0/1)	1.488	0.837	0.684	0.987	0.071*	0.045	25.114***	0.004*	0.000
	(1.401)	(1.474)	(1.692)	(1.810)	(0.039)	(0.028)	(8.160)	(0.003)	(0.001)
C. Within 10km radius									
Protestant (0/1)	-0.040	-0.801	-1.289	-0.935	0.038	0.038*	20.751***	-0.004	0.001
	(1.211)	(1.291)	(1.507)	(1.523)	(0.033)	(0.019)	(7.361)	(0.007)	(0.003)
D. Within enclave clusters									
Protestant (0/1)	2.974	1.876	1.163	1.254	0.195	-0.060	-6.366	0.012	0.002
	(4.148)	(4.273)	(4.700)	(4.601)	(0.153)	(0.070)	(30.173)	(0.010)	(0.005)
Control mean	19.225	18.722	24.762	24.115	0.549	0.053	336.382	0.336	0.328

Notes: Control means and coefficients on an indicator for Protestant parishes are reported. All regressions include a quadratic in longitude and latitude as well as geographic and pre-Reformation controls. Panel B includes county fixed effects, Panel C uses deviations from the mean within a 10 km neighborhood, and Panel D includes enclave-cluster fixed effects and an enclave indicator. Standard errors are clustered by county in Panels A–C and by enclave cluster in Panel D.

parishes, which likely share unobserved characteristics.

Perhaps most importantly, Protestantism is no longer positively associated with the industrial employment share in the within-county, neighborhood, and enclave regressions. Industrialization is thus an unlikely candidate for explaining that Protestant parishes saw the fertility transition first. This finding is significant because industrial employment shares embody core elements of historical economic development—as highlighted by the Weber thesis—including modernization, the spread of capitalism, and the division of labor (Becker and Woessmann, 2009).

The most robust (positive) association is between Protestantism and taxable income per capita. This finding means that we cannot fully rule out income as a partial channel. An increase in income may reduce fertility if the negative substitution effect from higher opportunity costs of raising children dominates the positive income effect (Becker, 1960). However, the evidence for the income-fertility nexus is mixed, and unified growth theory holds that the negative correlation at the aggregate level is spurious and driven by technical change (see e.g. Galor and Weil, 2000; Madsen and Strulik, 2023). In our setting, the effect of Protestantism on income is relatively small: depending on the specification, Protestantism increases taxable income per capita by between 4.4% and 7.5% relative to the mean in Catholic parishes.[16] Moreover, the point estimates become statistically insignificant in the enclave regressions, and taxable income is not accompanied by similarly robust Protestant advantages in industrial employment, non-agricultural employment, firm size, steam technology, or the gender wage gap. Overall, income differences may have contributed at the margin, but the evidence does not support income or broader economic modernization as the primary explanation for the earlier Protestant fertility decline.

We consider several additional mechanisms and potential confounders. First, changes in fertility could mechanically reflect changes in marriage behavior. Appendix Figure A.5 shows that marriage rates evolved very similarly in Protestant and Catholic parishes, and if anything were slightly higher in Protestant parishes for much of the period in which their fertility fell earlier. Appendix Figure A.10 confirms the expected mechanical relationship: higher lagged marriage rates, and more weakly lagged out-of-wedlock births, are associated with higher crude birth rates. These margins matter for the level of fertility, but they do not explain why Protestant parishes entered the decline earlier, since nuptiality was similar or slightly higher in Protestant parishes and thus works against an earlier Protestant decline. Infant mortality tells a similar story. Appendix Figure A.8 shows that it is positively related to fertility, with the association weakening over time, but this relationship does not vary systematically by denomination: the Protestant interaction in Panel (b) is small and rarely distinguishable from zero, so denomination does not change how mortality

[16]For comparison, the IV estimates in Becker and Woessmann (2009) suggest that an all-Protestant county in Prussia has 29.6% higher income tax revenues than an all-Catholic county.

maps into fertility.[17]

Second, we examine whether the Protestant fertility decline reflects other spatial or institutional channels. Appendix Figure A.9 asks whether parishes at religious borders follow different fertility trajectories. The estimates are informative but do not line up with a clean alternative explanation for the Protestant effect: border effects are imprecise and strongest early in the sample, while the main Protestant-Catholic fertility gap opens later. Appendix Figure A.11 considers distance to the nearest monastery as a proxy for the local historical presence of Catholic institutions. The estimates are mostly small and imprecise once county fixed effects and controls are included. Finally, Appendix Figure A.12 studies inheritance regimes. Closed inheritance is related to fertility, marriage, and out-of-wedlock births in some periods, but the pattern is noisy and the interaction with Protestantism is not stable over time. Taken together, these appendix results make it unlikely that marriage-market behavior, infant mortality, borders, monastic proximity, or inheritance systems are the primary reason for the earlier Protestant transition.

A related concern is human capital. Becker and Cinnirella (2020) show that the Protestant-Catholic fertility gap in Prussia disappears once local literacy is controlled, which raises the possibility that our denominational effect reflects education rather than denomination itself. We do not observe parish-level literacy or schooling, so we cannot control for it directly. Two features of our setting limit this concern. First, by our sample period compulsory primary schooling applied throughout the Kingdom of Württemberg to Protestant and Catholic children alike (Katein, 1956, 61–73), and contemporary accounts suggest that illiteracy in Württemberg was already rare in the nineteenth century (Sauer, 1997, 386). The cross-local variation in literacy that drives the Prussian result was therefore likely much weaker in Württemberg.

[17] A replacement mechanism could in principle still contribute through levels rather than slopes, if infant mortality fell earlier in Protestant parishes. Figure A.8 does not test this directly, and our within-county, neighborhood, and enclave comparisons remove only the infant-mortality variation that is common to nearby parishes. We therefore treat infant mortality as an unlikely primary explanation for the earlier Protestant transition, rather than as a channel we can exclude outright.

Second, our within-county, neighborhood, and enclave comparisons difference out any spatially smooth education gradient, exactly as they difference out geography; a residual literacy gap would have to open up between immediately adjacent parishes of different denomination to confound the enclave estimates.

The remaining interpretation is that religious denomination shaped the acceptability and timing of deliberate fertility control. This interpretation is supported by contemporary observers, who already described the fertility decline as a change in sexual norms rather than only as a response to material incentives. Wolf (1912a) explicitly characterized the decline as a rationalization of sexual life and linked lower birth rates to Protestant areas and to groups less bound by traditional church authority; in a related contribution, Wolf (1912b) connected fertility patterns to social-democratic milieus, which contemporaries often read as a marker of secularization. Offenbacher (1900) made a similar argument for neighboring Baden, emphasizing that Protestants had moved away from ecclesiastical traditionalism earlier than Catholics. These accounts are useful because they point to a behavioral margin that need not show up primarily through fewer marriages, higher out-of-wedlock births, or lower infant mortality: couples could remain married, but begin limiting births within marriage once the moral costs of fertility control declined. The earlier Protestant fertility decline appears less consistent with a single economic or institutional channel than with a denomination-specific weakening of religious constraints on family limitation.

1.8 Conclusion

To understand the relative importance of the factors driving the demographic transition, and in particular the fertility decline, we study the southwest German Kingdom of Württemberg. Our results provide the first parish-level evidence on the timing of the fertility decline in Germany, covering the universe of 1,764 localities. Using the average crude birth rate of the respective county in 1812–1866 as the pre-transition baseline, the mean fertility transition year is 1906; using the alternative parish-level baseline (1871 and 1878–1881), it is 1901. The standard deviation of transition dates is large in both cases, at 11.5 and 12.6 years. The heterogeneity within Württemberg is thus comparable to the dispersion documented across European provinces (Coale and Watkins, 1986)–a fact that more aggregated data would conceal.

We then document pronounced timing differences between Protestant and Catholic parishes. Protestant parishes experienced the onset of the fertility decline about a decade earlier than Catholic parishes. To enable a causal interpretation, we rely on the assumption that the determination of a parish's denomination by local rulers in the 16th century is arguably exogenous to fertility behavior centuries later. Balancing tests show no systematic differences in observable predetermined characteristics once we restrict the comparison to parishes within the same county, within a 10 km neighborhood, or to religious enclaves and their immediate neighbors. Reassuringly, the effect is largest where the comparison is most restricted: religious enclaves, surrounded since the 16th century by parishes of the other denomination, experience the transition roughly 10 years earlier than their neighbors, compared with between 7 and 8 years in the county and neighborhood designs.

Using these same three settings, we do not find consistent differences between Protestant and Catholic parishes in industrial or non-agricultural employment, firm size, the adoption of steam engines, or the gender wage gap. The one robust positive association outside the enclave specification is taxable income per capita, which is 4.4% to 7.5% higher in Protestant parishes relative to the Catholic control mean. This effect is small, however, and it is not accompanied by differences in the other markers of modernization; it also turns insignificant in the enclave comparison. We therefore consider it unlikely that the earlier Protestant decline operated through income or broader economic modernization—though the income result is of interest in its own right for the ongoing debate on the Weber hypothesis (Becker and Woessmann, 2009; Kersting et al., 2020).

We also examine a range of demographic and institutional channels, and none provides a compelling alternative. Marriage rates evolved similarly across denominations and were, if anything, slightly higher in Protestant parishes; out-of-wedlock births, infant mortality, proximity to religious borders, distance to the nearest monastery, and inheritance regimes likewise fail to explain the timing gap in any systematic way. Because Protestant and Catholic parishes were under the same administrative rule and experienced no discriminatory changes during the transition, differences in marriage law or other legal institutions cannot account for the gap either.

What remains is the interpretation that denomination shaped the acceptability, and hence the timing, of deliberate fertility control within marriage. Couples could remain married yet begin to limit births once the moral cost of doing so declined, a margin that doesn't need to show up as fewer marriages, more out-of-wedlock births, or lower infant mortality. This reading is most directly echoed by Wolf (1912a), who described the birth decline as a rationalization of sexual life and linked lower fertility to Protestant areas and to social-democratic, less tradition-bound milieus. Offenbacher (1900)'s analysis of neighboring Baden points to the same broader confessional mechanism: Protestantism was associated with a weaker attachment to inherited ecclesiastical traditionalism and with greater openness to modern economic and cultural life. We emphasize that we do not measure secularization directly; our case for it rests on the elimination of competing economic, demographic, and institutional channels, together with these contemporary accounts.

Overall, our design holds local economic and geographic conditions roughly fixed and still recovers a denominational difference in the timing of the fertility decline. This speaks to the long-standing debate over economic versus cultural drivers of the transition (Galor, 2012; Doepke et al., 2023), and complements recent work emphasizing culture and secularization (Spolaore and Wacziarg, 2022; Beach and Hanlon, 2023b; Blanc, 2023) by isolating the role of denomination within a common institutional setting.

Chapter 2

From Weimar to West Germany: Confessional Cleavages, Party Fragmentation, and Far-Right Voting[1]

2.1 Introduction

Consider two neighboring parishes in southwest Germany at the end of the Weimar Republic: the Catholic parish of Dielheim and the Protestant parish of Tairnbach were separated by only a few kilometers and a confessional boundary dating back to the Reformation. In the last free election of the republic in November 1932, Dielheim delivered 75% of its vote to the Catholic *Zentrum* (Center Party). By contrast, Tairnbach had no equivalent party to anchor its Protestant center-right electorate. After years of fragmented bourgeois voting, 87% of Tairnbach's voters

[1]This paper is currently circulating under the same title, under Kiel Working Paper Series, No. 2324, June 2026. It is a joint work with Sebastian Braun. Both authors contributed equally. Sarah Schaller and Barbara Schneider provided excellent research assistance. We would like to thank seminar and conference participants at the University of Bayreuth, IZA, the University of Mannheim, the 2025 Congress for Economic and Social History in Berlin, the 2025 World Economic History Congress in Lund and the 2025 European Historical Economics Society Conference in Hohenheim for their helpful comments and suggestions. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project Number 471335227 is gratefully acknowledged. All remaining errors are my own.

flocked to the *Nationalsozialistische Deutsche Arbeiterpartei* (NSDAP). Thirty years later, the landscape had fundamentally changed: the cross-confessional *Christlich Demokratische Union Deutschlands* (CDU) commanded the greatest share of the vote in both parishes, though it remained stronger in Catholic Dielheim.

This paper asks whether this local contrast reflects a broader pattern across south-western Germany and to what extent confessional fragmentation in voting persisted after 1945. It thus speaks to a central question in comparative politics and the study of party systems: how do social cleavages shape party-system fragmentation, and how does this effect depend on whether parties mobilize voters within or across these cleavages?

In the cleavage tradition of Lipset and Rokkan (1967), social divisions become politically consequential through their organizational expression (Bartolini and Mair, 1990). Religion is a classic case: confessional life embeds voters in dense associational milieus (Lepsius, 1966) and exposes them to political influence by the clergy (Spenkuch and Tillmann, 2018). While there is much evidence showing that NSDAP support was considerably weaker in Catholic areas (e.g. Falter, 2020), existing work focuses on who voted for Hitler and mostly stops in 1933. It therefore leaves open how confession shaped party-system fragmentation more broadly and whether the confessional structuring of electoral competition persisted into the postwar era.

We address these questions using the dense confessional mosaic of Baden and Württemberg. We first quantify how religious cleavages affected political fragmentation and the Nazi vote across parishes in the Weimar Republic, and assess to what extent the CDU's emergence as a cross-confessional party weakened the link between religion, political fragmentation, and support for the far right after the war. We then use survey evidence to distinguish competing explanations for the confessional gap in voting behavior that remained after 1945. In doing so, we provide new parish-level evidence on how the religious cleavage shaped party-system fragmentation—not only far-right voting—during both the collapse of Weimar and the consolidation of the postwar party system.

Denominational divisions created an asymmetric party landscape on the center-right in the Weimar Republic. Catholics formed a cohesive bloc around the Zentrum, sustained by a strong voting norm rooted in a distinct Catholic milieu and the organizational power of the Church (e.g. Burnham, 1972; Falter, 2020). Protestants, by contrast, lacked any equivalent organizing force on the center-right. Their vote was dispersed across the liberal *Deutsche Demokratische Partei* (DDP), the national-liberal *Deutsche Volkspartei* (DVP), the conservative-nationalist *Deutschnationale Volkspartei* (DNVP), and a shifting array of smaller bourgeois parties.

This asymmetry implied, as we show in a simple voting model and confirm in our data, that fragmentation followed an inverted U-shape in the local share of Catholics. It was highest where a significant Catholic minority coexisted with a Protestant majority, as in the German Reich as a whole. The cleavages between Catholics and Protestants thus decisively contributed to the fragmentation of Weimar's political system, which arguably made lasting coalitions difficult and undermined the stability of the system (Falter, 2013; Weir and Greenberg, 2022). When voter preferences radicalized under the weight of the Great Depression, the fragmented Protestant vote also provided fertile ground for Nazi mobilization. Catholic voters, in contrast, were largely immunized against the NSDAP through their voting norm for the Catholic Zentrum (Spenkuch and Tillmann, 2018; Falter, 2020).

The founding of the CDU as a cross-confessional *Volkspartei* (catch-all party) after World War II offered a potential solution to this problem. A central motivation was to overcome the denominational fragmentation on the center-right by uniting Catholic and Protestant voters under one party (e.g. Rudzio, 2019; Beckmann, 2022). Whether and to what extent the CDU succeeded in breaking the pattern we document for Weimar, and what explains the persistence of confessional differences when the far-right *Nationaldemokratische Partei Deutschlands* (NPD) rose in the late 1960s, are central empirical questions of this paper.

To address these questions and trace the evolution of party fragmentation from the Weimar Republic to West Germany, we assemble new data from over 3,000 parishes in Baden and Württemberg from 1928 to 1969. Parishes are the smallest administrative units in Germany. Existing research on Weimar electoral politics and the rise of the Nazi party has relied primarily on county-level data (e.g. Spenkuch and Tillmann, 2018; De Juan et al., 2024)[2] or data from larger cities (e.g. Galofré-Vilà et al., 2021; Doerr et al., 2022). Our parish-level dataset increases the number of observations by roughly a factor of thirty relative to county-level data for southwestern Germany. This level of granularity enables fine-grained comparisons between neighboring parishes that differ in their majority confession, providing new leverage for identifying the relationship between religious composition, party fragmentation, and far-right voting.

Our empirical analysis exploits the confessional patchwork that the Peace of Augsburg (1555) imprinted on southwestern Germany. The region was fragmented into hundreds of small territories during the Middle Ages. These territories adopted the religion of their ruler under the principle of *cuius regio, eius religio*. This produced sharp Catholic–Protestant boundaries between adjacent localities that persisted into the 20th century. We leverage this historical variation by comparing each parish to its neighboring parishes, thereby holding constant local geography and other neighborhood characteristics while isolating the effect of differences in religious composition.

We find that, in line with our predictions, political fragmentation in the Weimar period exhibited an inverted U-shaped relationship with the Catholic population share. Moreover, the Catholic population share alone explains more than half of the parish-level variation in the NSDAP vote in November 1932, with Nazi support

[2] Falter and Hänisch (1990a) compiled the election database underlying many of these studies. The database also offers parish-level data, but only for parishes with 2,000 or more inhabitants. Election returns for all parishes, including smaller rural localities, do not exist for Weimar Germany as a whole.

markedly lower in more Catholic parishes. In the postwar period, the fragmentation hump gradually weakens. Fragmentation declined most in confessionally mixed and Protestant parishes, as the CDU increasingly attracted Protestant voters. Nevertheless, Catholics continued to vote disproportionately for the CDU throughout the period, while Protestant areas remained more susceptible to the far-right NPD in the late 1960s.

To disentangle the mechanism behind the shrinking but persistent confessional gap in voting, we turn to individual-level survey evidence from the 1968 Baden-Württemberg state election. Protestants and Catholics held virtually identical views on the political questions most associated with NPD support—nationalism, anti-elitism, and resentment of Allied influence. The key difference lay instead in church attachment and the church's perceived political stance. The Catholic Church sent a clear anti-NPD, pro-CDU signal; the Protestant Church's positioning was perceived as more ambiguous. Among devout respondents, NPD support was low in both confessions. The aggregate confessional gap thus reflects both the clarity of the Catholic Church's signal and Catholics' stronger church attachment. Overall, our findings locate the religious cleavage primarily in its organizational dimension (Bartolini and Mair, 1990): confessional identities shaped electoral competition through their mobilization by parties and churches, not simply through differences in political attitudes.

Related literature. Our paper contributes to four strands of research on religious cleavages, party-system fragmentation, and far-right voting. A first literature studies how social divisions structure party systems in Western Europe (see Ford and Jennings, 2020, for a recent review). The classic account by Lipset and Rokkan (1967) identified religious cleavage as one of the most important social divisions shaping the formation and fragmentation of European party systems. Although secularization weakened religious voting across much of Europe, religious identities and church attachment continue to shape party choices (e.g. Elff, 2007; Goldberg, 2020; Gomez, 2022), including in Germany (Elff and Rossteutscher, 2011). We build on this literature by asking how the religious cleavage between Catholics and Protestants shaped not only individual party choice, but also local party-system fragmentation from Weimar to West Germany.[3]

A second literature studies the effect of religion on the electoral success of the NSDAP. The confessional divide is the strongest predictor of Nazi voting: Catholic areas were less receptive to the NSDAP than Protestant areas, even after accounting for economic distress, social structure, war casualties, and media exposure (King et al., 2008; Adena et al., 2015; Falter, 2020; Galofré-Vilà et al., 2021; Koenig, 2023; De Juan et al., 2024). Spenkuch and Tillmann (2018) provide the most direct quantitative evidence, showing that Catholic voters were insulated from the NSDAP because Catholic priests discouraged Nazi support. Moreover, Becker and Voth (2023) demonstrate that deep-rooted Christian religiosity, more prevalent in Catholic areas, made individuals less susceptible to Nazi propaganda. We revisit the evidence on religion and Nazi voting with parish-level data. This allows us to compare neighboring Catholic and Protestant localities and to connect the confessional gap in voting behavior to the fragmentation of the center-right.

Third, our paper contributes to the study of Christian democracy and the stabilization of the postwar party system. The CDU presented itself as a cross-confessional *Volkspartei*, uniting former Zentrum Catholics with Protestant conservatives and liberals instead of recreating the confessional party system of the Weimar Republic (Kalyvas, 1996; Mitchell, 2012; Rudzio, 2019; Beckmann, 2022). This cross-confessional strategy was designed to overcome the fragmentation of the center-right in the Weimar era. We provide quantitative evidence on whether it succeeded. Specifically, we demonstrate that the relationship between religious composition and political fragmentation weakened over time as the CDU increasingly attracted Protestant voters, thereby reducing electoral divisions along confessional lines.

[3] The empirical literature on the determinants of party-system fragmentation has focused primarily on ethnic cleavages (e.g. Ordeshook and Shvetsova, 1994; Mozaffar et al., 2003; Lublin, 2017), with little attention to religious cleavages (Raymond, 2016).

Finally, we contribute to research on confessional differences in far-right voting. When the NPD rose in the late 1960s, it attracted disproportionate support in Protestant areas (Klingemann, 1971), echoing the Weimar pattern. More recent studies likewise document confessional divides in support for the far right in Germany (Arzheimer and Berning, 2019; Haffert, 2022). The sources of these differences, however, remain contested. They may reflect underlying political preferences, the influence of religious milieus (Haffert, 2022), or church leaders' efforts to steer voters toward some parties and away from others (Spenkuch and Tillmann, 2018). Using individual-level survey data, we show that the confessional gap in NPD voting was driven primarily by the perceived clarity of church signals rather than differences in underlying preferences. Moreover, strong church attachment reduced NPD support, mirroring contemporary evidence for Western Europe (e.g. Immerzeel et al., 2013; Marcinkiewicz and Dassonneville, 2022).

2.2 Conceptual Framework

We adapt the logic of Spenkuch and Tillmann (2018) to illustrate our hypotheses on the impact of confessional composition on political fragmentation and far-right voting. We first explore local[4] voting patterns in the Weimar Republic, then turn to postwar West Germany.

2.2.1 Weimar Republic

Suppose that voters in the Weimar Republic could choose between the following six parties, ordered from left to right on the political spectrum: *Kommunistische Partei Deutschlands* (KPD), *Sozialdemokratische Partei Deutschlands* (SPD), DDP, Zentrum (Z), DNVP, and NSDAP. The five non-Nazi parties are often grouped into three blocs,

[4]Of course, the fragmentation of the Weimar parliament ultimately depends on Germany-wide vote shares. However, in Germany's proportional voting system, national-level vote shares are the sum of local votes, so local voting patterns are the building blocks of national vote shares (and fragmentation).

the left/Marxist (KPD, SPD), the Catholic (Z), and the Protestant bourgeois (DDP, DNVP) camps (Burnham, 1972; Shively, 1972).

A voter i of denomination $r \in \{c, p\}$ votes for the party X that maximizes her utility U_i^r, which depends (negatively) on the deviation of the party's ideological position from her own position t_i. Catholics derive additional utility from voting Z, reflecting a positive political voting norm grounded in a distinct Catholic culture and milieu (e.g. Burnham, 1972; Falter, 2020) or the directives of the clergy (Spenkuch and Tillmann, 2018). Protestants, on the other hand, lose utility by voting for Z as the party of Catholics. Utility can thus be expressed as

$$U_i^r(X) = \begin{cases} -(X - t_i)^2 + \lambda^r & \text{for } X = Z, \\ -(X - t_i)^2 & \text{otherwise,} \end{cases} \tag{2.1}$$

where $\lambda^c > 0$ and $\lambda^p < 0$ reflect the utility gain or loss from voting Zentrum as a Catholic or Protestant, respectively. We assume that Catholics and Protestants have the same distribution of preferences. This assumption deliberately rules out preference differences as the primitive source of confessional voting gaps: any difference in party choice arises from the voting norm, not from Catholics and Protestants occupying different ideological positions.[5]

We can integrate over voter preferences to calculate party vote shares for an all-Catholic electorate and an all-Protestant electorate. We use these vote shares to calculate the effective number of electoral parties N^r as a standard measure of political fragmentation (Laakso and Taagepera, 1979):

$$N^r = \frac{1}{\sum_{i=1}^{n}(p_i^r)^2} \text{ for } r \in \{c, p\}, \tag{2.2}$$

where n is the number of parties and p_i^r is the vote share of party i among voters of denomination r. If we are willing to assume that the voting behavior of Protestants

[5]The individual-level evidence we present for 1968 is consistent with this assumption in the postwar period, showing similar nationalist and anti-establishment views across confessions (see Section 2.6).

and Catholics does not depend on their local share of the population, we can also calculate the fragmentation index for denominationally mixed parishes:

$$N(share_c) = \frac{1}{\sum_{i=1}^{n}(share_c \cdot p_i^c + share_p \cdot p_i^p)^2}, \tag{2.3}$$

where $share_c$ and $share_p$ are the shares of Catholics and Protestants in the local population (we set $share_p = 1 - share_c$ for simplicity).

1928 election. We first consider the framework's predictions for a moderate electorate, as in the 1928 Reichstag election. Panel (a) of Figure 2.1 shows the utility-maximizing party choice for voters of different ideological types, separately for Catholic and Protestant voters.[6] Because of the positive political voting norm, Catholic center-right voters vote Zentrum. The Protestant center-right vote, on the other hand, is split between the more liberal DDP and the conservative DNVP. The Zentrum also draws votes from Catholic voters positioned on the moderate left and far right who, without the positive voting norm, would have voted SPD or NSDAP.

Figure 2.2 shows the key novel aspect of our analysis, the predicted fragmentation by local denominational composition; the red line refers to the 1928 election. The effective number of parties follows an inverted U-shape in the proportion of Catholics. An all-Catholic electorate is the least fragmented, given the dominance of the Zentrum. Protestant electorates are much more fragmented. Electorates with a significant Catholic minority are the most fragmented, reflecting the presence of two blocs on the center-right—a Catholic and a fragmented Protestant one.

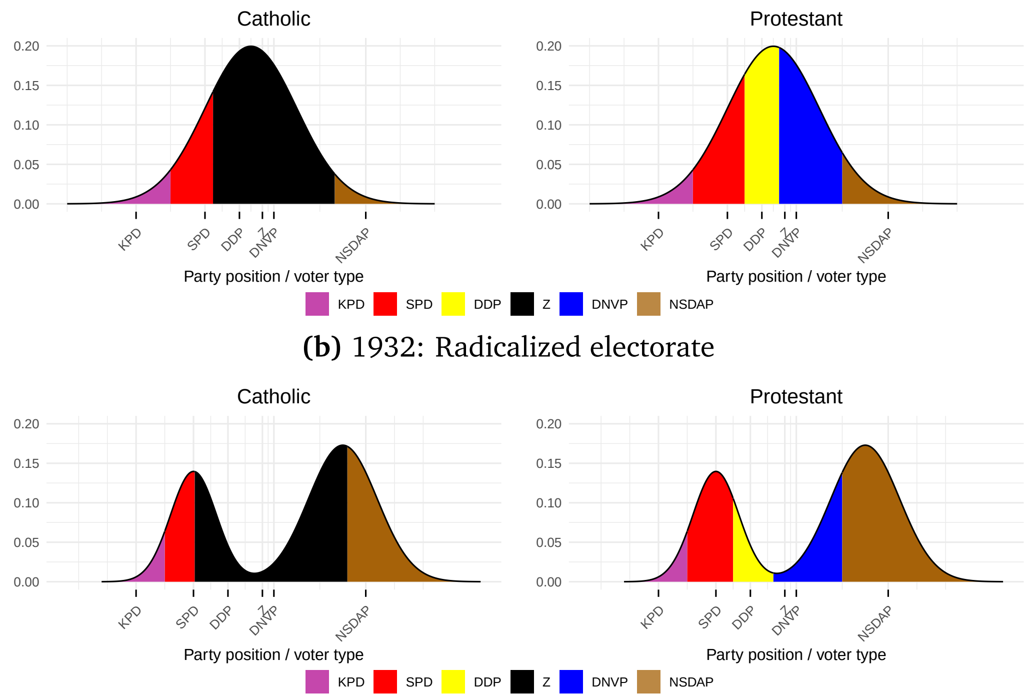
Appendix Figure B.1 shows the party vote shares underlying Figure 2.2 for an all-Catholic electorate, an all-Protestant electorate, and a mixed electorate that is one-third Catholic and two-thirds Protestant (roughly the share for the *Reich* as a whole). While the Catholic (Z) and Protestant bourgeois (DDP, DNVP) camps have clear

[6]The figure is for an electorate that is moderately center-right, as is likely in our mostly rural parishes.

Figure 2.1: Confessional voting in the Weimar Republic: Predictions

(a) 1928: Moderate electorate

(b) 1932: Radicalized electorate



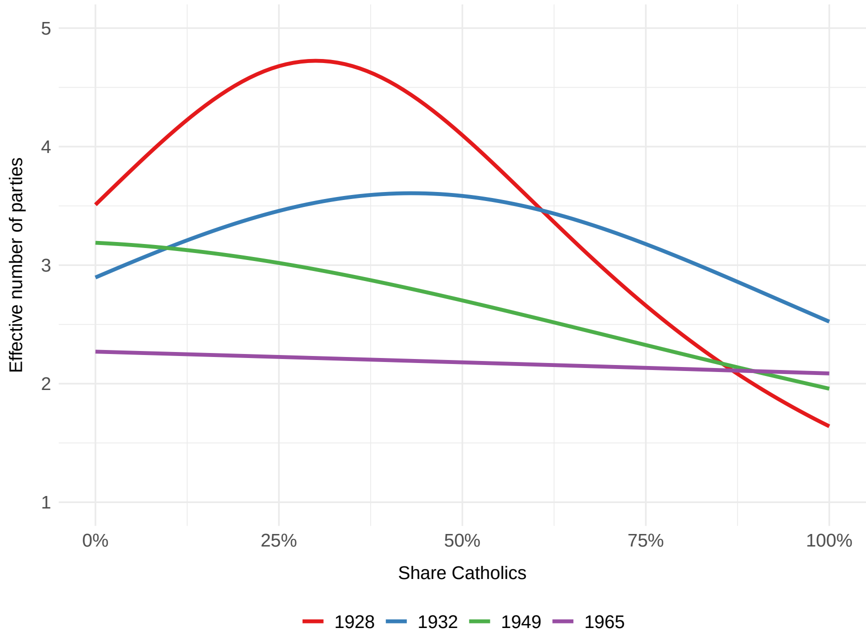
Notes: The figure sketches the predictions of our conceptual framework regarding the voting preferences of the electorate by voter type. We distinguish between Catholic and Protestant voters. Panel (a) illustrates the predictions for a moderate electorate, as in the 1928 election, and panel (b) illustrates the predictions for a radicalized electorate, as in the 1932 election.

majorities in their respective electorates, neither camp necessarily wins a majority in confessionally mixed settings.[7] Catholic-Protestant antagonism thus increased fragmentation and made stable coalition-building more difficult.

1932 election. Next, we consider the November 1932 Reichstag election, which was the last before the Nazis seized power in January 1933. As in Spenkuch and Tillmann (2018), we assume that in response to the Great Depression and the perceived inability of the political center to deal with the crisis, voter preferences bifurcate, with the extremes on both sides gaining traction relative to 1928. As long as their preferences are not too far to the right, Catholic voters are insulated from the NSDAP by their voting norm for the Zentrum. While the NSDAP makes large gains among both Catholics and Protestants, the gains are much larger among Protestants. The NSDAP vote share in 1932 is therefore decreasing in the local share of Catholics. At

[7]In more urban settings, of course, potential majorities for the left/Marxist camp would also apply.

Figure 2.2: Fragmentation by local denominational composition: Model Predictions



Notes: The figure shows the effective number of political parties as a function of the local share of Catholics (see equation (2.3)), as predicted by our conceptual framework. We distinguish between the predictions for the 1928, 1932, 1949, and 1965 elections.

the same time, the rise of the NSDAP reduces political fragmentation in Protestant parishes by consolidating most votes on the right of the political spectrum. In Catholic areas, by contrast, fragmentation increases as the electorate moves to the right.

Summary. We summarize the predictions of our conceptual framework for the Weimar Republic in Hypotheses 1a and 1b below, and test them empirically in Section 2.4.

Hypothesis 1a ***Fragmentation***—*Political fragmentation in the 1928 Weimar election follows an inverted U-shape in the proportion of Catholics: lowest in all-Catholic parishes and highest in parishes with a significant Catholic minority. Between 1928 and 1932, fragmentation decreases in all-Protestant parishes and increases in all-Catholic parishes.*

Hypothesis 1b ***Nazi Vote***—*Parishes with a higher proportion of Catholics show a smaller increase in the NSDAP vote between 1928 and 1932 and a lower NSDAP vote in 1932.*

2.2.2 Federal Republic of Germany

The most important innovation in the postwar political landscape was the founding of the CDU as a cross-confessional *Volkspartei*. Unlike the Zentrum, the CDU sought to integrate Catholic and Protestant voters under a common Christian label. One core motivation was to overcome the political fragmentation of Weimar on the center-right—and thus the political instability that had characterized the first German democracy (e.g. Rudzio, 2019; Beckmann, 2022). In the 1950s and 1960s, the CDU further consolidated the center-right by absorbing supporters of smaller right-wing parties such as the *Deutsche Partei* (DP).

We capture this change on the supply side by applying the same voting framework to the postwar party system. Voters again choose among parties ordered from left to right. In the early Federal Republic, the relevant choice set consists of KPD, SPD, *Freie Demokratische Partei* (FDP), CDU, and smaller right-wing parties such as the DP. The CDU replaces the Zentrum as the main center-right anchor, but with a crucial difference: it seeks to bind Catholic and Protestant voters within the same party. Over time, the KPD disappeared after its ban in 1956, while the DP lost electoral relevance as the CDU consolidated the center-right. For the 1968 state and 1969 Bundestag elections, we add the NPD as the relevant party at the right extreme.

The Catholic clergy supported the CDU early on, so that the refounded Zentrum was only a splinter party in West Germany. At the same time, this clerical support opened the CDU to accusations of representing primarily Catholic interests. The Protestant Church avoided party-political positioning and engaged openly with the SPD (Besier, 1990).

We therefore let λ^r denote the voting norm for the CDU among voters of denomination r in the postwar period. Whereas $\lambda^p < 0$ in the Weimar Republic captured Protestants' aversion to the Catholic Zentrum, we assume $\lambda^p \geq 0$ after 1949, reflecting the CDU's cross-confessional appeal. In the early postwar period, however, λ^c was still larger than λ^p, reflecting the Catholic clergy's stronger initial endorsement of the CDU.[8] Chancellor Adenauer emphasized the CDU's non-denominational character by balancing Protestants and Catholics across his cabinet ministries from the second cabinet onward. We model the CDU's increasing credibility as a cross-confessional party by assuming that λ^p approaches λ^c over time.

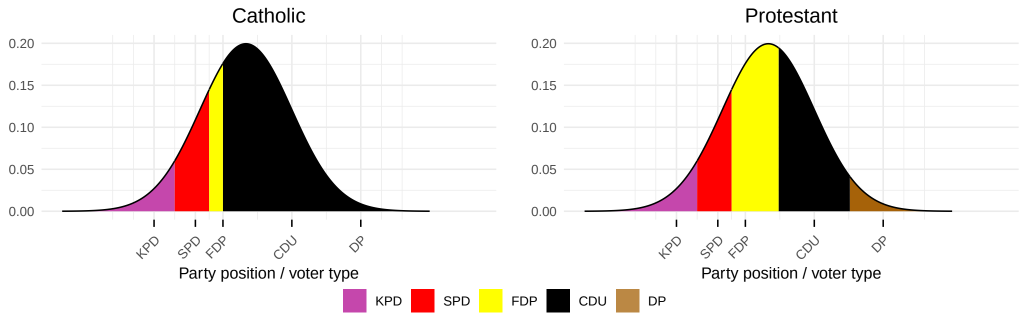
1949 and 1965 elections. Panels (a) and (b) of Figure 2.3 show the party choice of voters of different ideological types in the 1949 and 1965 Bundestag elections, separately for Catholic and Protestant voters. The green and purple lines in Figure 2.2 show the respective fragmentation index for the two elections. The key prediction is that the emergence of the CDU significantly reduced political fragmentation on the center-right. The decrease in fragmentation from the prewar to the postwar period should be strongest in Protestant and mixed parishes. The fragmentation hump should therefore flatten in the 1950s and 1960s as the CDU gains acceptance as a cross-confessional party.

1968 and 1969 elections. In the 1968 Baden-Württemberg state election, the NPD received 9.8% of the vote, its best result ever. The NPD's rise is typically attributed to Germany-wide factors, such as the first postwar recession in the mid-1960s, the formation of the Grand Coalition between the CDU and SPD in November 1966, and the onset of student protests. We use this election to test whether a strong norm of voting CDU had taken hold also among Protestant voters by the late 1960s. We again assume that voter preferences bifurcated in response to these crises. In the model, the center-right norm of voting CDU limits the NPD's rise. If the norm remained stronger among Catholics, the model still predicts lower NPD support in Catholic parishes (see panel (c) of Figure 2.3). The same reasoning applies to the 1969 Bundestag election, which forms the basis of our parish-level analysis in Section 2.5.

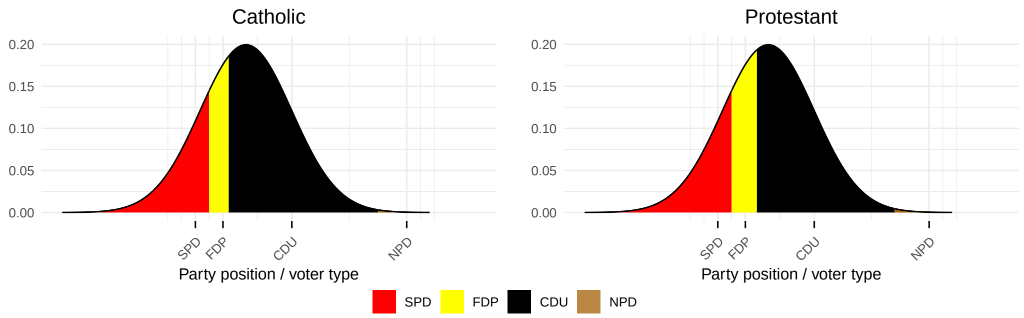
[8] Survey evidence from 1968 is consistent with this assumption: respondents of both confessions perceived their church as favorably disposed toward the CDU, but Catholics perceived a stronger pro-CDU signal than Protestants (see Figure 2.10, Appendix Figure B.7, and Section 2.6).

Figure 2.3: Confessional voting in the Federal Republic of Germany: Predictions

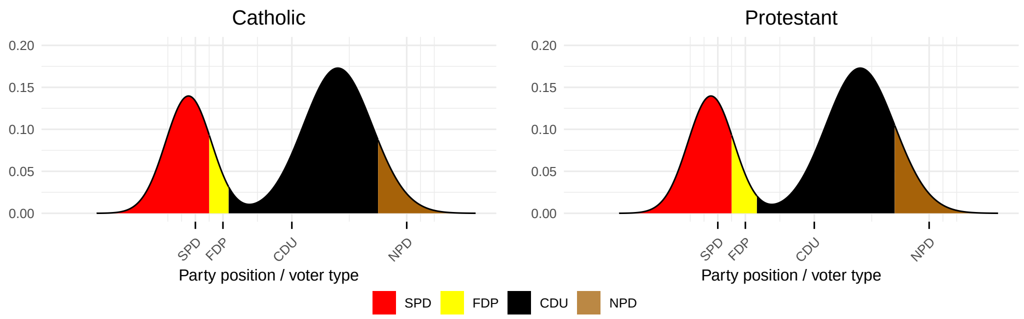
(a) 1949: Moderate electorate, $\lambda^c \gg \lambda^p > 0$



(b) 1965: Moderate electorate, $\lambda^c \geq \lambda^p > 0$



(c) 1968/69: Radicalized electorate, $\lambda^c \geq \lambda^p > 0$



Notes: The figure sketches the predictions of our conceptual framework regarding the voting preferences of the electorate by voter type. We distinguish between Catholic and Protestant voters. Panels (a) and (b) illustrate the predictions for a moderate electorate, as in the 1949 and 1965 elections, and panel (c) illustrates the predictions for a radicalized electorate, as in the 1968 state and 1969 Bundestag elections.

Summary. We summarize the predictions of our conceptual framework for postwar West Germany in Hypotheses 2a and 2b below, and test them empirically in Section 2.5.

Hypothesis 2a ***Fragmentation***—*Relative to the prewar period, political fragmentation decreases mainly in Protestant and denominationally mixed parishes. The fragmentation hump in the share of Catholics becomes increasingly flat in the 1950s and 1960s.*

Hypothesis 2b ***NPD Vote***—*Parishes with a higher proportion of Catholics show a lower NPD vote in the late 1960s. However, the confessional gap in NPD support is substantially smaller than the corresponding gap in Nazi support in 1932.*

2.3 Data and Empirical Strategy

Data. Our data cover all civil parishes (*Gemeinden*) in Baden and Württemberg. We use administrative borders as of 1939, merging some parishes to maintain consistency over 1928–1969. The resulting dataset covers 3,180 parishes. The median parish area is 8.56 square kilometers, and the median 1925 population size was 662. Most parishes were religiously homogeneous: 82 percent had more than 90 percent adherence to a single denomination.

Electoral data on federal and state elections come from several sources. For the Weimar period, our primary source is Falter and Hänisch (1990b), who provide parish-level data for Reichstag elections in Baden between 1924 and 1933. Following Spenkuch and Tillmann (2018), we focus on the November 1932 election. Comparable parish-level data for Württemberg are available only for the state election of 24 April 1932 (Württembergischen Statistischen Landesamt, 1932). Consequently, our prewar analysis focuses primarily on Baden. For the postwar period, we assemble parish-level data for Baden-Württemberg from the state's online archive, and cross-check it against published volumes (Landesarchiv Baden-Württemberg, 2026).

Baden-Württemberg was created in 1952 through the merger of the former states of Baden and Württemberg with the small Prussian province of Hohenzollern.

Religious composition, our main explanatory variable, and socioeconomic controls come from population and occupation censuses. These censuses provide information on population size, the female population share, and the sectoral composition of the workforce. For Baden, the prewar data come from the 1925 census (Falter and Hänisch, 1990b). For Württemberg, data are drawn from the 1933 census (Württembergischen Statistischen Landesamt, 1935), except for religious composition, which is available from the 1925 census. The 1933 Württemberg census also reports unemployment rates, which are unavailable for Baden. For the postwar period, we use the 1950, 1961, and 1970 censuses for Baden-Württemberg (Landesarchiv Baden-Württemberg, 2026). We also construct geographical controls, including soil type, river access, and terrain ruggedness.

We supplement the parish-level data with individual-level survey data from the 1968 Baden-Württemberg state election (Klingemann, 2012). The post-election survey sampled 895 respondents representative of the electorate in Baden-Württemberg. The survey contains information on respondents' voting behavior and focuses on questions related to the rise of the right-wing NPD, including respondents' attitudes toward the NPD and toward nationalism. Crucially, the survey also contains information on respondents' religious affiliation and church attachment, as well as their perceptions of the churches' party-political positions.

Appendix B.1 gives an overview of all sources used.

Specifications. Let y_{it} denote a political outcome of interest of parish i in year t, including the NSDAP vote share in Weimar Germany, the NPD vote share in West Germany, and measures of political fragmentation in both German states. Vote shares are measured as percentages of valid votes, and political fragmentation is measured by the effective number of electoral parties (see equation (2.2)). To assess the relationship between Catholic population share and these outcomes, we estimate OLS regression models of the following type:

$$y_{it} = \alpha + \beta CatholicShare_{it} + f(lat_i, lon_i) + \mathbf{X}_{it}\gamma + \epsilon_{it}. \tag{2.4}$$

Here, $CatholicShare_{it}$ refers to the percent share of Catholics in the population, $f(lat_i, lon_i)$ is a cubic polynomial in longitude and latitude to capture spatial trends, $\mathbf{X}_{it}$ is a (row) vector of control variables, and ϵ_{it} is an error term. For Weimar outcomes, we measure religious composition in 1925; for postwar outcomes, we use religious composition in 1950.

The parameter of interest is β. It indicates the (percentage) point change in vote shares and fragmentation associated with a one-percentage point increase in a parish's share of Catholics, holding other factors constant. For fragmentation outcomes, where our conceptual framework predicts an inverted U-shape, we also include quadratic and cubic terms in $CatholicShare_{it}$. Control variables include the local share of Jews, of women, of workers, population size (in logs), the employment share in industry, as well as geographical controls for soil types, elevation, and area, in line with the literature.

We address remaining concerns that unobserved variables are driving the relationship between the religious composition of parishes and political outcomes in three ways: first, by using only within-county variation; second, by demeaning the data with respect to a close neighborhood of parishes; and third, by comparing religious enclaves with their neighboring communities. As a robustness check, we also estimate IV regressions using parishes' religious confessions in 1619 as an instrumental variable, following Spenkuch and Tillmann (2018).

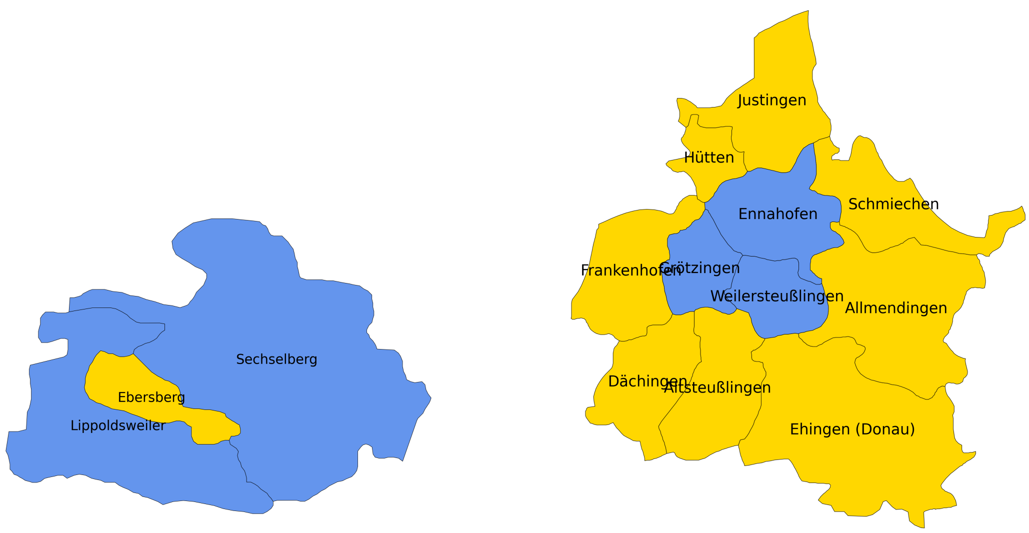
Within-county variation. Our first identification strategy uses only within-county variation, adding county-fixed effects to equation (2.4). At the time, Baden and Württemberg had 103 counties. With a median size of 322 square kilometers, these counties are comparatively small administrative units that serve as the unit of analysis in many papers on the Nazi Party's electoral success (e.g. Spenkuch and Tillmann, 2018; Falter, 2020; Becker and Voth, 2023). By comparing parishes within these counties, we exploit variation only in the religious composition of nearby parishes, which should differ little in observable–and, by extension, unobservable–characteristics. We provide evidence in line with this argument below.

Within-neighborhood variation. Our second identification strategy transforms equation (2.4) by subtracting the respective mean of all observations within a 10 km radius of the centroid of each parish from the dependent and independent variables. We choose this radius to ensure that each parish has at least one neighbor within it. This approach is similar to our first identification strategy; however, instead of comparing parishes within administrative boundaries, we compare parishes within geographic neighborhoods. Both identification strategies have also been proposed to deal with strong spatial autocorrelation (Becker et al., 2025; Müller and Watson, 2024).

Religious enclaves. Our third strategy relies on "historical accidents" that created several religious enclaves in Baden and Württemberg. We define a religious enclave as one or more parishes of one Christian denomination that are surrounded by parishes of the other denomination. Again, the idea behind the identification strategy is that the surrounding communities serve as a suitable control group because they likely differ from the enclaves only in their Christian denomination. Empirically, we restrict the sample to enclave parishes and their surrounding communities, include fixed effects for each enclave case, and add a control for enclave parishes. There are a total of 44 enclave cases in Baden and Württemberg, with a total sample size of 450 parishes (meaning the enclaves and their neighbors).

Figure 2.4 displays two examples (see Appendix B.2 for a list of all cases). The left panel shows the Catholic parish of Ebersberg, surrounded by the Protestant villages of Sechselberg and Lippoldsweiler. In 1654, Ebersberg was sold to Jeremias Vollmar Schenk von Winterstetten, who invoked *cuius regio, eius religio* to reintroduce Catholicism, pressuring inhabitants to convert; those unwilling emigrated. The right panel shows the "Lutheran Mountains" (*Lutherische Berge*), consisting of the Protestant communities of Ennahofen, Grötzingen, and Weilersteußlingen. The area takes its name from Duke Ludwig of Württemberg, who introduced the Reformation there in 1581 after the Steußlingen territory reverted to him as a fief, leaving the surrounding villages Catholic.

Figure 2.4: Religious Enclaves in Baden-Württemberg



Notes: The figure shows two examples of enclave cases in our data. Yellow parishes are Catholic; blue parishes are Protestant. See the main text for historical background on the two cases.

Instrumental variables regression. As a robustness check, we follow Spenkuch and Tillmann (2018) and use the religious situation at the beginning of the Thirty Years' War as an instrumental variable for the religion of voters 300 years later. Specifically, we assign each parish the religion of the lord who reigned over the corresponding area in 1619. For the exclusion restriction to hold, the choices of rulers in the 16th and early 17th centuries must be uncorrelated with unobserved factors that determine election outcomes in Weimar and West Germany. Spenkuch and Tillmann (2018, p. 27) argues that, conditional on covariates, "identification in our IV approach comes from idiosyncrasies in rulers' choices." While this is plausible, the instrument strongly predicts the endogenous variable in our case.[9] This strong first stage reduces concerns about weak instruments. At the same time, it implies that the variation isolated by the IV regressions is close to the variation used in our OLS specifications.

Standard errors. We report standard errors clustered at the county level to allow for arbitrary forms of correlation in the residuals of nearby parishes. In our regressions involving enclaves, we cluster standard errors at the case level where a case refers to the enclave and its surrounding control parishes.

Comparison between nearby parishes in predetermined characteristics. Our core identification strategies all exploit variation in religious composition across nearby parishes. We define "nearby" in three ways: being located in the same county, being within a 10-kilometer radius, or being adjacent to a religious enclave. The logic is straightforward. Neighboring parishes should be similar in predetermined characteristics, alleviating concerns that unobserved local factors drive the relationship between a parish's religious composition and its political outcomes.

Table 2.1 supports our empirical strategy. The table documents significant differences in predetermined characteristics between parishes that are more than 90% Catholic (Column 1) or Protestant (Column 2). Longitude and latitude data show that Catholic parishes, on average, are more likely to be located in the south and west of the study area. They are also more likely to be located at higher altitudes, in river valleys, and in hilly areas. Column (3) shows that many of these differences are statistically significant in unconditional regressions of predetermined characteristics on religious makeup in 1925.

Columns (4) through (6) show that our identification strategies level out these differences. Column (4) reports differences between Catholic and Protestant parishes

[9] Conditional on covariates and county-level fixed effects, the share of Catholics is 83 percentage points higher in parishes that were already Catholic in 1619. The residual R2 of the first stage regression is 0.91.

within counties (from regressions of each characteristic on the 1925 share of Catholics and county fixed effects). The differences shrink markedly for all characteristics, though smaller differences in soil type persist. Column (5) applies the same 10 km spatial-difference transformation to each characteristic and to the Catholic share. Again, the differences shrink considerably and mostly become statistically insignificant. Finally, Column (6) restricts the sample to religious enclaves and their surrounding parishes, comparing only neighboring parishes. None of the differences remain statistically significant at the 5% level (although the small sample size may contribute to this result).

Table 2.1: Comparison of Predetermined Characteristics between Protestant and Catholic Parishes

	Averages:		Catholic Coefficient:			
			Un-	Within transformation		
	Catholic	Protestant	conditional	Counties	Neighborhoods	Enclaves
	(1)	(2)	(3)	(4)	(5)	(6)
Longitude	8.911	9.120	-0.002*	0.000	-0.000	0.000
	(0.020)	(0.018)	(0.001)	(0.000)	(0.000)	(0.000)
Latitude	48.315	48.762	-0.005***	-0.000	-0.000*	-0.000
	(0.015)	(0.012)	(0.001)	(0.000)	(0.000)	(0.000)
Average elevation (*in m*)	518.721	424.052	1.057***	0.082	-0.007	-0.008
	(5.783)	(4.907)	(0.303)	(0.146)	(0.042)	(0.043)
River access (%)	8.416	6.093	0.018	0.030	0.025	0.032
	(0.739)	(0.691)	(0.022)	(0.029)	(0.017)	(0.042)
Area (*sq. km*)	10.489	9.738	0.008	0.001	-0.004	0.012
	(0.197)	(0.232)	(0.007)	(0.005)	(0.006)	(0.014)
Soil type (in %):						
Coastal area and bog soils	0.544	0.226	0.004	-0.004	-0.006	-0.007
	(0.123)	(0.097)	(0.003)	(0.004)	(0.004)	(0.007)
Broad river valleys	13.871	4.751	0.089***	0.003	0.002	-0.014
	(0.761)	(0.579)	(0.032)	(0.027)	(0.013)	(0.020)
Undulating lowlands and hilly areas	21.262	0.114	0.234***	0.006	0.007	0.017*
	(1.001)	(0.078)	(0.051)	(0.005)	(0.005)	(0.010)
Loess area	9.774	25.177	-0.174***	-0.077**	-0.051***	0.007
	(0.625)	(1.043)	(0.035)	(0.031)	(0.017)	(0.022)
Mountain and hill soils	54.163	69.446	-0.156**	0.070*	0.045**	-0.008
	(1.215)	(1.138)	(0.063)	(0.036)	(0.018)	(0.020)
Alpine soils	0.311	0.286	0.001	0.003	0.004**	0.005
	(0.052)	(0.051)	(0.001)	(0.002)	(0.002)	(0.004)

Notes: The table reports group mean values of predetermined local characteristics for Catholic parishes (Column (1)) and Protestant parishes (Column (2)), defined as parishes with at least 90% adherence to the respective denomination. Columns (3)–(6) present coefficients from regressions of each characteristic listed in the left column on the local share of Catholics in 1925. All specifications control for the 1925 shares of Jews and other religions. Column (4) adds county fixed effects. Column (5) reports coefficients from regressions using variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. Column (6) restricts the sample to religious enclaves and reports coefficients conditional on fixed effects for each enclave case, where each case consists of an enclave and its surrounding parishes. See the main text and appendix for data sources and further details. Standard deviations of the means are shown in parentheses in Columns (1) and (2). Robust standard errors clustered at the county and enclave-case level are shown in Columns (3)–(5) and (6), respectively. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Note that parishes of different denominations in close proximity need not share similar socioeconomic characteristics, such as sectoral employment structures. Following Max Weber, a large literature examines whether Protestantism affects economic development (see Becker et al., 2016, 2024, for recent literature reviews). Because local socioeconomic characteristics may themselves be outcomes of religious composition, they are, strictly speaking, bad controls. Nevertheless, following the existing literature, our regressions generally condition on socioeconomic parish characteristics. However, we also show that specifications estimated without these potentially endogenous controls produce very similar results.

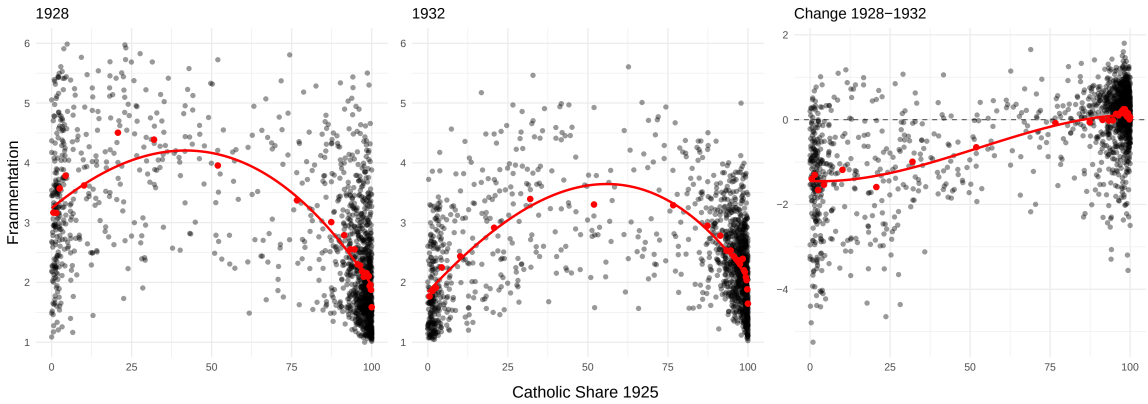
Ecological fallacy. Our estimates are based on ecological regressions at the parish level. Therefore, our results regarding voting outcomes do not have to align with the difference in Nazi support between Protestant and Catholic voters. However, Spenkuch and Tillmann (2018) show that instrumental variable regressions also recover the individual-level difference in Nazi support between Catholic and Protestant voters when the exclusion restriction holds. Furthermore, the ecological inference problem should be small in our context because we mostly compare parishes that are almost entirely Protestant or Catholic. Finally, we also report survey evidence at the individual level regarding the voting behavior of Protestants and Catholics in postwar Germany.

2.4 Results for the Weimar Republic

Fragmentation. Figure 2.5 plots our index of political fragmentation against the 1925 share of Catholics in a parish for the May 1928 and November 1932 Reichstag elections in Baden. A cubic polynomial regression fit is overlaid in red, together with red markers indicating mean values computed within 25 equally sized bins. The panel on the right shows the change in fragmentation between the two elections.

For Protestant parishes, the differences between the two elections are stark and closely align with Hypothesis 1a. In the May 1928 election, Protestant parishes were politically fragmented. In parishes that were 95% or more Protestant, the

Figure 2.5: Share Catholic and political fragmentation in Baden, 5/1928 and 11/1932 elections



Notes: The figure presents scatter plots of the political fragmentation index against the 1925 Catholic population share across parishes in Baden. From left to right, the panels correspond to the Reichstag elections of 20 May 1928 and 6 November 1932, and the change in fragmentation between the two elections. The reference line at zero in the change panel distinguishes parishes where fragmentation increased from those where it decreased. A cubic polynomial regression fit is overlaid in red, together with red markers indicating mean fragmentation values within 25 equally sized quantile bins.

effective number of electoral parties was 3.40, despite their homogeneous religious composition. This contrasts sharply with overwhelmingly Catholic parishes, where the effective number of parties was only 2.06.

As hypothesized, political fragmentation was greatest in predominantly Protestant parishes with a sizeable Catholic minority, resulting in an inverted U-shaped relationship between political fragmentation and the Catholic population share in the 1928 election. This pattern is noteworthy because the Reich as a whole had a similar configuration—predominantly Protestant, but with a large Catholic minority. The Catholic population share (and its squared and cubic terms) alone explains about 40% of the cross-parish variation in the fragmentation index. Appendix Table B.3 shows that the inverted U-shaped relationship remains intact when we add controls and exploit only variation within counties or neighborhoods.

In contrast, Protestant parishes were *less* politically fragmented than Catholic ones in the November 1932 election (see the middle panel of Figure 2.5). Their mean effective number of parties fell to just 1.86, compared to 2.19 in parishes with a

Catholic share of 95% or more. The Nazis thus succeeded in unifying the previously fragmented (rural) Protestant vote in the 1932 elections. In Catholic parishes, by contrast, fragmentation changed little due to the continued dominance of the Zentrum, despite some Nazi gains. The inverted U-shape between fragmentation and Catholic population share remains intact for the November 1932 election (see also the regression results in Appendix Table B.4).

Nazi vote share. We turn now to Hypothesis 1b. Table 2.2 summarizes our results on the relationship between local religious composition and the NSDAP vote share. We focus on the Reichstagswahl of 6 November 1932 in Baden. The univariate regression in Column (1) suggests that an increase in the local Catholic population share by one percentage point (pp) is associated with a 0.438 pp decrease in the NSDAP vote share. Put differently, the Nazi vote share was, on average, 43.8 percentage points lower in an all-Catholic parish than in an all-Protestant one. The Jewish population share also exhibits a strong negative correlation with Nazi voting. Religious composition alone explains about 53% of the parish-level variation in NSDAP vote shares.

Columns (2) to (4) successively add a cubic polynomial in longitude and latitude to account for spatial trends, as well as geographic and sociodemographic controls. The coefficient on the Catholic share remains largely unchanged across these specifications, indicating that omitted variables correlated with Catholic settlement patterns are unlikely to drive the observed relationship. The share of explained variation increases only modestly, reaching about 62% in specification (4) with the full set of controls.[10]

Columns (5) to (7) of Table 2.2 address remaining concerns that unobserved variables drive the relationship between parish-level religious composition and Nazi voting by relying on comparisons across nearby parishes. Specifically, we consider regressions

[10]Appendix Figure B.2 quantifies the central role of the Catholic population share in explaining local variation in Nazi voting using general dominance statistics. These statistics are Shapley values that decompose the overall variation explained in Column (4) into components attributable to each explanatory variable. The Catholic population share accounts for almost three quarters of the model's fit, making it by far the most important predictor of the NSDAP vote share in the Reichstag election of 6 November 1932.

that add county fixed effects (Column (5)), exploit only deviations from the mean within a circular neighborhood around each parish (Column (6)), or compare religious enclaves with their surrounding parishes (Column (7)). All three specifications confirm our baseline estimates, with point estimates ranging from –0.431 to –0.538, even in the enclave specification, which uses only about one tenth of the full sample.

We conduct three types of robustness checks for Baden. First, Column (8) shows that our results are quantitatively similar when we focus on the change in the NSDAP vote share between May 1928 and November 1932. Second, Column (9) confirms our findings using an alternative specification that replaces the Catholic population share with a dummy variable for Catholic parishes and restricts the sample to religiously homogeneous locations—those with at least 95% Catholic or 95% Protestant populations. The estimates imply that the NSDAP vote share was, on average, 48.1 pp lower in Catholic than in Protestant parishes. Third, Columns (10) and (11) confirm that our results remain virtually unchanged when we follow Spenkuch and Tillmann (2018) and use the religious situation in 1619 as an instrumental variable for voters' religion in 1925.

Table 2.2: Religious composition and NSDAP voting across parishes in Baden, November 1932

	OLS 11/1932 (1)	OLS 11/1932 (2)	OLS 11/1932 (3)	OLS 11/1932 (4)	OLS 11/1932 (5)	OLS 11/1932 (6)	OLS 11/1932 (7)	OLS Δ 1928-32 (8)	OLS 11/1932 (9)	2SLS 11/1932 (10)	2SLS 11/1932 (11)
Catholics (%)	-0.438***	-0.460***	-0.451***	-0.457***	-0.490***	-0.538***	-0.431***	-0.397***		-0.437***	-0.468***
	(0.028)	(0.029)	(0.027)	(0.026)	(0.030)	(0.033)	(0.063)	(0.030)		(0.028)	(0.032)
Jews (%)	-1.231***	-1.061***	-0.936***	-0.697***	-0.785***	-0.803***	-1.112	-1.052**		-0.465*	-0.552***
	(0.293)	(0.233)	(0.263)	(0.247)	(0.226)	(0.198)	(0.834)	(0.429)		(0.260)	(0.183)
Others (%)	0.078	0.019	0.033	-0.038	0.012	-0.097	-0.551	0.111		-0.011	0.052
	(0.248)	(0.225)	(0.203)	(0.167)	(0.139)	(0.151)	(0.855)	(0.145)		(0.161)	(0.128)
Catholic parish (0/1)									-0.481***		
									(0.032)		
Obs.	1436	1436	1436	1436	1436	1436	146	1436	866	1162	1162
R2	0.530	0.551	0.564	0.618	0.667	0.485	0.738	0.567	0.688		
Kleibergen-Paap F										3504	2329
Spatial trend	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geo. controls	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes
Sociodem. controls	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Within transformation	none	none	none	none	county	radius	enclave	county	county	none	county

Notes: The table shows the relationship between religious composition–as measured by the population share of Catholics, Jews, and other religious groups, with the residual category being Protestants–and the Nazi vote share across parishes in Baden. Results are for the *Reichstagswahl* on 6 November 1932 in Baden. Columns (1) to (9) present OLS estimates, while Columns (10) and (11) report 2SLS estimates using parish religion in 1619 as an instrument for the Catholic population share in 1925. Parishes lacking information on religious composition in 1619 are excluded from the 2SLS specifications. Column (1) is for a parsimonious model without further controls. The remaining columns successively add a cubic polynomial in longitude and latitude to control for spatial trends as well as geographic and sociodemographic control variables, respectively. Columns (5), (8), (9) and (11) add county fixed effects. Column (6) reports coefficients from regressions using variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. Column (7) restricts the sample to enclaves and surrounding parishes, adding fixed effects for each enclave case. Column (8) uses the change in the Nazi vote share between the elections in May 1928 and November 1932 as the dependent variable. Column (9) keeps only parishes that are either 95% or more Catholic or Protestant, dropping parishes with a mixed religious composition. It replaces population shares of the different religions with a dummy variable for Catholic parishes. Robust standard errors, clustered at the enclave-case level in Column (7) and at the county level in all other columns, are reported in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table B.5 replicates our analysis for the state election in Württemberg held on 24 April 1932. The results resemble those of the Reichstag election in Baden six months later, with substantially higher Nazi vote shares in Protestant parishes. However, the Protestant–Catholic differences in Nazi voting were smaller in Württemberg than in Baden. The regression with county fixed effects suggests that an increase in the local Catholic population share by one pp decreased the NSDAP vote share by 0.295 pp in Württemberg, compared with 0.490 pp in Baden. One possible explanation is that the *Württembergischer Bauern- und Weingärtnerbund* played a role for the rural Protestant population in Württemberg similar to that of the Zentrum for Catholics in Baden. This relatively close partisan alignment prevented the NSDAP from fully consolidating the Protestant vote in Württemberg.

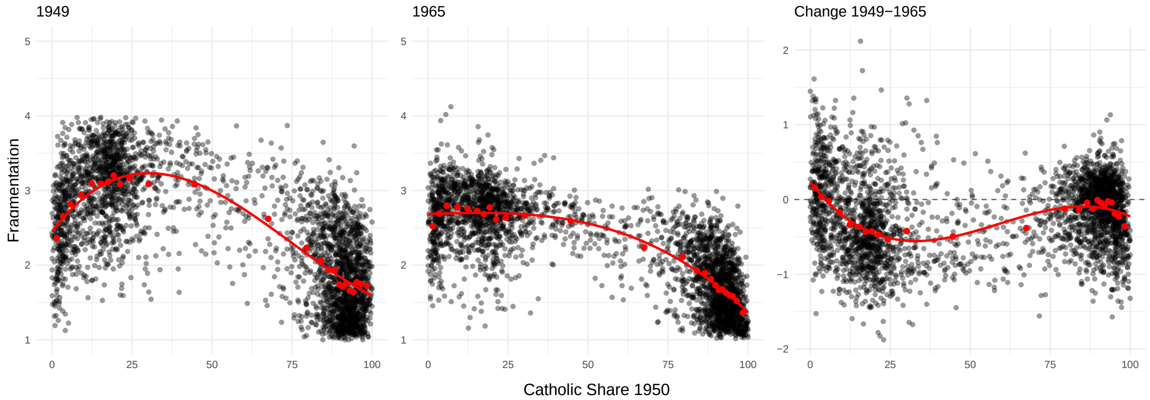
2.5 Results for West Germany

Fragmentation. Figure 2.6 shows that the relationship between fragmentation and the Catholic population share changed substantially between 1949 and 1965. In 1949, the pattern still resembled the hump-shaped relationship observed in the 1928 *Reichstag* election. Mean fragmentation stood at 2.36 in the most Protestant parishes, increased to 3.20 in parishes with around 19% Catholic population, and then decreased sharply to 1.72 in the most Catholic parishes.[11]

By 1965, the fragmentation hump had largely disappeared, consistent with Hypothesis 2a. Fragmentation at the Protestant end of the distribution was nearly flat, with only small differences between almost entirely Protestant parishes and parishes with a Catholic population share around the former peak. Catholic parishes remained less fragmented, with mean values between 1.36 and 1.68 among bins with Catholic shares above 90%. The change panel shows that the decline was largest in confession-

[11]One reason the hump-shaped pattern persisted in 1949 that is not captured by our conceptual framework is that independent candidates who mostly represented the interests of displaced persons were particularly successful in Protestant areas that received Catholic refugees after 1945. With time, support for refugee parties waned, and they were no longer represented in the *Bundestag* since 1957.

Figure 2.6: Share Catholic and political fragmentation in Baden-Württemberg, 8/1949 and 9/1965 elections



Notes: The figure presents scatter plots of the political fragmentation index against the 1950 Catholic population share across parishes in Baden-Württemberg. From left to right, the panels correspond to the Bundestag elections of 14 August 1949 and 19 September 1965, and the change in fragmentation between the two elections. The reference line at zero in the change panel distinguishes parishes where fragmentation increased from those where it decreased. A cubic polynomial regression fit is overlaid in red, together with red markers indicating mean fragmentation values within 25 equally sized quantile bins.

ally mixed parishes, where fragmentation fell by 0.53 among parishes with a Catholic population share around 24%. On average, parish-level fragmentation declined by 0.21 between 1949 and 1965.[12]

By separating Baden from Württemberg, Appendix Figure B.3 reveals a qualitatively similar pattern in both parts of the state: a hump-shaped relationship in 1949 that largely flattens in the Protestant and mixed-religion range by 1965. Subtle differences exist, however. First, the hump-shaped pattern in 1949 was more pronounced in Württemberg, where the religious distribution was more bimodal in 1950 with fewer religiously mixed parishes. Second, the gap between the Protestant and Catholic ends of the distribution widened slightly more in Baden, with modest increases in fragmentation at the Protestant end and decreases at the Catholic end.

[12]Looking at Baden only suggests that much of the decline in fragmentation between the Weimar Republic and the Federal Republic had already occurred by 1949. The mean change between the *Reichstag* and *Bundestag* elections was -0.427 between 1928 and 1949 and -0.644 between 1928 and 1965. The decline by 1949 partly reflects the introduction of the 5% electoral threshold, which applied separately to the federal states in the 1949 election.

CDU and NPD vote shares. Table 2.3 summarizes our results on the relationship between local religious composition and the vote shares of the CDU, the main innovation on the center-right of the political spectrum, and the NPD, the main far-right party in the mid- and late-1960s.

The first key result is that the CDU still won considerably higher vote shares in Catholic parishes. In the 1949 Bundestag election, a one pp increase in the Catholic population share was associated with a 0.496 pp higher CDU vote share. This association weakened somewhat over time, but the coefficient is still at 0.310 pp in 1965.

Figure 2.7 illustrates the second key finding by comparing the Zentrum's vote share in 1928 with the CDU's in 1965 across parishes in Baden. While the Zentrum received virtually no votes in predominantly Protestant parishes in 1928 (left panel), the CDU gained considerable support across the entire religious spectrum in 1965 (center panel). Overall, the CDU gained ground everywhere relative to the Zentrum, but the increase was most significant in predominantly Protestant parishes (right panel). The CDU's integrative success thus lay not in completely eliminating confessional differences, but in creating a common center-right party that could bind both Catholic and Protestant voters.

The third key result, which speaks to Hypothesis 2b, relates to the rise of the NPD in the late 1960s. Table 2.3 shows that NPD support was lower in Catholic parishes. The coefficient is already negative in 1965 (-0.013) and becomes larger in magnitude by 1969 (-0.039), when the NPD reached its strongest federal result. This finding reveals a limitation of the CDU's integrative success. Despite consolidating the center-right vote across denominational lines, the CDU did not make Protestant parishes immune to the appeal of new right-wing parties. The NPD's disproportionate support in Protestant areas shows that the confessional divide in far-right voting persisted into the late 1960s. However, the scale differed sharply from Weimar: far-right

Table 2.3: CDU & NPD Support in 1949, 1965 and 1969

	(1) CDU 1949	(2) CDU 1965	(3) NPD 1965	(4) CDU 1969	(5) NPD 1969
Share of Catholic Population (%)	0.496***	0.310***	-0.013***	0.268***	-0.039***
	(0.03)	(0.07)	(0.00)	(0.06)	(0.01)
Population (*in logs*)	-2.986***	-3.184***	0.101**	-3.671***	-0.607***
	(0.42)	(0.32)	(0.05)	(0.29)	(0.10)
Industry (%)	-0.195***	-0.139***	-0.000	-0.049***	-0.011**
	(0.04)	(0.02)	(0.00)	(0.01)	(0.01)
Females (%)	0.029	-0.060	0.004	0.010	-0.015
	(0.10)	(0.11)	(0.01)	(0.08)	(0.03)
Observations	3052	3117	3117	3117	3117
R-Squared	0.795	0.711	0.378	0.683	0.390
Spatial Trend	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes
Within Transf.	radius	radius	radius	radius	radius

Notes: The table shows the relationship between religious composition, as measured by the population share of Catholics, and CDU and NPD vote shares across parishes in Baden-Württemberg. Results are for the *Bundestagswahl* on 14 August 1949 (Column (1)), 19 September 1965 (Columns (2) and (3)), and 28 September 1969 (Columns (4) and (5)). All regression models use variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. All models include a cubic polynomial in longitude and latitude to control for spatial trends as well as geographic and sociodemographic control variables. Column (1) is missing 65 observations due to missing data in the original data source. Robust standard errors, clustered at the county level, are reported in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

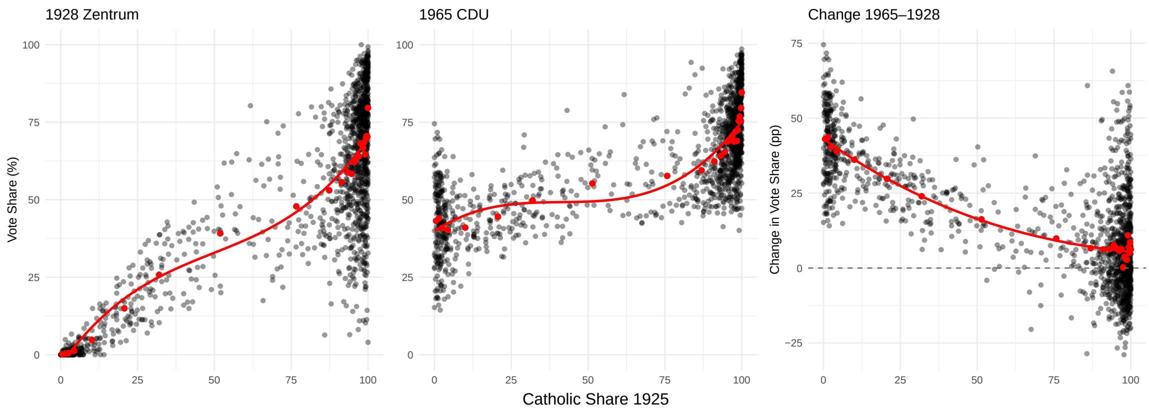
mobilization remained much more contained in the postwar party system.

2.6 Mechanisms: Individual-Level Evidence from 1968

The parish-level evidence documents a robust negative relationship between the local Catholic population share and support for the NPD. Yet these ecological correlations do not by themselves identify the individual-level mechanism. The confessional gap could reflect church influence: Catholic clergy may have steered followers away from right-wing extremism and toward the CDU (Spenkuch and Tillmann, 2018). Alternatively, Catholics and Protestants may simply have held different political views. This section uses survey evidence from the 1968 Baden-Württemberg state election to distinguish these competing explanations for the confessional gap in voting.

Protestants were more likely to vote NPD. Before we turn to mechanisms, Figure 2.8 confirms that Protestants report substantially higher rates of NPD support

Figure 2.7: Zentrum and CDU vote shares in Baden: 1928, 1965, and change



Notes: The figure presents scatter plots of vote shares against the 1925 Catholic population share across parishes in Baden. The left panel shows the Zentrum vote share in the Reichstag election of May 1928; the center panel shows the CDU vote share in the Bundestag election of September 1965; the right panel shows the change in vote share between the two elections (CDU 1965 minus Zentrum 1928). The reference line at zero in the change panel distinguishes parishes where the CDU gained relative to the Zentrum from those where it lost votes. A cubic regression fit is overlaid in red, together with red markers indicating mean vote shares within 25 equally sized quantile bins.

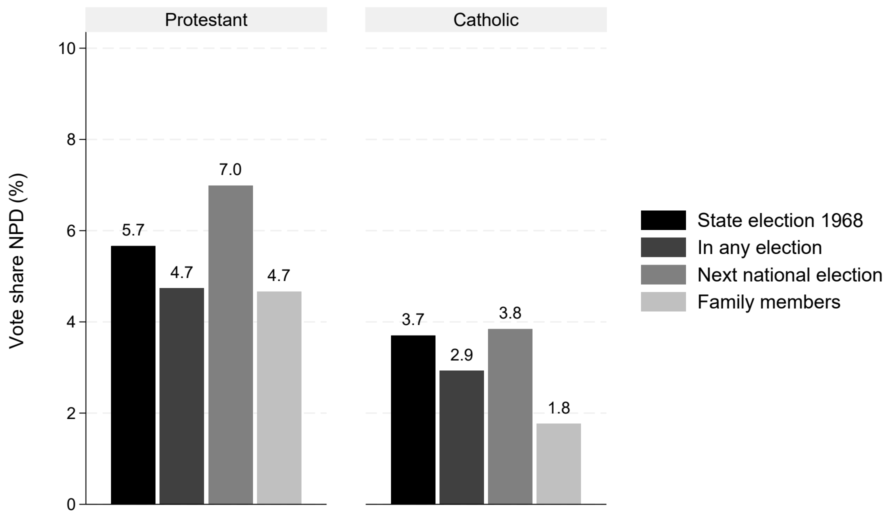
than Catholics across four indicators: having voted for the NPD in the 1968 state election, having voted for the NPD in any past election, intending to vote for the NPD in the next national election, and having family members who support the NPD.[13] These results suggest that the parish-level patterns documented in Section 2.5 indeed reflect differences in individual voting behavior. Appendix Table B.6 also suggests that these raw differences persist conditional on socioeconomic controls. Conversely, Catholics were considerably more likely to have voted for the CDU in 1968, consistent with the aggregate evidence from the previous section.[14]

Similar policy preferences. One explanation for the confessional gap is that Catholics and Protestants have different interests and political views. However, the survey evidence points against this interpretation. Figure 2.9 shows agreement with a set of NPD-aligned positions—including nationalism approval, support for a strong national leader, anti-elitism, resentment of US dominance, and the view that

[13]Following Klingemann (1971), we also use indirect measures of NPD support to mitigate potential underreporting due to respondents' reluctance to reveal an extremist vote in face-to-face interviews.

[14]Appendix Figure B.4 shows stated voting behavior for all four major parties in the 1968 state election, broken down by religious affiliation.

Figure 2.8: NPD support, by confession



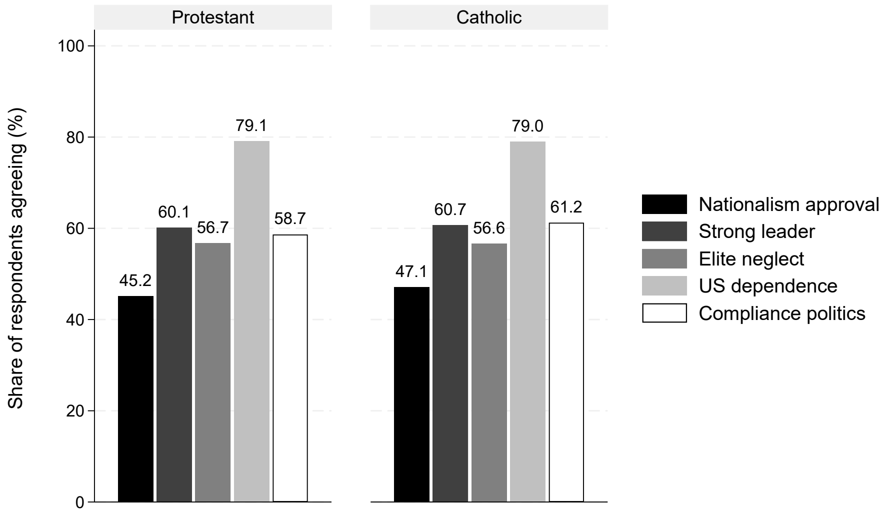
Notes: The figure shows the percentage of respondents who indicated that they or their family support the NPD. The first bar shows the percentage of respondents who state that they voted for the NPD in Baden-Württemberg's 1968 state election, conditional on voting. The second bar indicates the percentage of respondents who state that they voted for the NPD in any past election. The third bar indicates the percentage of voters who say they will vote for the NPD in the next national election. Finally, the fourth bar indicates the percentage of respondents whose family members generally support the NPD.

Germany pursues a compliance policy toward the victorious powers—broken down by confession. The share of respondents agreeing with these positions is strikingly similar between Catholics and Protestants.[15] For example, 45% of Protestants and 47% of Catholics agreed that nationalism is a good party characteristic, while 60% of Protestants and 61% of Catholics agreed that there is nothing wrong with a strong national leader.

Elite influence of the clergy. The elite mechanism implies that the Catholic Church's opposition to the NPD influenced its followers' voting behavior, despite their political preferences. For the mechanism to work, Catholics had to perceive the Catholic Church as opposing the NPD and uphold an electoral norm of voting CDU instead. At the same time, we would expect those who value the church more highly to act on that signal.

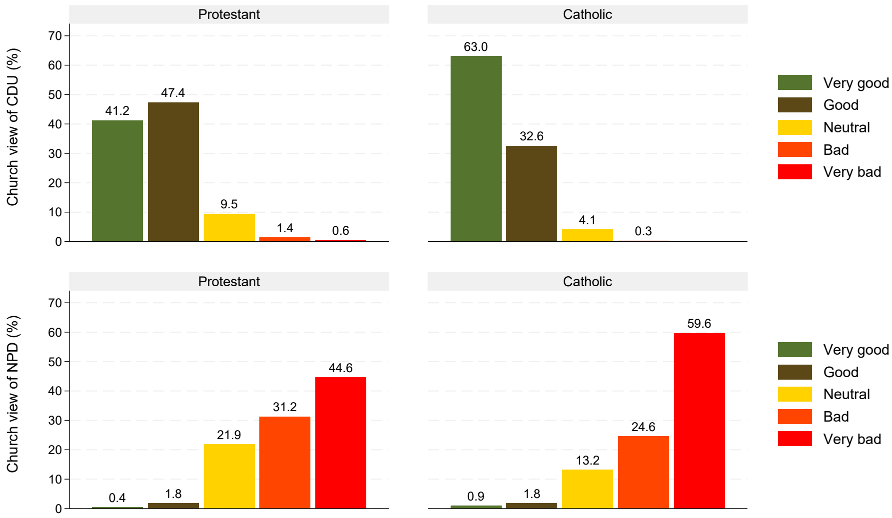
[15]Appendix Figures B.5 and B.6 and Appendix Tables B.7 and B.8 extend this finding to a broader set of policy topics and statements: the issues that Catholics and Protestants consider important, and the policy goals they endorse, show only small and largely statistically insignificant differences. The main exception is confessional schools, which are more important to Catholics.

Figure 2.9: Support for NPD positions, by confession



<u>Notes</u>: The figure shows the percentage of respondents who agree with statements supported by the NPD. We aggregate the proportion of respondents expressing weak and strong agreement. The 'nationalism approval' statement holds that nationalism is a good party characteristic. The 'strong leader' statement asserts that there is nothing wrong with a strong national leader, such as those that Germany used to have, provided they are fair. The 'elite neglect' statement is that wealthy people do not care about ordinary people. The 'US dependence' statement asserts that it is detrimental to the German economy to be under American control. The 'compliance politics' statement asserts that Germany conducts compliance policy (*Erfüllungspolitik*) toward the victorious powers. The left panel shows the results for Protestant respondents, and the right panel shows the results for Catholic respondents.

Figure 2.10: Perception of CDU and NPD by the Protestant and Catholic Churches



Notes: The figure shows the percentage of respondents who assessed their churches' stances toward the CDU (upper panel) and the NPD (lower panel) as very good, good, neutral, bad, and very bad. The left panel shows the results for Protestant respondents, while the right panel shows the results for Catholic respondents.

Figure 2.10 shows moderate but visible differences in how respondents perceived their own churches' stances toward the CDU and the NPD. For Catholic respondents, the message is clear: the vast majority viewed the Catholic Church's stance on the NPD as 'bad' or 'very bad' (84%), while viewing its stance on the CDU as 'good' or 'very good' (96%). Among Protestant respondents, the perceived stance on the NPD is more ambiguous, with 24% viewing their churches' stance on the party as neutral or even positive. Most Protestants view their churches' stance on the CDU positively, which is a clear sign of the CDU's growing cross-confessional appeal. However, the proportion of Protestants who perceive their churches' view of the CDU as 'very good' is substantially lower than among Catholics.

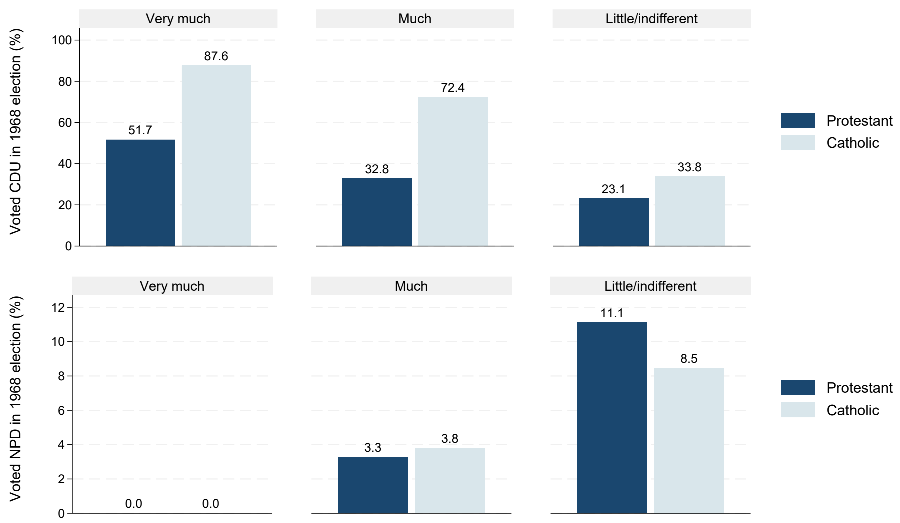
Overall, the Catholic Church sent a clearer anti-NPD and pro-CDU signal to its followers than the Protestant churches did. Appendix Figure B.7 corroborates this pattern. A significant proportion of Protestants viewed their churches' stance on the liberals and Social Democrats as being as favorable as their position on the CDU. In contrast, the majority of Catholics identified the CDU as the party most strongly favored by their church.

Figure 2.11 shows that church attachment was closely related to voting behavior. Among Catholics, respondents who placed high importance on their church were more likely to vote CDU and less likely to vote NPD than Catholics who were largely indifferent towards the church. A similar pattern holds among Protestants: those who considered their church very important also reported very low NPD vote shares, comparable to their Catholic counterparts (consistent with previous evidence in Klingemann, 1971; Liepelt, 1967). The aggregate confessional gap in NPD voting is therefore partly compositional. Catholics were more likely to be devout: 29% of Catholic respondents said the church was very important to them, compared with 18% of Protestants. Since NPD support was low among devout respondents in both confessions, Catholics' stronger church attachment contributed to their lower aggregate NPD support.

Appendix Table B.9 confirms these patterns in a regression framework, conditioning on age, gender, education, marital status, and community size. Columns (1)–(4) show that church importance is a significant positive predictor of CDU voting. The association is present in both confessions but stronger among Catholics, as indicated by the positive interaction between church importance and Catholic affiliation in Column (4). This pattern is consistent with the stronger pro-CDU signal Catholics received from their church, but also shows that the CDU made inroads among devout Protestants. By contrast, local club membership does not predict CDU voting, even among Catholics. This is consistent with Haffert (2022), who argues that southern German Catholic mobilization relied more on the church itself than on the dense associational networks usually associated with the Catholic milieu.

Columns (5)–(8) show that high church importance is associated with significantly lower NPD support. Unlike for CDU voting, the interaction between church importance and Catholic affiliation in Column (8) is small and not statistically significant. This suggests that church attachment reduced NPD support similarly among Catholics and Protestants. The pattern matches Figure 2.11: among respondents who placed

Figure 2.11: Share of CDU and NPD voters by individual importance of church and confession



Notes: The figure shows the percentage of respondents who voted for the CDU (top panel) or the NPD (bottom panel) in the 1968 election in Baden-Württemberg. Respondents are categorized based on the importance they attach to their church (very much, much, or little/indifferent) and their religious affiliation. Non-voters are excluded.

a very high importance on their church, NPD support was essentially zero in both confessions. The confessional gap in NPD voting therefore partly reflected differences in religiosity.

Taken together, the individual-level evidence favors the elite influence explanation over the preference explanation. Despite holding similar nationalist and anti-establishment views, Catholics were considerably less likely to vote NPD—a gap driven by two reinforcing mechanisms. First, the Catholic Church sent a clearer political signal: Catholics overwhelmingly perceived it as opposing the NPD and favoring the CDU, whereas Protestants viewed their church's stance as less clear. Second, Catholics were more likely to attach very high importance to their church.

2.7 Conclusion

It has long been argued that the political instability that brought down the Weimar Republic was rooted in party fragmentation on the center-right (e.g. Hermens, 1968; Jones, 1972), and that overcoming this fragmentation was key to West Germany's postwar stability (e.g. Smith, 1982). This paper empirically demonstrates how confessional politics shaped both Weimar fragmentation and its postwar attenuation, using the rich confessional mosaic of southwestern Germany for identification.

In Weimar, the Catholic-Protestant divide produced a hump-shaped fragmentation pattern, peaking in mixed parishes where a fragmented Protestant bloc competed with the Catholic Zentrum. As preferences polarized during the Great Depression, the NSDAP consolidated the fragmented Protestant vote while the Zentrum remained the main anchor for the Catholic vote. The CDU altered this dynamic after the war by appealing to Catholic and Protestant voters on the center-right, thereby gradually dissolving the fragmentation hump.

However, confessional voting did not disappear completely, and Protestant areas were more susceptible to the NPD in the late 1960s. Survey evidence reveals that Catholics were less likely than Protestants to vote for the NPD, despite holding similar nationalist and anti-establishment attitudes. This gap is explained by two reinforcing mechanisms, consistent with the elite influence explanation: the Catholic Church conveyed a stronger anti-NPD and pro-CDU message than the Protestant Church, and Catholics placed greater importance on their church.

Taken together, these findings highlight how organizational structures shape the effect of religious divisions on party fragmentation and election outcomes. During Weimar, Catholic voters were loyal to the Zentrum, while Protestant center-right voters lacked an equivalent political home. In West Germany, the CDU reduced this asymmetry by providing a cross-confessional anchor. More broadly, the German case suggests that cross-cleavage parties can mitigate the fragmentation risks associated with social divisions—an organizational lesson for maintaining stability in divided societies.

Chapter 3

Reproducible LLM Workflows in Economic History: Building a Dataset from Degener's *Wer Ist's?*[1]

3.1 Introduction

Research in economic history is often not constrained by the scarcity of primary sources but by the cost of transforming abundant material into usable data. Even when the research questions we pose are qualitative or theoretical, many projects require large-scale data work, and as a result, many ideas never leave the brainstorming stage. Recent developments in AI tools can lower these costs, but integrating them safely and reproducibly into research practice remains challenging. In this paper, I argue that transparent and reproducible large language model (LLM) workflows can address some of these concerns, provided they are documented end-to-end and

[1]This paper is accepted and forthcoming with the same title in a special issue in the Economic History Yearbook 01/2027. This is a slightly modified version for this dissertation. It differs only by the addition of a new appendix and corresponding references in the text. I thank Sibylle Lehmann-Hasemeyer, Felix Selgert, Sebastian Braun, Mario Larch, Hartmut Egger, participants at 2nd CEPR Applied Micro-Economic History Workshop in St. Gallen and EHS Centennial Conference in London as well as the seminar participants at the University of Bayreuth for their comments. All remaining errors are my own.

designed so that outputs can be checked and rerun.

Using LLMs is not inherently reliable: hallucinations, misunderstandings, and formatting issues can lead to errors that are nearly impossible to backtrack. The rapid and sometimes untraceable changes in proprietary models further threaten the reproducibility of results. I address these issues with an agentic pipeline that separates tasks into clear, easy-to-understand, and implementable categories: recognizing text, identifying relevant information, extracting it, validating it, and flagging potential problems for human researchers. The goal is not to find or use a perfect model, but to build a workflow that is robust to imperfect steps.

I argue that reproducibility remains achievable despite LLMs' inherent stochasticity. To showcase this reproducibility workflow, I use a challenging case where the text is dense and filled with abbreviations. *Wer Ist's?* is a book published in 1911 by Herrmann A. L. Degener (Degener, 1911) that includes biographies of living contemporaries relevant to German high society. The first edition was published in Leipzig, Germany, in 1905, and a total of ten editions were published until 1935. The 1911 Edition lists around 16,000 people, totaling 1,686 pages. Each biography can contain up to 18 variables, ranging from basic demographic information to career details and contemporary residence address. Building a dataset of every person mentioned within, along with their biographies, would have been very costly just a few years ago.

To extract these biographies and create a dataset, I built an agentic LLM system in Python using Mistral's and OpenAI's models and APIs. All the code and data for this system are publicly available on GitHub (Öztürk, 2026a,b). Appendix C explains how the repository can be inspected and rerun.[2] The system assigns each workflow step a clear objective and a fail-safe: every step logs its decisions and flags uncertainty, so the researcher can identify exactly where the pipeline struggled and intervene

[2] Some sections of Appendix C document the *Wer Ist's?* workflow directly, while others use additional exercises to isolate specific methodological issues such as version stability, prompt sensitivity, open-weight model behavior, and occupation-coding benchmarks; see Appendix C.1.

before the data are used. The full pipeline was processed in 66 minutes, with API costs totaling 159.24 Euros.

I begin with a brief summary of recent advances in AI and their relationship to the field of economic history in Section 3.2. Section 3.3 outlines the definitions of reproducibility and the need for transparent, accessible methods. I then introduce the source material, *Wer Ist's?* in Section 3.4. Section 3.5 presents the workflow, its guiding principles, and its main components in detail. Section 3.6 reports the workflow metrics, presents the dataset from Wer Ist's?, and discusses good practices and open-weight alternatives. Section 3.7 concludes.

3.2 A Survey of Recent Literature

Recently, the attention of the literature has pivoted towards a task-based understanding of AI and many aspects of human-AI collaboration. For instance, new research emphasizes the importance of designing AI systems that can collaborate effectively with humans rather than simply replace them (Brynjolfsson et al., 2025). The new AI products are also designed with this in mind. Instead of receiving a prompt and predicting a "correct" answer, the new AI products rely on multiple calls and agentic systems to verify knowledge and maintain accuracy. They can access augmented tools that can crawl the internet, read the provided information, and run code. The same principles can be applied to our workflows in economic history as well.

Economic historians have struggled to handle scalable data extraction, largely due to financial or time constraints. Against this backdrop, there have been many attempts to unify or create novel methods to alleviate extraction issues in economic history (Beach and Hanlon, 2023a; Abramitzky et al., 2020; Shen et al., 2021). Traditionally, software solutions such as Tesseract or Transkribus have been the primary tools for this task (or hiring a firm that specializes in digitization). These systems rely on optical character or word recognition algorithms to produce text. While effective for modern text, such solutions often prove insufficient for economic historians. Recent work using LLMs, however, has shown that many of the problems and limitations previously encountered are disappearing. Some projects are already showcasing the strength of LLMs in vision and information-extraction tasks (Lagakos et al., 2025; Voth and Yanagizawa-Drott, 2024).

By leveraging LLMs' vision and text understanding capabilities, it is possible to reliably extract information from documents. Recent work demonstrates that LLMs can extract information from printed documents with high accuracy and achieve acceptable results even on unstructured layouts (Ma et al., 2024). However, model-level benchmarks improve so rapidly that comparisons published even a year ago may no longer reflect the state of the art (Greif et al., 2025). Since many current models already approach near-perfect accuracy on standard OCR tasks, I argue that the binding constraint has shifted from model selection to workflow design (Griesshaber and Streb, 2026). Yet using LLMs in a reproducible way has received little attention in economic history so far, which this paper aims to address. For approaches outside economic history, see Cui and Alexander (2026). The data extraction workflows in economic history are especially fragile to AI agents, given the stochasticity and black-box nature of these products.

3.3 Definitions and Direction

To achieve reproducible research, one needs to define what reproducibility entails. The broader social science literature offers several perspectives on this, and a good candidate for discussion here is the work of Christopher Barrie, Alexis Palmer, and Arthur Spirling (Barrie et al., 2025). They present a comprehensive view of replication for the political scientists who would like to use LLMs in their workflows, categorizing it into three distinct categories. First, there is deterministic replication, which is the classical replication we might think of in economics, where we have static data and code. Second, closer to economic history, there is the *stochastic* replication. This type of replication is based on crowdsourcing, whether it involves a company typing the cells in the tables, research assistants, or the researcher themselves. Third, there is the replication of static categorical assignments, which is more specific to political science and less important for our context, since economic historians typically extract unstructured information rather than apply predetermined coding schemes. I build upon their definition of *stochastic reproducibility*.

Economic historians mostly deal with and are interested in the extraction of historical information, statistics, and unstructured text if LLMs are involved. This process, especially if done by humans, is usually subject to many corrections and checks. This is why the end result, which is the dataset, being correct and reproducible, is more important than getting every LLM step reproducibly. This, however, does not mean that there should be more blind trust in the system; on the contrary, since we know that LLMs' results at each step have some degree of stochasticity, we should build systems that address these issues.

Simple *computational replication* of workflows does not suffice in economic history research. Just because another researcher can run the same workflow does not imply that it is a feasible option. For example, one of the solutions proposed to the stochasticity problem of LLMs is using an open-weight model with set seeds or low temperatures.[3] This, however, is misleading as the vast majority of researchers do not have constant access to the high-performance computing needed to run open-weight models. Nor are the models themselves designed to perform as well as proprietary models, as lower model temperatures drastically degrade the qualitative performance of open-weight models. This is why I also focus on comparing open-weight and proprietary language models in Section 3.5.

Another important aspect is model choice. I argue that, if possible, the focus on

[3]Setting a seed fixes the initial internal parameters of the model and leads it to be almost perfectly reproducible, especially if combined with a low temperature setting, which is the *creativity* setting of the model.

specific proprietary models or products of a single company or organization should be avoided. Economic historians are mostly interested in *extracting* data from primary sources, and we can always go back and verify the extraction results. Therefore, why should the model choice matter semantically if we can verify the end result? Popular model families such as Gemini, Llama, and GPT, of course, have different prices, speeds, and quality when it comes to generation, but our workflows in economic history are not about generating data that does not exist, but about reliably extracting information from primary sources.

Transparency and accessibility are two other essential aspects of LLM-based economic history projects, enabling clear evaluation of our decisions, results, and reasoning. The design decisions made during the ideation phase should be transparent, including model and parameter choices. If one step of the process is automatically verified and the other is not, that should also be clearly explained. Finally, the code, intermediate results, failures, LLM logs including the processing times and token costs, and final results should be easily accessible and inspectable.[4] This study follows open-science principles, and all materials are available on GitHub (Öztürk, 2026a).

3.4 Introducing the Source: Degener's Wer Ist's?

Wer Ist's? (German for "Who is it?") was a widely used biographical directory of notable living persons in the German Empire, founded and edited by Herrmann A. L. Degener. The first edition was published in 1905 as a German counterpart to Britain's *Who's Who*? There are ten editions in total, the last in 1935. Degener issued new editions at irregular intervals; the first five editions came out within six years, expanding and updating entries as public figures and their careers changed. According to Herrmann A. L. Degener, the first edition was almost sold out. The 1911 edition has around 16,000 biographies, indicating that Degener was quite generous with the numbers. Publication paused during World War I and resumed with two

[4] I argue that it is essential to see how the output quality evolves and the improvements are made.

additional editions in the Weimar Republic in 1922 and 1928. The tenth and last edition was published in 1935; Degener died in 1943. Since this book was a product of Degener's initiative, there were no other editions. Only after the Second World War was the concept revived: a new volume appeared in 1948 under the title *Wer ist Wer?* (Who is Who?) In 1951, an editor named Walter Habel formally continued the series under that name, which lives to this day.

Although we don't know how the selection into the book was made, it was in no way representative, and in some cases, it was not even voluntary. According to the modern-day publisher, those listed in the lexicon do not pay for inclusion; entries are selected on merit. I was unable to determine whether this was the case before World War II. In Degener's time, a part of the business model involved selling volumes to libraries, institutions, and private buyers. Degener's professional networks may also have strengthened both his finances and his reputation (or vice versa): from 1911 to 1927, he served as treasurer of the Central Office for German Personal and Family History (*Zentralstelle für deutsche Personen- und Familiengeschichte*). Degener also earned public recognition for the project; in the 1912 edition, he could advertise that the book had won a Grand Prix at the World Exposition in Brussels.

The lexicon offers detailed biographical entries, including a person's birth date and place, family background, education and career, major works or accomplishments, personal interests, political party affiliation, memberships in societies, and contemporary address. It aimed to cover influential figures in public life, but it also includes many artists, such as opera singers and painters, in these editions. One of the most common occupations is *that of Schriftsteller* (authors). Women who had gained public distinction - such as artists, writers, opera singers, or educators - were included, even though men dominated the entries, reflecting contemporary patterns of public office; nonetheless, the lexicon did acknowledge some women, as in the entries for the soprano Therese Malten and others in the 1905 edition.

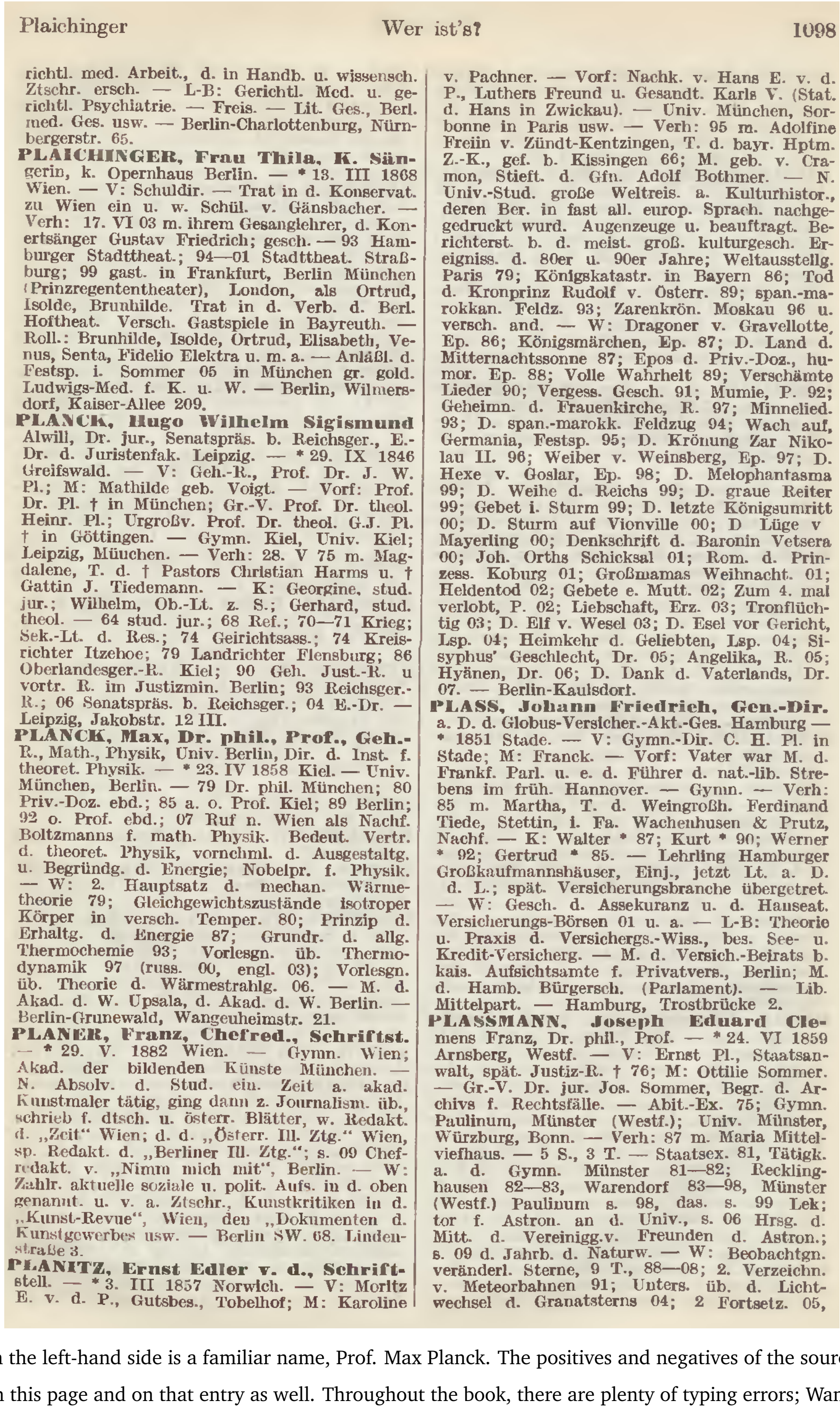
Interestingly, the book also incorporated reader feedback. The sixth edition (1912) included a form for users to list notable individuals they "missed" in the book and return it to the publisher for possible inclusion. Across successive editions, new entries were added, while those who had died were removed, since *Wer Ist's?* documented only living contemporaries. Each edition thus offers a snapshot of who was considered noteworthy by the editor at that time.

All in all, this book and this edition were selected for digitization and for this paper partly because they are printed in an older, dense typeface and partly because their abbreviations are non-standard, both of which have discouraged earlier large-scale digitization. Entries also frequently break across columns or pages, and the source itself contains minor typesetting errors, such as Wangeuheimstr. for Wangenheimstrasse in the entry for Prof. Max Planck shown in Figure 3.1. Combined with roughly 16,000 biographies across 1,686 pages, these features defeat both traditional OCR and rule-based abbreviation expansion, which is why a workflow that can flag and show uncertainty is more useful here than a single high-accuracy model.

3.5 Workflow for the Digitization of *Wer Ist's?*

An agentic framework is simply a chain of LLMs, where each LLM takes the output of the previous one as input. The number of agents is as many as the workflow requires, as the task defined for an LLM agent can be anything, including image recognition, OCR, labeling, classification, and simply inference. An essential component of the workflow is automatic validation. Since LLMs are known to hallucinate under certain conditions, one can build a smarter supervisor agent whose sole job is to detect whether the previous agent did everything correctly and can be as conservative as needed. As the number of tasks increases, it is inescapable that most of the validation comes from agentic LLMs. Figure 3.2 displays the full workflow for *Wer Ist's?* and the key details of this approach, which could be used to extract structured information

Figure 3.1: Page 1098 of Degener's *Wer Ist's?*


Plaichinger Wer ist's? 1098

richtl. med. Arbeit., d. in Handb. u. wissensch. Ztschr. ersch. — L-B: Gerichtl. Med. u. gerichtl. Psychiatrie. — Freis. — Lit. Ges., Berl. med. Ges. usw. — Berlin-Charlottenburg, Nürnbergerstr. 65.

PLAICHINGER, Frau Thila, K. Sängerin, k. Opernhaus Berlin. — * 13. III 1868 Wien. — V: Schuldir. — Trat in d. Konservat. zu Wien ein u. w. Schül. v. Gänsbacher. — Verh: 17. VI 03 m. ihrem Gesanglehrer, d. Konertsänger Gustav Friedrich; gesch. — 93 Hamburger Stadttheat.; 94—01 Stadttheat. Straßburg; 99 gast. in Frankfurt, Berlin München (Prinzregententheater), London, als Ortrud, Isolde, Brunhilde. Trat in d. Verb. d. Berl. Hoftheat. Versch. Gastspiele in Bayreuth. — Roll.: Brunhilde, Isolde, Ortrud, Elisabeth, Venus, Senta, Fidelio Elektra u. m. a. — Anläßl. d. Festsp. i. Sommer 05 in München gr. gold. Ludwigs-Med. f. K. u. W. — Berlin, Wilmersdorf, Kaiser-Allee 209.

PLANCK, Hugo Wilhelm Sigismund Alwill, Dr. jur., Senatspräs. b. Reichsger., E.-Dr. d. Juristenfak. Leipzig. — * 29. IX 1846 Greifswald. — V: Geh.-R., Prof. Dr. J. W. Pl.; M: Mathilde geb. Voigt. — Vorf: Prof. Dr. Pl. † in München; Gr.-V. Prof. Dr. theol. Heinr. Pl.; Urgroßv. Prof. Dr. theol. G.J. Pl. † in Göttingen. — Gymn. Kiel, Univ. Kiel; Leipzig, Müuchen. — Verh: 28. V 75 m. Magdalene, T. d. † Pastors Christian Harms u. † Gattin J. Tiedemann. — K: Georgine, stud. jur.; Wilhelm, Ob.-Lt. z. S.; Gerhard, stud. theol. — 64 stud. jur.; 68 Ref.; 70—71 Krieg; Sek.-Lt. d. Res.; 74 Geirichtsass.; 74 Kreisrichter Itzehoe; 79 Landrichter Flensburg; 86 Oberlandesger.-R. Kiel; 90 Geh. Just.-R. u vortr. R. im Justizmin. Berlin; 93 Reichsger.-R.; 06 Senatspräs. b. Reichsger.; 04 E.-Dr. — Leipzig, Jakobstr. 12 III.

PLANCK, Max, Dr. phil., Prof., Geh.-R., Math., Physik, Univ. Berlin, Dir. d. Inst. f. theoret. Physik. — * 23. IV 1858 Kiel. — Univ. München, Berlin. — 79 Dr. phil. München; 80 Priv.-Doz. ebd.; 85 a. o. Prof. Kiel; 89 Berlin; 92 o. Prof. ebd.; 07 Ruf n. Wien als Nachf. Boltzmanns f. math. Physik. Bedeut. Vertr. d. theoret. Physik, vornehml. d. Ausgestaltg. u. Begründg. d. Energie; Nobelpr. f. Physik. — W: 2. Hauptsatz d. mechan. Wärmetheorie 79; Gleichgewichtszustände isotroper Körper in versch. Temper. 80; Prinzip d. Erhaltg. d. Energie 87; Grundr. d. allg. Thermochemie 93; Vorlesgn. üb. Thermodynamik 97 (russ. 00, engl. 03); Vorlesgn. üb. Theorie d. Wärmestrahlg. 06. — M. d. Akad. d. W. Upsala, d. Akad. d. W. Berlin. — Berlin-Grunewald, Wangeuheimstr. 21.

PLANER, Franz, Chefred., Schriftst. — * 29. V. 1882 Wien. — Gymn. Wien; Akad. der bildenden Künste München. — N. Absolv. d. Stud. ein. Zeit a. akad. Kunstmaler tätig, ging dann z. Journalism. üb., schrieb f. dtsch. u. österr. Blätter, w. Redakt. d. „Zeit" Wien; d. d. „Österr. Ill. Ztg." Wien, sp. Redakt. d. „Berliner Ill. Ztg."; s. 09 Chefredakt. v. „Nimm mich mit", Berlin. — W: Zahlr. aktuelle soziale u. polit. Aufs. in d. oben genannt. u. v. a. Ztschr., Kunstkritiken in d. „Kunst-Revue", Wien, den „Dokumenten d. Kunstgewerbes usw. — Berlin SW. 68. Lindenstraße 3.

PLANITZ, Ernst Edler v. d., Schriftstell. — * 3. III 1857 Norwich. — V: Moritz E. v. d. P., Gutsbes., Tobelhof; M: Karoline v. Pachner. — Vorf: Nachk. v. Hans E. v. d. P., Luthers Freund u. Gesandt. Karls V. (Stat. d. Hans in Zwickau). — Univ. München, Sorbonne in Paris usw. — Verh: 95 m. Adolfine Freiin v. Zündt-Kentzingen, T. d. bayr. Hptm. Z.-K., gef. b. Kissingen 66; M. geb. v. Cramon, Stieft. d. Gfn. Adolf Bothmer. — N. Univ.-Stud. große Weltreis. a. Kulturhistor., deren Ber. in fast all. europ. Sprach. nachgegedruckt wurd. Augenzeuge u. beauftragt. Berichterst. b. d. meist. groß. kulturgesch. Ereigniss. d. 80er u. 90er Jahre; Weltausstellg. Paris 79; Königskatastr. in Bayern 86; Tod d. Kronprinz Rudolf v. Österr. 89; span.-marokkan. Feldz. 93; Zarenkrön. Moskau 96 u. versch. and. — W: Dragoner v. Gravellotte, Ep. 86; Königsmärchen, Ep. 87; D. Land d. Mitternachtssonne 87; Epos d. Priv.-Doz., humor. Ep. 88; Volle Wahrheit 89; Verschämte Lieder 90; Vergess. Gesch. 91; Mumie, P. 92; Geheimn. d. Frauenkirche, R. 97; Minnelied. 93; D. span.-marokk. Feldzug 94; Wach auf, Germania, Festsp. 95; D. Krönung Zar Nikolau II. 96; Weiber v. Weinsberg, Ep. 97; D. Hexe v. Goslar, Ep. 98; D. Melophantasma 99; D. Weihe d. Reichs 99; D. graue Reiter 99; Gebet i. Sturm 99; D. letzte Königsumritt 00; D. Sturm auf Vionville 00; D Lüge v Mayerling 00; Denkschrift d. Baronin Vetsera 00; Joh. Orths Schicksal 01; Rom. d. Prinzess. Koburg 01; Großmamas Weihnacht. 01; Heldentod 02; Gebete e. Mutt. 02; Zum 4. mal verlobt, P. 02; Liebschaft, Erz. 03; Tronflüchtig 03; D. Elf v. Wesel 03; D. Esel vor Gericht, Lsp. 04; Heimkehr d. Geliebten, Lsp. 04; Sisyphus' Geschlecht, Dr. 05; Angelika, R. 05; Hyänen, Dr. 06; D. Dank d. Vaterlands, Dr. 07. — Berlin-Kaulsdorf.

PLASS, Johann Friedrich, Gen.-Dir. a. D. d. Globus-Versicher.-Akt.-Ges. Hamburg — * 1851 Stade. — V: Gymn.-Dir. C. H. Pl. in Stade; M: Franck. — Vorf: Vater war M. d. Frankf. Parl. u. e. d. Führer d. nat.-lib. Strebens im früh. Hannover. — Gymn. — Verh: 85 m. Martha, T. d. Weingroßh. Ferdinand Tiede, Stettin, i. Fa. Wachenhusen & Prutz, Nachf. — K: Walter * 87; Kurt * 90; Werner * 92; Gertrud * 85. — Lehrling Hamburger Großkaufmannshäuser, Einj., jetzt Lt. a. D. d. L.; spät. Versicherungsbranche übergetret. — W: Gesch. d. Assekuranz u. d. Hanseat. Versicherungs-Börsen 01 u. a. — L-B: Theorie u. Praxis d. Versichergs.-Wiss., bes. See- u. Kredit-Versicherg. — M. d. Versich.-Beirats b. kais. Aufsichtsamte f. Privatvers., Berlin; M. d. Hamb. Bürgersch. (Parlament). — Lib. Mittelpart. — Hamburg, Trostbrücke 2.

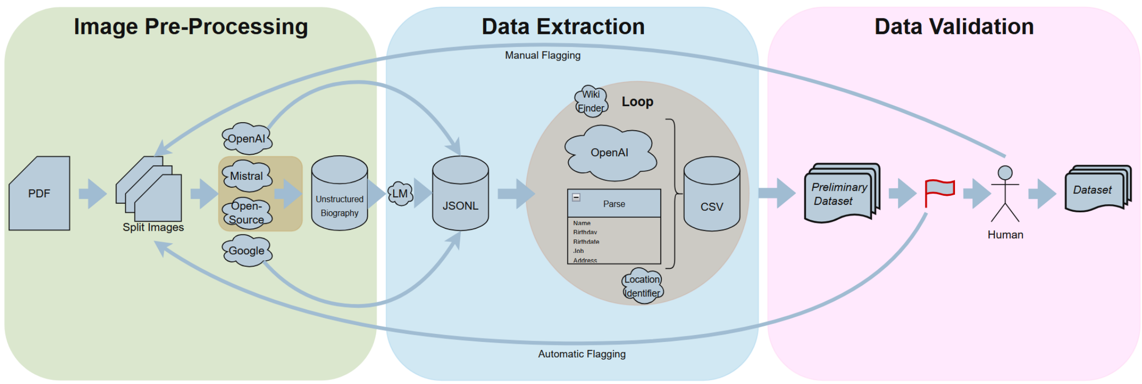
PLASSMANN, Joseph Eduard Clemens Franz, Dr. phil., Prof. — * 24. VI 1859 Arnsberg, Westf. — V: Ernst Pl., Staatsanwalt, spät. Justiz-R. † 76; M: Ottilie Sommer. — Gr.-V. Dr. jur. Jos. Sommer, Begr. d. Archivs f. Rechtsfälle. — Abit.-Ex. 75; Gymn. Paulinum, Münster (Westf.); Univ. Münster, Würzburg, Bonn. — Verh: 87 m. Maria Mittelviefhaus. — 5 S., 3 T. — Staatsex. 81, Tätigk. a. d. Gymn. Münster 81—82; Recklinghausen 82—83, Warendorf 83—98, Münster (Westf.) Paulinum s. 98, das. s. 99 Lektor f. Astron. an d. Univ., s. 06 Hrsg. d. Mitt. d. Vereinigg.v. Freunden d. Astron.; s. 09 d. Jahrb. d. Naturw. — W: Beobachtgn. veränderl. Sterne, 9 T., 88—08; 2. Verzeichn. v. Meteorbahnen 91; Unters. üb. d. Lichtwechsel d. Granatsterns 04; 2 Fortsetz. 05,

Notes: On the left-hand side is a familiar name, Prof. Max Planck. The positives and negatives of the source material can be seen in this page and on that entry as well. Throughout the book, there are plenty of typing errors; Wangeuheimstr. 21 does not exist, and it should have been Wangenheimstr. 21.

Source: H.A.L. Degener (Ed.), *Wer Ist's? Unsere Zeitgenossen*, 5th edition, Leipzig 1911, p. 1098.

from any economic history source, provided there is a high-quality PDF source.

Figure 3.2: Workflow to extract information from historical documents



Notes: The workflow components are broken down into broad categories. The main takeaway is twofold: (1) keep each step as simple as possible, and (2) always make it possible to automatically or manually review the outputs of the LLM. *Source*: Author's own.

Image Preprocessing. I start the workflow by splitting the PDF pages into PNG images.[5] There are many benefits to using images over full PDFs, but the most basic reasoning, and a recurring one, is the context window of LLMs. Context window, simply put, is the limit of the model's understanding per inquiry. The less context we provide in our initial prompt, the more the model has to think and develop. Although there is no one-size-fits-all strategy for prompting and context windows, it is usually good practice to keep each LLM step as simple as possible.

Next, after obtaining the single images, my code takes each PNG and automatically turns a two-column photo into two cleaner page images that are easier to read: it first straightens the picture if the scan is slightly tilted, then finds the middle "column fold" area to separate the left and right pages, trims extra blank space around the text while keeping the inner edge near the fold, and saves the results as two new files (one ending in _left.png and one in _right.png). This procedure yields two primary advantages: (1) a reduced context window, which allows the LLM to operate more reliably, and (2) smaller file sizes, which marginally reduce processing time and

[5] This is done via pdf2image (Belval, 2026). I decided to do this manually via this module to ensure I can modify every parameter during conversion, including image quality, and to maintain the transparency of every step of my code.

computational cost in subsequent stages.

The proper format for the LLM use is to convert our preprocessed PNG images into *encoded text*, which essentially converts the pixel information from the PNG image into a format the machine can understand, including the patterns within. LLMs are compatible with such encodings (this is how LLMs *see*). This is done again with a simple Python package and requires little processing (Python Software Foundation, 2026). Each PNG image is converted into an encoded string.

At this point, we have a textual representation of the source material and can proceed with the first LLM-based task. There are several ways to extract text using LLMs. The first approach is to send the encoded images directly to a general-purpose LLM such as Gemini or GPT along with a prompt (Greif et al., 2025; Griesshaber and Streb, 2026; Lagakos et al., 2025). For instance, asking the model to extract every biography on the page. This requires prompt engineering, experimentation, and post-hoc validation. The second way, which is becoming more common, is to use a dedicated OCR LLM, such as those provided by Mistral or z.ai. The fundamental difference is that the latter simply recognizes every character and its location, as traditional OCR software does, and does not require any thinking from the LLM. I decided to go with *mistral-ocr-latest* because I was already familiar with their products, and it was the cheapest option to process these pages at the time. As mentioned before, I argue that workflows should be independent of endorsing specific products. For quality comparisons, one can refer to AI leaderboards or simply experiment with different solutions.[6] An important reminder is that the model's performance can vary dramatically across different source materials, and the decision for Mistral for this paper was also made after experimentation.

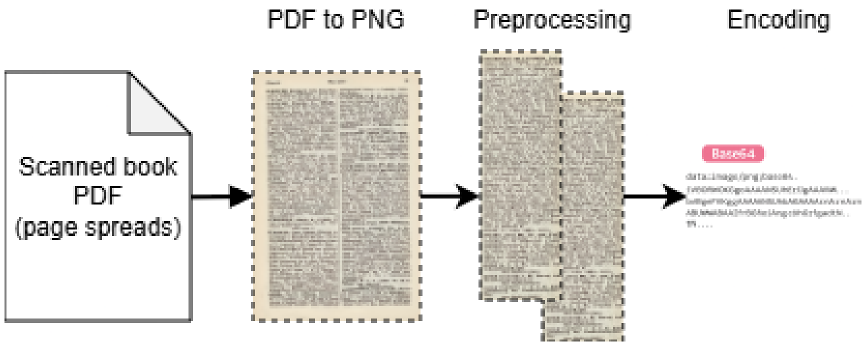
Because the book spans 1,686 pages, I process API calls in parallel to keep the total runtime manageable. After receiving responses for all relevant pages from Mistral,

[6]Although every company does its own benchmarks and claims performance superiority, it might be a good idea to experiment with leaderboards first, i.e., for OCR-oriented model comparisons, see OCR Arena (2026); Arena AI (2026).

raw model responses are stored in a JSONL database alongside the extracted text and metadata (page number, model version, and prompt configuration), so that any output can be traced back to the exact call that produced it and reprocessed if needed.

Finally, a forward-looking note is warranted. At the time of writing, there are no widely available, high-quality, and open-weight multimodal LLMs that perform reliably on both printed and non-printed documents. Many existing multimodal models are relatively small-scale and insufficiently trained for such tasks. This situation is likely to change soon, reducing reliance on cloud-based services. The final section of this paper will discuss which open-weight LLMs become viable and how they can be integrated into similar workflows.

Figure 3.3: The diagram for image preprocessing



Source: Author's own.

Biography Preparation and Sanity Checks. After the previous preprocessing step, we are left with the full text content of each page. To extract biographies from these pages and ensure we capture the correct span of text, I carefully assemble them line by line, using only Python. There are many minor details that require attention, including cases where a biographical entry spans more than one page, meaning it continues from one column or page to another. This step is relatively error-free, thanks to a simple principle: every biography starts with an uppercase surname and ends with a paragraph break. A quality screening flags entries that look unreliable: records with missing surnames, entries that are too short or too long, text that appears broken in the middle, or names that fall out of the expected alphabetical order. These flagged cases are set aside for human verification to keep the working dataset as dependable as possible for analysis. At the end, I identified 16,001 biographies and set aside 164 entries for human verification.

Data Extraction. After the biographies have been identified, the second component of the workflow begins. While this raw text is already valuable, we are interested in the structured variables that can be used for other purposes, such as addresses, occupations, and birth dates. I decided to extract the requested information from the questionnaire, namely the reader forms mentioned earlier[7].

The extraction process is divided into four steps, which are essentially the *agentic* part of this paper. Figure 3.4 depicts these four steps. The first step is what I call an *identifier* step. This LLM call sends the biography text and asks which variables are present in that biography. This step is useful because it grounds the next agent by indicating which variables are present and reduces the amount of thinking it must do. The other agents are also configured to be aware of the possibility of mistakes by previous agents. The result of the first step is 18 binary variables, indicating true for the present variables, and false for the ones that do not exist. The surname is 100 percent true for all biographies, as each biography starts with a surname. The least commonly identified variables are collections (6 percent) and ancestor information (11.4 percent).

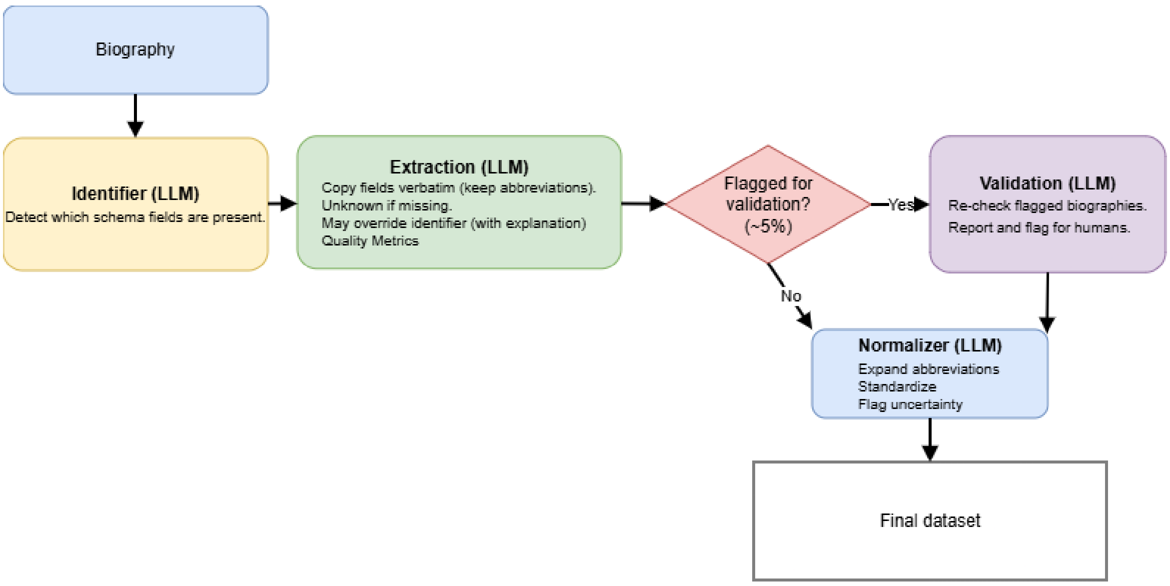
The second step is to extract the variables that were identified by the previous agent; this stage is called *extraction*. This step gets the relevant text exactly as it appears, so that all abbreviations remain as they are at this stage. If a variable does not exist, it is marked as "Unknown". The extraction agent can also override the identified fields, but it must provide an explanation when it does and must inform the researcher. This way, the potential problems caused by the identified fields can be mitigated, and, most importantly, we can track changes between agents. The extraction step also reports

[7]These are the following pieces of information on the form that Degener provides: Surname, First Name, Occupation, Birth Data and Place, Father Info, Mother Info, Ancestor Info, Education, Spouse, Children, Career Info, Publications or Works, Specialization, Hobbies, Collections, Personal Notes, Political Party, Memberships, and Current Address. See H.A.L. Degener (Ed.), Wer Ist's? Unsere Zeitgenossen, 5th Edition, Leipzig 1911.

quality measures, namely OCR confidence, the quality of variable separation, and reasoning why these variables are filled by it, if any issues were detected. The prompts are built conservatively, so usually there are more false negatives than positives.

This brings us to the validation step. The validation step is performed only on the biographies under scrutiny by the previous agents. In our case, around 9 percent of the biographies are flagged for validation, whether due to OCR issues, potential quality issues, missing variables, or mismatches between identified and extracted variables. The *validator* takes the biography, the decisions of previous agents, and the original image of the page where the biography is located, and determines whether the flagged issue should be kept as-is or corrected to the values it deems correct.

Figure 3.4: The four agentic extraction steps used to identify, extract, validate, and normalize variables from the biographies



Source: Author's own.

Finally, the last step is the *normalizer*. This step takes the full biography and the extracted variables and expands the abbreviations into clearer wording. It does this carefully and conservatively, and flags uncertainty if the abbreviations are not clear, or were flagged as OCR problems by previous agents. Initially, it seemed like a good idea to use the abbreviation list given in the book to expand all of them without the LLM. However, many abbreviations are ambiguous, and not to mention the potential OCR mistakes, historical spellings, and free-form fields like career. Rule-based abbreviations can help with birth dates, where one has to change the Roman numerals, but they struggle with fields such as career and publications, where we have too many idiosyncratic abbreviations. This is why I also included the last normalizer step to reach full biographies, but the workflow remains conservative to prevent semantic mistakes.[8]

The workflow is designed to generalize beyond just these biographical entries in *Wer Ist's?* Although this paper focuses on biographical entries and a fixed number of variables, the workflow should be portable to other document-extraction tasks because it separates four clear steps: (1) identifying variables, (2) extracting those fields as verbatim, (3) validating the quality, and (4) normalizing raw text. Adapting this workflow to a new source requires careful thought about what is necessary to keep. In practice, not all of these are necessary in every setting. Tabular data, for instance, is regular enough that the identifier step can be simplified or dropped entirely. Newspapers, by contrast, are heterogeneous enough that identification becomes more valuable, and a larger share of entries would likely require validation.

Model Choice. I experimented with several options for OCR and large language models.[9] A traditional OCR without an LLM, PaddleOCR, works well but requires significant post-processing. For economic historians who regularly build large-scale datasets with precision, this is not a valid option because hard-coding and manual corrections make the process almost impossible. One can implement a strategy like the one used by Gavin Greif, Niclas Griesshaber, and Robin Greif, using the output of PaddleOCR as input to an LLM, which should improve performance but might lead to hallucinations and a bias towards the LLM's ground truth. However, these methods

[8]Appendix C.2 compares repeated extractions from the *Wer Ist's?* biographies. The exercise shows that short, constrained fields are comparatively stable, while longer career fields are more sensitive to span selection, paraphrasing, and abbreviation expansion. This motivates keeping the normalization step conservative and auditable.

[9]Appendix C.3 shows how much generated occupation and social-standing scores change when prompts, models, temperatures, or seeds change. These scores are are used to show how prompt design, model choice, temperature, and seed settings affect LLM-generated measurement variables.

are quickly becoming outdated as LLM performance improves (Greif et al., 2025).

For the text recognition step, I ultimately chose a cloud-based LLM. For this, there were two realistic candidates given my experience with them. Mistral OCR offers excellent performance and is highly cost-effective, charging one Euro per 1,000 pages of OCR.

For the extraction steps, I decided to go with OpenAI's models for several reasons. First, OpenAI offers daily free usage, which is quite generous. This is not only helpful for reducing total costs, but also for experimenting without worrying about costs. Second, OpenAI APIs offer "*flex*" processing, which means reduced speed but a 50 percent discount on processing costs. Other companies offer similar schemes with batch processing, but these can take up to 24 hours. Flex processing, on the other hand, can reduce processing speeds; however, this is not noticeable for workflows such as this paper. Third, and the most important reason, was that I was used to their software development kit (SDK). I believe most researchers will get used to one or the other model and continue using it, unless the performance benefit of switching outweighs the costs. Again, these reasons should not be understood as endorsements. Models such as Gemini and its family could have done the same job, as the literature already shows. I argue that the important part of the model choice is to be transparent about it and to provide relevant information to make future researchers aware of the process.

Open-Weight. The replication could also have been done with a local solution. The local solutions, however, are not (1) as good as the cloud-based models, and (2) it is almost impossible to reach the speeds of the cloud service. I tested and evaluated several open-weight models across two systems. First, a workstation with an NVIDIA 5000 ADA GPU and 512 GB RAM, and second, the high-performance computer cluster at the University of Bayreuth, called *festus* (Bayreuth Centre for High Performance Computing, 2026). Still, both solutions are incredibly slow, as they need to process thousands of pages and requests one by one, whereas one can send 10,000 API calls to OpenAI and receive results in a few seconds. The bottom line is that both local and cloud solutions can be feasible for such workflows, albeit with varying qualities and speeds. A simple data extraction task from a relatively uncomplicated tabular structure may be accomplished with an open-weight model.[10] This trade-off does not need to be resolved the same way at every step. Because the workflow is modular, each stage can use a different model, whether local or cloud-based. In practice, I would recommend experimenting locally, where trying things costs little, and running the full pipeline through a cloud API, where parallel processing turns days of work into minutes. This balance will shift as open-weight models improve at reasoning tasks.

Transparency and Accessibility. LLM-based workflows can feel like a black box, which is why I make the code, prompts, logs, and all outputs openly available on GitHub, including the version history. Every change is recorded and reversible, and readers can inspect what changed between early drafts and the final version used in the paper. The four steps mentioned before, identification, extraction, validation, and normalization, are stored separately. In addition to the final dataset, I provide these intermediate outputs and their prior versions. Each biography has a unique identifier, so every dataset row can be traced back to the underlying biography text and original image. The repository includes a more detailed replication guide explaining how to rerun the workflow and generate a new version of the *Wer Ist's?* dataset.[11]

The workflow produces variables of two kinds, and they carry different risks. Extracted variables are copied from the page and can be checked against it: a birth year, an address, the honorifics in a name. Generated variables are the model's own inference. Some are simple, such as the gender dummy inferred from a first name; others are interpretive, such as a 0 to 100 occupation-prestige or social-standing score the

[10] The open-weight comparison in Appendix C.4 is a narrower score-generation comparison rather than a full replication of the *Wer Ist's?* pipeline. Its purpose is to show how local models behave relative to proprietary models when the same generated measurement task is repeated.

[11] Appendix C.1 gives a map of the files, logs, prompts, outputs, and rerun commands.

model assigns rather than reads. An extracted field is wrong only when the page is misread, whereas an interpretive score can move with the prompt, model, or temperature even when the page is read correctly. The regression in the next section uses extracted variables and one simple inferred dummy, so its robustness rests mainly on extraction stability (Appendix C.2). Appendix C then checks three risks for the harder interpretive scores: whether generated scores depend on prompt, model, temperature, or seed settings (Appendix C.3), whether open-weight models behave similarly (Appendix C.4), and whether occupation measures align with external or human benchmarks (Appendices C.5 and C.6).

3.6 Results

The primary output of the workflow described in Section 3.5 is a structured dataset of 16,001 biographical entries extracted from the 1911 edition of Degener's *Wer Ist's?*. Each entry corresponds to a single biography in the source and is stored as a row containing 18 variables, from basic demographic information to career details, family background, and contemporary address. The full pipeline ran in 66 minutes at a total API cost of 159.24 Euros. Before turning to specific applications of the data, I describe the dataset, its coverage, and its limitations.

Table 3.1 summarizes the coverage rates of the main variable groups. Across all 16,001 biographies and 18 variables, 57.09 percent of them are known, meaning that a typical biography has roughly 13 fields with non-missing values. This reflects the source material rather than extraction failures: many biographies simply do not contain information on every field because the respondents chose not to provide it. For example, information about ancestors is available in only 11 percent of the biographies. Because the pipeline defaults to marking a field as "Unknown" rather than guessing, as I described in the extraction and normalization steps of Section 3.5, these figures are likely a lower bound on actual coverage.

It is worth emphasizing that the biographies in *Wer Ist's?* are not representative of the general population. A typical entry concerns a person with some degree of public standing, and the coverage patterns that follow should be read with this selection in mind.

Since the book does not explicitly record gender, the workflow infers it from available first names during extraction. Over 92 percent of entries are classified as male, roughly 5 percent as female, and the remaining small percentage could not be identified. This distribution is consistent with the expected composition of public life in Germany at the time, where women's access to professional and public positions was restricted. The female entries are dominated by artists, writers, and singers.

Many respondents provided information about their family background, though coverage varies considerably across fields. Information about fathers appears roughly 10 percentage points more often than information about mothers. Spouse information falls between the two, and information about children is reported even less frequently.

Political party affiliation is reported for about 12 percent of entries, with the National Liberals and the Catholic Centre Party being the most frequently mentioned. This low overall rate likely reflects both the fact that many listed individuals were not politically active and the possibility that some respondents preferred not to disclose their affiliation. Memberships in societies and organizations are considerably more widespread, appearing in roughly 40 percent of biographies. The most frequently mentioned affiliations include the *Verein Berliner Presse*, the *Reichstag*, and various honorary knighthoods. Combined, over 43 percent of entries report either a party affiliation or an organizational membership.

Career information is among the most consistently available fields, albeit the most difficult one to extract and normalize. Publications and hobbies, by contrast, appear only when the respondent has written works or chooses to share personal interests. The least commonly populated field is collections, present in approximately 6 percent

Table 3.1: Coverage, classification, and workflow diagnostics

Variable or indicator	Count	Percent
Panel A. Generated Information		
Male	14,863	92.9%
Female	715	4.5%
Unknown	423	2.6%
Panel B. Extracted biography coverage		
Birth date	13,558	84.7%
Birthplace	13,041	81.5%
Address	14,543	90.9%
Education	10,848	67.8%
Career history	11,605	72.5%
Works/publications	9,791	61.2%
Political party	1,893	11.8%
Memberships	6,402	40.0%
Panel C. Family information		
Any family information reported	9,362	58.5%
Father	8,228	51.4%
Mother	6,740	42.1%
Spouse	7,195	45.0%
Children	4,925	30.8%
Ancestors	1,804	11.3%
Panel D. Workflow		
Explicit quality issue	556	3.5%
Routed for extraction review (automatic review by LLM)	13,197	82.5%
Panel E. Address and Occupation		
Street-level address	9,048	56.5%
Geocoded address	13,838	86.5%
Outside Modern Germany	4,839	35.0%
Occupation not stated	455	2.8%
Unclear, mixed, or title-only occupation	868	5.4%

Notes: The full dataset contains 16,001 biographies. In total, there are 352,022 possible biography fields. Percentages use all biographies as the denominator except Outside Modern Germany, which is based on geocoded addresses. Occupation not stated covers rows with unknown source occupation text. Review flags describe the workflow, not necessarily confirmed data errors. For more information on error rates, see the next subsection.

Source: Author's own.

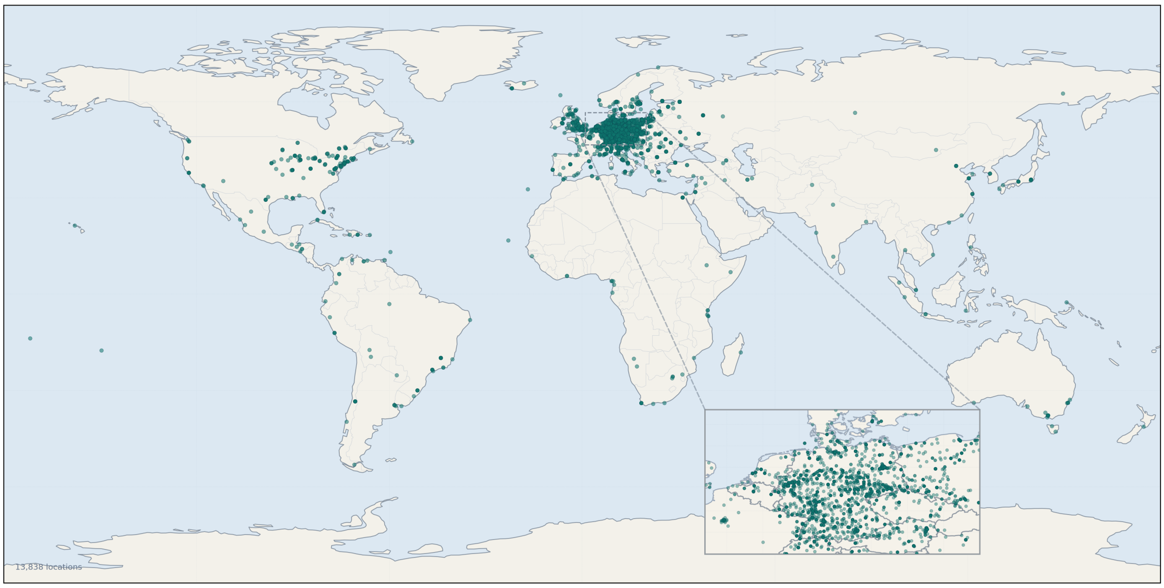
of entries.

The flagging system is designed to cast a broad net. Of the 16,001 entries, over 82 percent are routed for automatic review. This rate reflects the conservative prompting strategy described in Section 3.5: every entry where the extraction or normalization agents report even minor uncertainties is routed for review. Only about 4 percent of entries carry an explicit quality issue. As the next section shows, this rate is comparable to the functional error rate calculated by the human research assistants who evaluated a sample of this dataset. An additional 5 percent of entries are cross-reference entries that redirect to another biography rather than containing a full record.

Addresses. One of the highlights of the source material is the location information available for most biographical entries. Around 56 percent of the entries have detailed address information available at the street-number level, and around 86 percent have a location that can be geolocated. Locating these entries by how they are provided has a usual problem: most of the names are written in German, which makes one-to-one matching of their current international names difficult. I get around this problem by processing the addresses with an LLM to the smallest possible current international unit, usually a town or a city (*Lüttich* to Liège), and many locations simply no longer exist or exist under a different name (*Breslau* to Wrocław, and the street names in Wrocław are also different). This way, we get information on where these people lived, including their towns or cities, and can then geolocate their coordinates. This process uses the GeoNames web services (GeoNames, 2026a,b). The disadvantage of this method is that we lose information on the exact address for more than 50 percent of the biographies, but in return, we can locate them very closely. Both addresses are reported in the dataset. Given that pre-World War II addresses are notoriously difficult to match one-to-one in the territories of the German Empire, I believe this is a good compromise. Figure 3.5 represents the distribution of these addresses visually on a world map.

This geocoding process, too, is not without a human touch. Since smaller towns or old German names might not be reliably captured by the LLM, I also let it generate a confidence score between 0 and 1 and manually verify the addresses if the score is below 0.75. Only 4 percent of the addresses are flagged as low confidence.

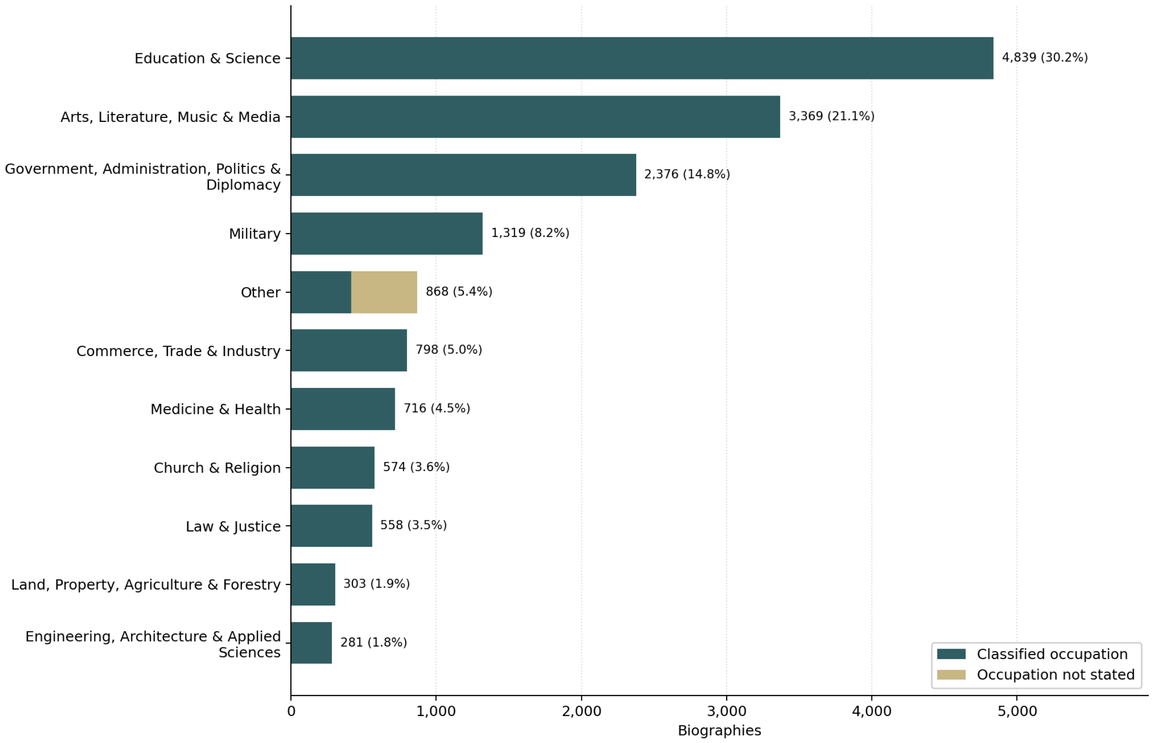
Figure 3.5: The distribution of addresses in the world



Notes: Although most addresses in this book are in the German Empire, 35 percent of the addresses are outside the modern Federal Republic of Germany. The inset shows the German Empire.
Source: Author's own.

Occupations. Occupation is one of the most informative dimensions of the extracted dataset. The classification procedure assigns an occupational category to every biography, though this should not be read as implying that every entry contains equally clear professional information. In a small number of cases, the underlying occupation text is unknown because it was left blank in the biography. For the vast majority, however, the category follows from substantive occupation text. The resulting distribution is concentrated in a few public-facing fields. Education and science account for nearly a third of all entries, followed by arts, literature, music, and media at roughly a fifth, and state administration, politics, and diplomacy at about 15 percent. Together, these three categories make up most of the directory and underline the extent to which *Wer Ist's?* captured a highly selective segment of educated, visible, and institutionally embedded contemporaries rather than any representative cross-section of society. The full distribution across categories is reported in Figure 3.6.

Figure 3.6: The distribution of occupations as categories



Notes: The categories are assigned by the LLM over all biographies. These are a coarse descriptive scheme the model assigns for summary purposes; the classification prompt and code are in the repository. Their accuracy is not separately benchmarked here beyond the overall extraction error rate. As similar evidence that the workflow can code occupations against an external standard, Appendix C.5 reports HISCO coding accuracy on a coded reference sample.
Source: Author's own.

Occupational coding in this source requires careful interpretation, because professional roles in the German Empire were often tied to rank, office, and honorific title. The classification step prioritizes the dominant occupational identity while retaining ambiguity when the source does not support a clear assignment. Roughly 5 percent of cases were classified as unclear, mixed, or title-only. Nobility and court-title markers are tracked separately and appear in about 10 percent and 15 percent of entries, respectively. The occupational results should therefore be understood as a disciplined simplification of a socially coded source: they provide a useful empirical summary of the directory's structure while preserving an explicit record of the cases in which status language is more salient than profession itself.[12]

[12]Appendix C.5 reports additional HISCO and occupational-prestige exercises. These exercises

Error Rates. One of the central questions is the errors LLMs make at each step of the process. The literature usually refers to the Character Error Rates for this calculation. The rise of LLMs blurs this metric to some extent, due to the sheer amount of data we can now process. One must imagine that, after digitizing millions of characters and encountering only a few punctuation errors, the resulting dataset yields a very low CER and does not reflect the actual quality of the dataset.

To tackle this problem, I decided to take a sample of 450 biographies (9,450 variables) and give the results to research assistants educated in historical methods and fluent in German. They were tasked to find (1) OCR Errors so a CER can be calculated, and (2) semantic mistakes during the process, meaning the variable extracted by the pipeline is wrong. The results confirmed the problem above: the CER for simple biographies was extremely low, as LLMs rarely make mistakes in those. Combined with the longer biographies, which are more likely to have critical problems, the CER for 9,450 variables (over 500,000 characters) was 1.07 percent[13].

When we look past this measure and calculate the results of the second task of the verification, the functional error rate, meaning the ratio of "wrong" variables[14] to all variables, was 4.49 percent, which is much more realistic and informative than the simple CER. These results reflect hallucinations, OCR errors, segmentation issues, and missed information. In summary, LLMs seem to be good enough, and sometimes even better than market solutions, in terms of error rates, but the pipelines that use them still need to be built carefully and scientifically to ensure reliable, transparent results.

Exploring Downstream Consequences. To economic historians, perhaps the most

are not drawn from the *Wer Ist's?* biographies and should not be read as direct validation of every occupation assigned in the Degener dataset. Instead, they show how the same staged classification procedure behaves on external historical occupation-coding tasks, which motivates caution when using LLMs to categorize text.

[13]The repository goes into a deeper discussion of calculation of CER and OCR mistakes, as this is not as straightforward as finding every single mistake, as it needs manual check for semantics. For further discussion of CER and OCR-error calculation, see Öztürk (2026a).

[14]Wrong means that the human did not agree with the pipeline's result.

relevant part of using LLMs as tools is the effects different workflows might have on outcomes. If we treat LLMs like any other tool on the market and expect reproducibility on par with other tools, then we must explore the potential consequences of LLMs' stochasticity.

To do this, using the dataset I extracted from the *Wer Ist's?*, I run a simple regression on the extracted dataset and then re-run it four more times on datasets produced by alternative extraction models. The aim is not to draw any substantive conclusion from either side of the equation; it is to ask whether the same specification, applied to data produced by five different model configurations, returns the same coefficients[15].

$$\Pr\left(\text{HasWiki}_i = 1 \mid X_i\right) = \Lambda\left(\alpha + \beta\text{Elite}_i + \gamma\text{Age}_i + \delta\text{Male}_i\right). \tag{3.1}$$

where HasWiki_i is a binary indicator for whether individual i has a Wikipedia page in German or English, and $\Lambda(\cdot)$ denotes the logistic cumulative distribution function (for details on how these are constructed, see Öztürk, 2026a). Elite_i indicates whether the person has honorific or status markers in their name, Age_i is the individual's age in 1911 computed from the extracted birth date, and Male_i is a gender dummy produced by the LLM. Each variable is chosen to stress a different part of the pipeline. The elite indicator depends on honorifics such as Doktor, Professor, Graf, and von, which are often abbreviated in the source and must be expanded by the normalization agent; any inconsistency in how the LLM resolves these abbreviations would directly affect who is classified as elite. Age is calculated from the extracted birth year, and because OCR can struggle with individual digits, a single misread character could shift a person's age by decades, making this variable sensitive to small extraction errors. Male_i is the only variable in the regression that is generated rather than extracted. If model choice or reasoning configuration affects how the LLM assigns gender to ambiguous or uncommon names, this should show up in the coefficient and

[15] I decided to sample 500 biographies that have information on all relevant variables, so that there is a clear comparison group. The sample size is kept small to keep replication costs low.

its standard error. Together, the three variables expose the pipeline to normalization errors, OCR noise, and inference, which is why they are useful for a robustness exercise.

Table 3.2: Regression results across extraction models

	GPT-5.1	GPT-5 Mini	Gemini-2.5 Flash	GPT-5 Nano	GPT-5 Nano
Elite (0/1)	-1.57***	-1.47***	-1.47***	-1.59***	-1.76***
	(0.25)	(0.25)	(0.24)	(0.26)	(0.45)
Age	0.00	0.00	-0.00	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Male (0/1)	-0.39	-0.28	-0.32	-0.34	0.77*
	(0.48)	(0.48)	(0.48)	(0.50)	(0.46)
Observations	500	500	500	500	500
Mean Dep. Var.	0.64	0.64	0.64	0.64	0.64
Pseudo R-Sq.	0.06	0.06	0.06	0.06	0.03
Reasoning	Default	Default	Default	Default	Minimal

Notes: Each column reports the results of estimating Equation (3.1) with the set of variables generated by the model indicated in the column header. The GitHub repository also includes results that demonstrate within-model stability. Standard errors are in parentheses. The final column uses reduced reasoning capacity to test the sensitivity of the estimates to model configuration. *p < 0.10, **p < 0.05, ***p < 0.01.
Source: Author's own.

Table 3.2 reports the results of estimating Equation (3.1) across five specifications, each generated by a different model or reasoning configuration. The result is reassuring: the elite coefficient is negative, statistically significant, and stable across the first four columns, ranging from roughly -1.5 to -1.6, whether the extraction was done by GPT-5.1, GPT-5 Mini, Gemini-2.5 Flash, or GPT-5 Nano with default reasoning. Age is effectively zero in all specifications, and the gender coefficient is small and insignificant. The estimates remain relatively consistent across models that differ in both design and price, suggesting that the careful pipeline design absorbs most of the randomness from individual models before it appears in the final results. In the last column, where the extraction is done with GPT-5 Nano with minimal reasoning, the elite coefficient shifts noticeably, the gender coefficient changes sign and becomes marginally significant, and the pseudo-R-squared drops by half. This is the kind of degradation one would expect when the extraction model is given less room to reason through ambiguous abbreviations and honorific titles, which are the most central variables to this regression. The comparison, therefore, illustrates both the robustness of the workflow under reasonable model choices and the point at which that robustness begins to break down.

3.7 Conclusion

This paper developed a transparent and reproducible LLM-based workflow for extracting structured data from historical print sources, and demonstrated it by digitizing Degener's *Wer Ist's?* (1911), a dense and heavily abbreviated source containing 16,001 biographies. The agentic workflow separates text recognition, variable identification, extraction, validation, and normalization into distinct and verifiable steps. The resulting workflow ran for 66 minutes at a cost of 159.24 Euros.

The key insight is that LLM workflows can be designed to be robust to imperfect steps through traceability and flagging. Each agent in the pipeline has a narrow task and logs its decisions, so the researcher can trace any output back to the original source and intervene where the system struggles. Validation and normalization extend the same logic: human effort is concentrated on flagged cases, and uncertainty is marked rather than guessed away. These design choices reflect a simple principle: it is better to flag a problem than to produce a wrong answer silently.

The regression exercise in Section 3.6 puts this approach to the test: it asks whether the same specification, when applied to datasets generated by different extraction models, yields the same results. Swapping the AI model produced broadly stable coefficient estimates, but with considerable differences when the reasoning capacity was throttled. This exercise shows that model choice is not irrelevant; a well-structured pipeline can absorb much of the stochasticity introduced by individual models.

Several limitations should be acknowledged. First, the exercises in this paper were designed with demonstration in mind and kept relatively simple. Future projects can strengthen this workflow by systematically testing LLM parameters such as temperature, seed, and reasoning level, and by benchmarking performance across a wider range of sources. The results of such tests might be more informative about the model choices and the workflow's limits.

Second, the dataset's human verification is based on a single random subsample of 450 biographies. The error rates this produced, 1.07 percent for CER and 4.49 percent for semantic mistakes, are informative for this pipeline and this source, but they are not benchmarks for LLM-based extraction in general. Both the number and type of errors depend on the source, the extraction task, the workflow design, and the models used. The contribution of this paper is therefore not the specific error rate, but the validation and flagging architecture that makes such errors observable.

Lastly, the workflow was developed for a specific type of source, dense biographical entries with a known structure, and adapting it to less structured documents, such as newspapers or handwritten records, will require modifications. Reliance on proprietary APIs introduces a further deprecation risk: if a model is retired or silently updated, exact replication becomes impossible, and archiving model version metadata mitigates but does not eliminate this concern. Despite these limitations, the broader contribution of this paper is methodological. The workflow's modular structure can be adapted to other document-extraction tasks in economic history, from financial records to census tables, provided each step is kept simple enough to understand and transparent enough to inspect. If transparent and reproducible workflows of this kind become routine, many historical sources that were previously too costly to digitize can now be considered for future work.

Conclusion

The first two chapters of this thesis exploited the historically determined Catholic-Protestant mosaic of the region at the parish level, first within Württemberg and then across Baden and Württemberg. The analysis of the fertility transition showed that Protestant parishes entered the decline about a decade earlier than otherwise similar Catholic neighbors, and that this gap is not explained by economic modernization, the gender wage gap, marriage behavior, infant mortality, or institutional differences. With these channels eliminated, and consistent with the accounts of contemporary observers, the most plausible interpretation is that denomination shaped the acceptability, and hence the timing, of deliberate fertility control within marriage. The analysis of political competition carried the same cleavage into the twentieth century. Because Catholic voters were organized around the *Zentrum* while Protestant voters lacked a comparable anchor, fragmentation was highest where a cohesive Catholic bloc coexisted with a dispersed Protestant electorate, and it was this fragmented Protestant vote that the NSDAP later consolidated. After 1945, the CDU offered both confessions a common anchor on the center-right and thereby weakened the link between religious composition and fragmentation. Yet the divide left traces: Catholic areas remained less receptive to the far right when the NPD emerged in the late 1960s, a persistence that survey evidence ties to church attachment and cues from religious authorities rather than to differences in nationalist or anti-establishment attitudes.

The third chapter addressed the foundation on which such analyses rest. Historical sources are rich in information but costly to convert into structured data, and AI tools that promise to lower this cost raise new problems of transparency, reproducibility, and verification. The agentic LLM workflow developed in the chapter answered these problems by separating text recognition, variable identification, extraction, validation, and normalization into auditable steps. Applied to Degener's *Wer Ist's?* of 1911, it produced a structured dataset of 16,001 biographies, demonstrating that large-scale historical data creation and methodological rigor are compatible.

Two broader lessons follow from these essays. First, long-run demographic and political change cannot be read off economic conditions alone. In the first two chapters, communities with similar economies diverged because religious norms and the institutions that carried them differed, and this divergence is visible only at a resolution that aggregated data would conceal. Second, the tools that now make such fine-grained evidence attainable at scale must meet the same standards of transparency and reproducibility as the analysis built on them.

Bibliography

Abramitzky, Ran; Mill, Roy; and Pérez, Santiago. Linking individuals across historical sources: A fully automated approach. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 53(2):94–111, 2020.

Acemoglu, Daron; Johnson, Simon; and Robinson, James A. The colonial origins of comparative development: An empirical investigation. American Economic Review, 91(5):1369–1401, 2001. doi: 10.1257/aer.91.5.1369.

Adena, Maja; Enikolopov, Ruben; Petrova, Maria; Santarosa, Veronica; and Zhuravskaya, Ekaterina. Radio and the rise of the Nazis in prewar Germany. The Quarterly Journal of Economics, 130(4):1885–1939, 2015.

Alsan, Marcella and Goldin, Claudia. Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. Journal of Political Economy, 127(2):586–638, April 2019. ISSN 1537-534X. doi: 10.1086/700766.

Arena AI. Vision arena ocr leaderboard, 2026. URL `https://arena.ai/leaderboard/vision/ocr`.

Arzheimer, Kai and Berning, Carl C. How the alternative for Germany (AfD) and their voters veered to the radical right, 2013–2017. Electoral Studies, 60:102040, 2019. ISSN 0261-3794. doi: 10.1016/j.electstud.2019.04.004. URL `https://www.sciencedirect.com/science/article/pii/S0261379418305158`.

Badischen Statistischen Landesamt. Badische Gemeindestatistik. Badischer Kommunalverlag, Karlsruhe, 1927.

Barrie, Christopher; Palmer, Alexis; and Spirling, Arthur. Replication for language models: Problems, principles, and best practices for political science, 2025. URL https://arthurspirling.org/documents/BarriePalmerSpirling_TrustMeBro.pdf. Working paper, draft of 4 November 2025. Accessed 12 May 2026.

Bartolini, Stefano and Mair, Peter. Identity, Competition, and Electoral Availability: The Stabilisation of European Electorates, 1885–1985. Cambridge University Press, Cambridge, 1990.

Bayreuth Centre for High Performance Computing. festus cluster, 2026. URL https://www.bzhpc.uni-bayreuth.de. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 523317330.

Beach, B. and Hanlon, W. W. Historical newspaper data: A researcher's guide. Explorations in Economic History, 90:101541, 2023a.

Beach, Brian and Hanlon, W Walker. Culture and the Historical Fertility Transition. The Review of Economic Studies, 90(4):1669–1700, July 2023b. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdac059. URL https://academic.oup.com/restud/article/90/4/1669/6678991.

Becker, Gary S. An economic analysis of fertility. In Demographic and economic change in developed countries, pages 209–240. Columbia University Press, 1960.

Becker, Sascha O. and Cinnirella, Francesco. Prussia disaggregated: the demography of its universe of localities in 1871. Journal of Demographic Economics, 86 (3):259–290, September 2020. ISSN 2054-0892, 2054-0906. doi: 10.1017/dem.2020.12. URL https://www.cambridge.org/core/product/identifier/S2054089220000127/type/journal_article.

Becker, Sascha O. and Voth, Hans-Joachim. From the death of god to the rise of Hitler. CESifo Working Paper Series 10730, CESifo, 2023.

Becker, Sascha O. and Woessmann, Ludger. Was Weber wrong? a human capital theory of Protestant economic history. Quarterly Journal of Economics, 124(2): 531–596, May 2009. ISSN 1531-4650. doi: 10.1162/qjec.2009.124.2.531.

Becker, Sascha O.; Cinnirella, Francesco; and Woessmann, Ludger. Does women's education affect fertility? Evidence from pre-demographic transition Prussia. European Review of Economic History, 17(1):24–44, February 2013. ISSN 1474-0044, 1361-4916. doi: 10.1093/ereh/hes017. URL `https://academic.oup.com/ereh/article-lookup/doi/10.1093/ereh/hes017`.

Becker, Sascha O.; Pfaff, Steven; and Rubin, Jared. Causes and consequences of the Protestant Reformation. Explorations in Economic History, 62:1–25, 2016. doi: 10.1016/j.eeh.2016.07.007.

Becker, Sascha O.; Rubin, Jared; and Woessmann, Ludger. Religion and growth. Journal of Economic Literature, 62(3):1094–1142, September 2024. doi: 10.1257/jel.20231666. URL `https://www.aeaweb.org/articles?id=10.1257/jel.20231666`.

Becker, Sascha O.; Boll, Paul David; and Voth, Hans-Joachim. Spatial unit roots in regressions: a practitioner's guide and a Stata package. CESifo Working Paper Series 11651, CESifo, 2025.

Beckmann, Christopher. Parteigründung in Ost und West. In Lammert, Norbert, editor, Handbuch zur Geschichte der CDU, pages 59–77. wbg Academic, Darmstadt, 2022.

Belval, E. pdf2image, 2026. URL `https://github.com/Belval/pdf2image`.

Besier, Gerhard. Christliche Parteipolitik und Konfession: Zur Entstehung des Evangelischen Arbeitskreises der CDU/CSU. Kirchliche Zeitgeschichte, 3(1):166–187, 1990.

Blanc, Guillaume. The cultural origins of the demographic transition in france, 2023. Working Paper.

Braun, Sebastian Till and Franke, Richard. Railways, growth, and industrialization in a developing german economy, 1829–1910. The Journal of Economic History, 82(4): 1183–1221, December 2022. ISSN 1471-6372. doi: 10.1017/s0022050722000377.

Brown, John C. and Guinnane, Timothy W. Fertility transition in a rural, Catholic population: Bavaria, 1880–1910. Population Studies, 56(1):35–49, January 2002. ISSN 0032-4728, 1477-4747. doi: 10.1080/00324720213799. URL `https://www.tandfonline.com/doi/full/10.1080/00324720213799`.

Brown, John C. and Guinnane, Timothy W. Regions and time in the European fertility transition: problems in the Princeton Project's statistical methodology. The Economic History Review, 60(3):574–595, August 2007. ISSN 0013-0117, 1468-0289. doi: 10.1111/j.1468-0289.2006.00371.x. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0289.2006.00371.x`.

Brynjolfsson, Erik; Korinek, Anton; and Agrawal, Ajay K. A research agenda for the economics of transformative ai. NBER Working Paper 34256, National Bureau of Economic Research, 2025. URL `https://doi.org/10.3386/w34256`.

Bundesamt für Kartographie und Geodäsie. Digital terrain model grid width 200m. DGM200, 2017. URL `http://www.bkg.bund.de`.

Bundesanstalt für Geowissenschaften und Rohstoffe. Bodenübersichtskarte der Bundesrepublik Deutschland 1:1.000.000, 2013. URL `https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/buek1000_node.html`.

Burnham, Walter Dean. Political immunization and political confessionalism: The United States and Weimar Germany. Journal of Interdisciplinary History, 3(1): 1–30, 1972. ISSN 0022-1953. doi: 10.2307/202460.

Cantoni, Davide. The economic effects of the protestant reformation: Testing the weber hypothesis in the german lands. Journal of the European Economic Association, 13(4):561–598, 2015. doi: 10.1111/jeea.12117. URL `https://doi.org/10.1111/jeea.12117`.

Chapman, Jonathan. The contribution of infrastructure investment to Britain's urban mortality decline, 1861–1900. The Economic History Review, 72(1):233–259, February 2019. ISSN 1468-0289. doi: 10.1111/ehr.12699.

Cinnirella, Francesco. Marital fertility and investment in children's education. In Diebolt, Claude; Rijpma, Auke; Carmichael, Sarah; Dilli, Selin; and Störmer, Charlotte, editors, Cliometrics of the Family, pages 33–54. Springer International Publishing, Cham, 2019. ISBN 9783319994802. doi: 10.1007/978-3-319-99480-2_3.

Cleland, John and Wilson, Christopher. Demand theories of the fertility transition: An iconoclastic view. Population Studies, 41(1):5–30, 1987. doi: 10.1080/0032472031000142516. URL `https://doi.org/10.1080/0032472031000142516`.

Coale, Ansley Johnson and Watkins, Susan Cotts. Decline of Fertility in Europe. Princeton University Press, 1986. ISBN 9780691629278.

Conley, Timothy G. and Kelly, Morgan. The standard errors of persistence. Journal of International Economics, 153:104027, January 2025. ISSN 0022-1996. doi: 10.1016/j.jinteco.2024.104027.

Cui, J. and Alexander, R. Same prompt, different outcomes: Evaluating the reproducibility of data analysis by llms, 2026. URL `https://doi.org/10.48550/arXiv.2602.14349`.

Cummins, Neil. Marital fertility and wealth during the fertility transition: rural France, 1750-1850: Fertility and Wealth in Transition-Era France. The Economic History Review, 66(2):449–476, May 2013. ISSN 00130117. doi: 10.1111/j.1468-0289.2012.00666.x. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0289.2012.00666.x`.

De Juan, Alexander; Haass, Felix; Koos, Carlo; Riaz, Sascha; and Tichelbaecker, Thomas. War and nationalism: How WW1 battle deaths fueled civilians' support for the Nazi Party. American Political Science Review, 118(1):144–162, 2024.

Degener, H. A. L., editor. Wer Ist's? Unsere Zeitgenossen. Leipzig, 5 edition, 1911.

Doepke, Matthias; Hannusch, Anne; Kindermann, Fabian; and Tertilt, Michèle. The economics of fertility: A new era. In Handbook of the Economics of the Family, volume 1, pages 151–254. Elsevier, 2023. doi: 10.1016/bs.hefam.2023.01.003.

Doerr, Sebastian; Gissler, Stefan; Peydró, José-Luis; and Voth, Hans-Joachim. Financial crises and political radicalization: How failing banks paved Hitler's path to power. The Journal of Finance, 77(6):3339–3372, 2022.

Elff, Martin. Social structure and electoral behavior in comparative perspective: The decline of social cleavages in Western Europe revisited. Perspectives on Politics, 5 (2):277–294, 2007. doi: 10.1017/S1537592707070788.

Elff, Martin and Rossteutscher, Sigrid. Stability or decline? class, religion and the vote in germany. German Politics, 20(1):107–127, 2011. doi: 10.1080/09644008. 2011.554109.

Falter, Jürgen W. Hitlers Wähler: Die Anhänger der NSDAP 1924–1933. Frankfurt, New York: Campus Verlag, 2020.

Falter, Jürgen W. and Hänisch, Dirk. Wahl- und Sozialdaten der Kreise und Gemeinden des Deutschen Reiches von 1920 bis 1933. GESIS Datenarchiv, Köln. ZA8013 Datenfile Version 1.0.0, https://doi.org/10.4232/1.8013, 1990a.

Falter, Jürgen W. and Hänisch, Dirk. Wählerbewegungen zum Nationalsozialismus. Wahl- und Sozialdaten der Gemeinden Badens in der zweiten Hälfte der Weimarer Republik. GESIS Datenarchiv, Köln. ZA8116 Datenfile Version 1.0.0, https://doi.org/10.4232/1.8116, 1990b.

Falter, Jürgen W. The social bases of political cleavages in the Weimar Republic, 1919-1933 (1992). Historical Social Research, Supplement, 25:194–216, 2013.

Fertig, Georg; Schlöder, Christian; Gehrmann, Rolf; Langfeldt, Christina; and Pfister, Ulrich. Das postmalthusianische Zeitalter: Die Bevölkerungsentwicklung in Deutschland, 1815–1871. Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, 105(1):6–33, 2018. ISSN 2365-2136. doi: 10.25162/vswg-2018-0001.

Ford, Robert and Jennings, Will. The changing cleavage politics of western europe. Annual Review of Political Science, 23(Volume 23, 2020):295–314, 2020. doi: 10.1146/annurev-polisci-052217-104957.

Galloway, Patrick R.; Hammel, Eugene A.; and Lee, Ronald D. Fertility Decline in Prussia, 1875–1910: A Pooled Cross-Section Time Series Analysis. Population Studies, 48(1):135–158, March 1994. ISSN 0032-4728, 1477-4747. doi: 10.1080/0032472031000147516. URL `http://www.tandfonline.com/doi/abs/10.1080/0032472031000147516`.

Galofré-Vilà, Gregori; Meissner, Christopher M; McKee, Martin; and Stuckler, David. Austerity and the rise of the Nazi Party. The Journal of Economic History, 81(1):81–113, 2021.

Galor, Oded. Unified Growth Theory. Princeton University Press, 2011. doi: 10.1515/9781400838868. URL `https://doi.org/10.1515/9781400838868`.

Galor, Oded. The demographic transition: causes and consequences. Cliometrica, 6(1):1–28, January 2012. ISSN 1863-2505, 1863-2513. doi: 10.1007/s11698-011-0062-7. URL `http://link.springer.com/10.1007/s11698-011-0062-7`.

Galor, Oded and Weil, David N. The gender gap, fertility, and growth. The American Economic Review, 86(3):374–387, 1996. ISSN 00028282. URL `http://www.jstor.org/stable/2118202`.

Galor, Oded and Weil, David N. Population, technology, and growth: from Malthusian stagnation to the demographic transition and beyond. American Economic Review, 90(4):806–828, September 2000. ISSN 0002-8282. doi: 10.1257/aer.90.4.806.

GeoNames. Geonames, 2026a. URL https://www.geonames.org/.

GeoNames. Geonames web services documentation, 2026b. URL https://www.geonames.org/export/web-services.html.

Goldberg, Andreas C. The evolution of cleavage voting in four western countries: Structural, behavioural or political dealignment? European Journal of Political Research, 59(1):68–90, 2020. doi: 10.1111/1475-6765.12336.

Goldin, Claudia. Understanding the Gender Gap: An Economic History of American Women. Oxford University Press, New York, 1990.

Gomez, Raul. When the chickens come home to roost: The long-term impact of party positions on religious voting. European Journal of Political Research, 61(2): 566–585, 2022. doi: 10.1111/1475-6765.12462. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.12462.

Greif, G.; Griesshaber, N.; and Greif, R. Multimodal llms for ocr, ocr post-correction, and named entity recognition in historical documents, 2025. URL https://doi.org/10.48550/arXiv.2504.00414.

Griesmeier, Josef. Die Entwicklung der Wirtschaft und der Bevölkerung von Baden und Württemberg im 19. und 20. Jahrhundert. Ein statistischer Rückblick auf die Zeit des Bestehens der Länder Baden und Württemberg. Jahrbücher für Statistik und Landeskunde von Baden-Württemberg, 1:121–242, 1954.

Griesshaber, N. and Streb, J. Multimodal llms for historical dataset construction from archival image scans: German patents (1877-1918). SAFE Working Paper 466, SAFE, Frankfurt a. M., 2026. URL https://hdl.handle.net/10419/335887.

Guinnane, Timothy W. The Historical Fertility Transition: A Guide for Economists. Journal of Economic Literature, 49(3):589–614, September 2011. ISSN 0022-0515. doi: 10.1257/jel.49.3.589. URL https://www.aeaweb.org/articles?id=10.1257/jel.49.3.589.

Guinnane, Timothy W.; Okun, Barbara S.; and Trussell, James. What Do We Know About the Timing of Fertility Transitions in Europe? Demography, 31(1):1–20, February 1994. ISSN 0070-3370, 1533-7790. doi: 10.2307/2061905. URL https://read.dukeupress.edu/demography/article/31/1/1/170921/What-Do-We-Know-About-the-Timing-of-Fertility.

Guriev, Sergei and Papaioannou, Elias. The political economy of populism. Journal of Economic Literature, 60(3):753–832, September 2022. doi: 10.1257/jel.20201595. URL https://www.aeaweb.org/articles?id=10.1257/jel.20201595.

Haffert, Lukas. The long-term effects of oppression: Prussia, political catholicism, and the Alternative für Deutschland. American Political Science Review, 116(2):595–614, 2022. doi: 10.1017/S0003055421001040.

Hermens, Ferdinand Aloys. Demokratie oder Anarchie? Untersuchung über die Verhältniswahl. VS Verlag für Sozialwissenschaften, Wiesbaden, 1968. ISBN 978-3322979292.

Immerzeel, Tim; Jaspers, Eva; and Lubbers, Marcel. Religion as catalyst or restraint of radical right voting? West European Politics, 36(5):946–968, 2013. doi: 10.1080/01402382.2013.797235.

Jones, Larry Eugene. "the dying middle": Weimar germany and the fragmentation of bourgeois politics. Central European History, 5(1):23–54, 1972.

Kalyvas, Stathis N. The Rise of Christian Democracy in Europe. Cornell University Press, Ithaca and London, 1996.

Katein, Werner. Das verhältnis von staat, kirche und volksschule im königreich württemberg. Zeitschrift für württembergische Landesgeschichte, 15:53–117, 1956.

Kersting, Felix; Wohnsiedler, Iris; and Wolf, Nikolaus. Weber revisited: The Protestant ethic and the spirit of nationalism. The Journal of Economic History, 80(3):710–745, August 2020. ISSN 1471-6372. doi: 10.1017/s0022050720000364.

King, Gary; Rosen, Ori; Tanner, Martin; and Wagner, Alexander F. Ordinary economic voting behavior in the extraordinary election of Adolf Hitler. The Journal of Economic History, 68(4):951–996, 2008.

Klingemann, Hans-Dieter. Politische und soziale Bedingungen der Wählerbewegungen zur NPD: Fallstudie Baden-Württemberg. In Sozialwissenschaftliches Jahrbuch für Politik, volume 2, pages 563–601. Olzog, München, 1971.

Klingemann, Hans-Dieter. Landtagswahl in Baden-Württemberg 1968 (Nachuntersuchung). (ZA0327; Version 2.0.0) [Data set]. GESIS, Köln. https://doi.org/10.4232/1.11197, 2012.

Klüsener, Sebastian; Dribe, Martin; and Scalone, Francesco. Spatial and Social Distance at the Onset of the Fertility Transition: Sweden, 1880–1900. Demography, 56(1):169–199, February 2019. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-018-0737-9. URL `https://read.dukeupress.edu/demography/article/56/1/169/167954/Spatial-and-Social-Distance-at-the-Onset-of-the`.

Knodel, John. Law, marriage and illegitimacy in nineteenth-century Germany. Population Studies, 20(3):279–294, March 1967. ISSN 1477-4747. doi: 10.1080/00324728.1967.10409964.

Knodel, John E. The Decline of Fertility in Germany, 1871-1939. Princeton University Press, Princeton, NJ, 1974.

Koenig, Christoph. Loose cannons: war veterans and the erosion of democracy in Weimar Germany. The Journal of Economic History, 83(1):167–202, 2023.

Kommission für geschichtliche Landeskunde in Baden-Württemberg. Historischer Atlas von Baden-Württemberg. Die Kommission, Stuttgart, 1988. ISBN 3-921201-10-1. Mehrbändiges Werk / 11 Lieferungen (1972–1988). Wissenschaftliche Gesamtleitung: Karl Heinz Schröder.

Kommission für geschichtliche Landeskunde in Baden-Württemberg. Historischer Atlas von Baden-Württemberg. Karte II,2: Gemeindegrenzenkarte, 1972. URL `https://www.leo-bw.de/themen/historischer-atlas-von-baden-wurttemberg`.

Kommission für geschichtliche Landeskunde in Baden-Württemberg. Historischer Atlas von Baden-Württemberg. Karte IV,4: Städte des Mittelalters. 1973. URL `https://www.leo-bw.de/themen/historischer-atlas-von-baden-wurttemberg`.

Kommission für geschichtliche Landeskunde in Baden-Württemberg. Historischer Atlas von Baden-Württemberg. Karte XI,2: Marktorte des Spätmittelalters und der frühen Neuzeit 1250-1828. 1982. URL `https://www.leo-bw.de/themen/historischer-atlas-von-baden-wurttemberg`.

Kommission für geschichtliche Landeskunde in Baden-Württemberg. Historischer Atlas von Baden-Württemberg. Beiwort zu den Karten XII, 2 & 4: Bevölkerungsdichte der Gemeinden 1834 und Bevölkerungsdichte der Gemeinden 1970. 1988. URL `https://www.leo-bw.de/themen/historischer-atlas-von-baden-wurttemberg`.

Königliches Statistisches Landesamt. Königlich-Württembergisches Hof- und Staats-Handbuch. Johann Friedrich Steinkopf, Stuttgart, 1824.

Königliches Statistisches Landesamt. Die Standorte der Gewerbe Württembergs nach Gemeinden am 14. Juni 1895: Gewerbetopographie. W. Kohlhammer Verlag, Stuttgart, 1900.

Königliches Statistisches Landesamt. Württembergische Gemeindestatistik. Zweite Ausgabe nach dem Stand vom Jahr 1907. W. Kohlhammer Verlag, Stuttgart, 1910.

Kull. Beiträge zur Statistik der Bevölkerung des Königreichs Württemberg. Württembergische Jahrbücher für Statistik und Landeskunde, 1874(1):1–232, 1875.

Kunz, Andreas and Zipf, Alexander. HGIS Germany: Historical information system of the German states, 1820 to 1914. Technical report, Leibniz-Institut für Europäische Geschichte and i3mainz - Institut für Raumbezogene Informations- und Messtechnik, 2008.

Laakso, Markku and Taagepera, Rein. "Effective" number of parties: A measure with application to West Europe. Comparative Political Studies, 12(1):3–27, 1979. doi: 10.1177/001041407901200101.

Lagakos, D.; Michalopoulos, S.; and Voth, H.-J. American life histories. NBER Working Paper 33373, National Bureau of Economic Research, 2025. URL `https://doi.org/10.3386/w33373`.

Landesarchiv Baden-Württemberg. LEO-BW: Landeskunde entdecken online. `https://www.leo-bw.de`, 2026. Online portal, accessed 7 May 2026.

Lepsius, M. Rainer. Parteiensystem und Sozialstruktur: Zum Problem der Demokratisierung der deutschen Gesellschaft. In Abel, Wilhelm et al., editors, Wirtschaft, Geschichte und Wirtschaftsgeschichte: Festschrift zum 65. Geburtstag von Friedrich Lütge, pages 371–393. Gustav Fischer Verlag, Stuttgart, 1966.

Lesthaeghe, Ron and Wilson, Chris. Chapter 6: Modes of Prodcution, Secularization, and the Pace of the Fertility Decline in Western Europe, 1870-1930. In The decline of fertility in Europe, pages 261–292. Princeton University Press, 1986.

Liepelt, Klaus. Anhänger der neuen Rechtspartei: Ein Beitrag zur Diskussion über das Wählerreservoir der NPD. Politische Vierteljahresschrift, 8(2):237–271, 1967.

Lipset, Seymour Martin and Rokkan, Stein. Cleavage structures, party systems, and voter alignments. an introduction. In Lipset, Seymour Martin and Rokkan, Stein, editors, Party Systems and Voter Alignments. Cross"=National Perspectives, pages 1–64. Free Press, New York, 1967.

Lublin, David. Electoral systems, ethnic heterogeneity and party system fragmentation. British Journal of Political Science, 47(2):373–389, 2017. doi: 10.1017/S0007123415000137.

Luchman, Joseph N. Determining relative importance in Stata using dominance analysis: domin and domme. The Stata Journal, 21(2):510–538, 2021. doi: 10.1177/1536867X211025837. URL `https://doi.org/10.1177/1536867X211025837`.

Ma, Z.; Santos, J. E.; Lackey, G.; Viswanathan, H.; and O'Malley, D. Information extraction from historical well records using a large language model. Scientific Reports, 14:31702, 2024. doi: 10.1038/s41598-024-81846-5. URL `https://doi.org/10.1038/s41598-024-81846-5`.

Madsen, Jakob and Strulik, Holger. Testing unified growth theory: Technological progress and the child quantity-quality tradeoff. Quantitative Economics, 14(1): 235–275, 2023. ISSN 1759-7323. doi: 10.3982/qe1751.

Marcinkiewicz, Kamil and Dassonneville, Ruth. Do religious voters support populist radical right parties? opposite effects in Western and East-Central Europe. Party Politics, 28(3):444–456, 2022. doi: 10.1177/1354068820985187.

Marquardt, Ernst. Geschichte Württembergs. Deutsche Verlags-Anstalt, Stuttgart, 1985.

Max Planck Institute for Demographic Research (MPIDR) and Chair for Geodesy and Geoinformatics, University of Rostock (CGG). MPIDR Population History GIS Collection (partly based on Bundesamt für Kartographie und Geodäsie), 2011. URL `http://censusmosaic.org/`. Mosaic Project.

McQuillan, Kevin. When does religion influence fertility? Population and

Development Review, 30(1):25–56, 2004. doi: 10.1111/j.1728-4457.2004.00002.x. URL https://doi.org/10.1111/j.1728-4457.2004.00002.x.

Messeri, Lisa and Crockett, M. J. Artificial intelligence and illusions of understanding in scientific research. Nature, 627:49–58, 2024. doi: 10.1038/s41586-024-07146-0.

Mitchell, Maria D. The Origins of Christian Democracy: Politics and Confession in Modern Germany. University of Michigan Press, Ann Arbor, 2012. doi: 10.3998/mpub.323749.

Mokyr, Joel. A Culture of Growth: The Origins of the Modern Economy. Princeton University Press, Princeton, NJ, 2016.

Mooney, Graham. Diagnostic spaces: Workhouse, hospital, and home in mid-victorian london. Social Science History, 33(3):357–390, 2009. doi: 10.1215/01455532-2009-005.

Mooney, Graham. Mooney_1866_London_occupational_codes. IISH Data Collection, dataset, version 1.2, 2016. URL https://hdl.handle.net/10622/ERGY0V. CC BY-SA 4.0.

Mozaffar, Shaheen; Scarritt, James R.; and Galaich, Glen. Electoral institutions, ethnopolitical cleavages, and party systems in africa's emerging democracies. American Political Science Review, 97(3):379–390, 2003. doi: 10.1017/S0003055403000753.

Müller, Ulrich K. and Watson, Mark W. Spatial correlation robust inference in linear regression and panel models. Journal of Business & Economic Statistics, 41(4):1050–1064, 2023. ISSN 1537-2707. doi: 10.1080/07350015.2022.2127737.

Müller, Ulrich K. and Watson, Mark W. Spatial unit roots and spurious regression. Econometrica, 92(5):1661–1695, 2024. ISSN 0012-9682. doi: 10.3982/ecta21654.

Nord, Marina; Altman, David; Fernandes, Tiago; Good God, Ana; and Lindberg, Staffan I. Democracy report 2026: Unraveling the democratic era? Technical report, V-Dem Institute, University of Gothenburg, 2026. URL `https://www.v-dem.net/publications/democracy-reports/`.

OCR Arena. Ocr arena leaderboard, 2026. URL `https://www.ocrarena.ai/leaderboard`.

Offenbacher, Martin. Konfession und soziale Schichtung: Eine Studie über die wirtschaftliche Lage der Katholiken und Protestanten in Baden. J. C. B. Mohr (Paul Siebeck), Tübingen and Leipzig, 1900.

Ordeshook, Peter C. and Shvetsova, Olga V. Ethnic heterogeneity, district magnitude, and the number of parties. American Journal of Political Science, 38(1):100–123, 1994.

Öztürk, Timur. Degener digitization workflow repository, 2026a. URL `https://github.com/timuroeztuerk/degener`.

Öztürk, Timur. Degener interactive website, 2026b. URL `https://timurozturk.de/degener/`.

Putzger, Friedrich Wilhelm. Putzger historischer Schul-Atlas. via Wikipedia Common, 1905.

Python Software Foundation. base64 - base16, base32, base64, base85 data encodings, 2026. URL `https://docs.python.org/3/library/base64.html`.

Raymond, Christopher D. Not all social cleavages are the same: On the relationship between religious diversity and party system fragmentation. Politics and Religion, 9(2):364–388, 2016. doi: 10.1017/S1755048316000250.

Rothenbacher, Franz and Fertig, Georg. Bevölkerung, haushalt und familien. In Rahlf, Thomas, editor, Deutschland in Daten: Zeitreihen zur historischen Statistik, pages 30–46. Bundeszentrale für politische Bildung, Bonn, 2., aktualisierte auflage edition, 2022.

Rudzio, Wolfgang. Das politische System der Bundesrepublik Deutschland. Wiesbaden: Springer, 2019.

Sauer, Paul. Reformer auf dem Königsthron: Wilhelm I. von Württemberg. Deutsche Verlags-Anstalt, Stuttgart, 1997. ISBN 3-421-05084-8.

Schaab, Meinrad. Die Anfänge einer Landesstatistik im Herzogtum Württemberg in den Badischen Markgrafschaften und in der Kurpfalz. Zeitschrift für Württembergische Landesgeschichte, 26:89–112, 1967.

Seybold, Gerhard. Württembergs Industrie und Außenhandel vom Ende der Napoleonischen Kriege bis zum Deutschen Zollverein, volume 74 of Veröffentlichungen der Kommission für geschichtliche Landeskunde in Baden-Württemberg Reihe B, Forschungen. W. Kohlhammer Verlag, Stuttgart, 1974.

Shen, Z.; Zhang, R.; Dell, M.; Lee, B. C. G.; Carlson, J.; and Li, W. Layoutparser: A unified toolkit for deep learning based document image analysis. In Lladós, J.; Lopresti, D.; and Uchida, S., editors, Document Analysis and Recognition - ICDAR 2021, pages 131–146, Cham, 2021.

Shively, W. Phillips. Party identification, party choice, and voting stability: The Weimar case. American Political Science Review, 66(4):1203–1225, December 1972. ISSN 1537-5943. doi: 10.2307/1957174.

Smith, Gordon. The german volkspartei and the career of the catch-all concept. In Döring, Herbert and Smith, Gordon, editors, Party Government and Political Culture in Western Germany, pages 59–76. Palgrave Macmillan UK, London, 1982.

Spenkuch, Jörg L and Tillmann, Philipp. Elite influence? religion and the electoral success of the Nazis. American Journal of Political Science, 62(1):19–36, 2018.

Spolaore, Enrico and Wacziarg, Romain. Fertility and Modernity. The Economic Journal, 132(642):796–833, February 2022. ISSN 0013-0133, 1468-0297. doi: 10.1093/ej/ueab066. URL https://academic.oup.com/ej/article/132/642/796/6357639.

Statistisch-Topographisches Bureau. Württembergische Jahrbücher für Statistik und Landeskunde. 1874. Stuttgart: H. Lindemann, 1875.

Statistisches Landesamt Baden-Württemberg. Volkszählungen in Württemberg 1834 bis 1925, 2008. URL https://www.statistik-bw.de/Service/Veroeff/Statistische_Daten/900208001.bs.

Statistisches Landesamt Baden-Württemberg. 150 Jahre Amtliche Statistik in Baden-Württemberg. Stuttgart: Statistisches Landesamt Baden-Württemberg, 1970.

Statistisches Reichsamt. Statistisches Jahrbuch für das Deutsche Reich 1934. Berlin: Reimar Hobbing Verlag, 1934.

United Nations Department of Economic and Social Affairs, Population Division. World population prospects 2024: Summary of results. Technical report, United Nations, New York, 2024. URL https://population.un.org/wpp/.

von Hippel, Wolfgang. Auswanderung aus Südwestdeutschland: Studien zur württembergischen Auswanderung und Auswanderungspolitik im 18. und 19. Jahrhundert. Industrielle Welt; 36. Klett-Cotta, Stuttgart, 1984.

Voth, H.-J. and Yanagizawa-Drott, D. Image(s). CEPR Discussion Paper 19219, Centre for Economic Policy Research, Paris, 2024. URL https://cepr.org/publications/dp19219.

Wanamaker, Marianne H. Industrialization and Fertility in the Nineteenth Century: Evidence from South Carolina. The Journal of Economic History, 72(1):168–196, March 2012. ISSN 0022-0507, 1471-6372. doi: 10.

1017/S0022050711002476. URL https://www.cambridge.org/core/product/identifier/S0022050711002476/type/journal_article.

Weber, Max. Die protestantische Ethik und der "Geist" des Kapitalismus. Archiv für Sozialwissenschaft und Sozialpolitik, 20:1–54, 1904.

Weir, Todd H. and Greenberg, Udi. Religious cultures and confessional politics. In The Oxford Handbook of the Weimar Republic. Oxford University Press, 01 2022. ISBN 9780198845775. doi: 10.1093/oxfordhb/9780198845775.013.28. URL https://doi.org/10.1093/oxfordhb/9780198845775.013.28.

Wolf, Julius. Der Geburtenrückgang. Die Rationalisierung des Sexuallebens in unserer Zeit. Verlag von Gustav Fischer, Jena, 1912a.

Wolf, Julius. Kinderzahl und Sozialdemokratie in Deutschland. In Die Volkswirtschaft: der Gegenwart und Zukunft, pages 285–298. A. Deichert'sche Verlagsbuchhandlg., Leipzig, 1912b.

Württembergischen Statistischen Landesamt. Die Ergebnisse der Württembergischen Landtagswahl vom 24. April 1932. Kohlhammer, Stuttgart, 1932.

Württembergischen Statistischen Landesamt. Württembergische Gemeinde- und Bezirkstatistik. Kohlhammer, Stuttgart, 1935.

Württembergisches Statistische Landesamt. Württembergische Gemeinde- und Bezirksstatistik. Dritte Ausgabe nach dem Stand vom Jahre 1933. Stuttgart: W. Kohlhammer, 1935.

Zimmermann, Wolfgang; Däumling, Gabriele; Grosse, Julia; Prangen, Melanie; Jautz, Andrea; Koch-Richter, Regina; Hierath, Claudia; Heizmann, Bodo; and Lenz, Florian. Censuses in Wuerttemberg, 1834 to 1925. GESIS, Cologne. ZA8654 Data file Version 1.0.0, https://doi.org/10.4232/1.14072, 2023.

Appendix A

Protestantism, Industrialization, and the Fertility Decline: Evidence from Religious Enclaves in Württemberg, 1871-1933

A.1 Data Sources

This section provides a detailed description of the sources, scope, and construction of variables in our data.

Demography:

Live births: Annual number of live births from 1871 to 1946, digitized from archival records (*StAL E 258 VII Bü 122*).

Stillbirths: Annual number of stillbirths from 1871 to 1930, digitized from archival records (*StAL E 258 VII Bü 119*).

Out-of-wedlock births: Annual number of out-of-wedlock births from 1871 to 1946, digitized from archival records (*StAL E 258 VII Bü 118*).

Deaths: Annual number of deaths from 1871 to 1946, digitized from archival records (*StAL E 258 VII Bü 120*). Figures from 1914 onward also separately report the number of war dead.

Infant deaths: Annual number of children who died in their first year of life from 1871 to 1946, digitized from archival records (*StAL E 258 VII Bü 120*).

Population: Total population in 1871, 1875, 1880, 1885, 1890, 1895, 1900, 1905, 1910, 1919, 1925, and 1933, as reported in the respective population census Statistisches Landesamt Baden-Württemberg (2008); Württembergisches Statistische Landesamt (1935). The censuses from 1871 to 1919 recorded the population present in a locality area (*ortsanwesende Bevölkerung*), the censuses of 1925 and 1933 the residential population (*Wohnbevölkerung*). We interpolate population for years between census dates.

Crude birth rate: The number of births per 1,000 population from 1871 to 1930. As an alternative indicator, we use the number of live births per 1,000 population, which we have for 1871 to 1933.

Crude death rate: The number of deaths per 1,000 population from 1871 to 1933.

Infant mortality: The number of children who die in their first year of life over the number of live births.

Fertility transition date: First date for which the 5-year moving average of the crude birth rate is at least 10% below its pre-transition level and never returns to the pre-transition level. We use two measures of pre-transition fertility. The baseline measure is the average crude birth rate of the county (*Oberamt*) in which a parish is located from 1812 to 1866, taken from (Kull, 1875). The second measure is the average crude birth rate at the parish level in the first five years of our parish data, but excluding the years of unusually high fertility from 1872 to 1877. In other words, this second measure of pre-transition fertility is the average crude birth rate of a parish in 1871, 1878, 1879, 1880, and 1881.

Fertility transition status: Indicator for years at or after a parish's fertility transition date.

Religion:

Protestants, Catholics: Number of Protestants and Catholics in the total population for 1822, 1843, and 1871. Figures for 1822 are from Königliches Statistisches Landesamt (1824); figures for 1843 and 1871 are from the respective population censuses Statistisches Landesamt Baden-Württemberg (2008).

Protestant parish: Indicator for parishes with a population share of Protestants of 90% or more in 1843.

Catholic parish: Indicator for parishes with a population share of Catholics of 90% or more in 1843.

Mixed parishes: Indicator for parishes where neither Protestants nor Catholics make up 90% or more of the population in 1843.

Industrialization:

Steam engine: Indicator for having at least one steam engine installed in 1869 from archival records (*StAl E 170 Bü 272*).

Total steam power installed: Total steam power installed per 1,000 persons in 1869 from archival records (*StAl E 170 Bü 272*).

Industrial employment: Number of full-time employees in industry per 100 persons from the 1895 and 1907 occupational censuses (Königliches Statistisches Landesamt, 1900, 1910).

Control variables:

Altitude: Average parish elevation above sea level (meters), based on Bundesamt für Kartographie und Geodäsie (2017).

Ruggedness: Parish terrain ruggedness, measured from the variation in parish-level elevation using Bundesamt für Kartographie und Geodäsie (2017).

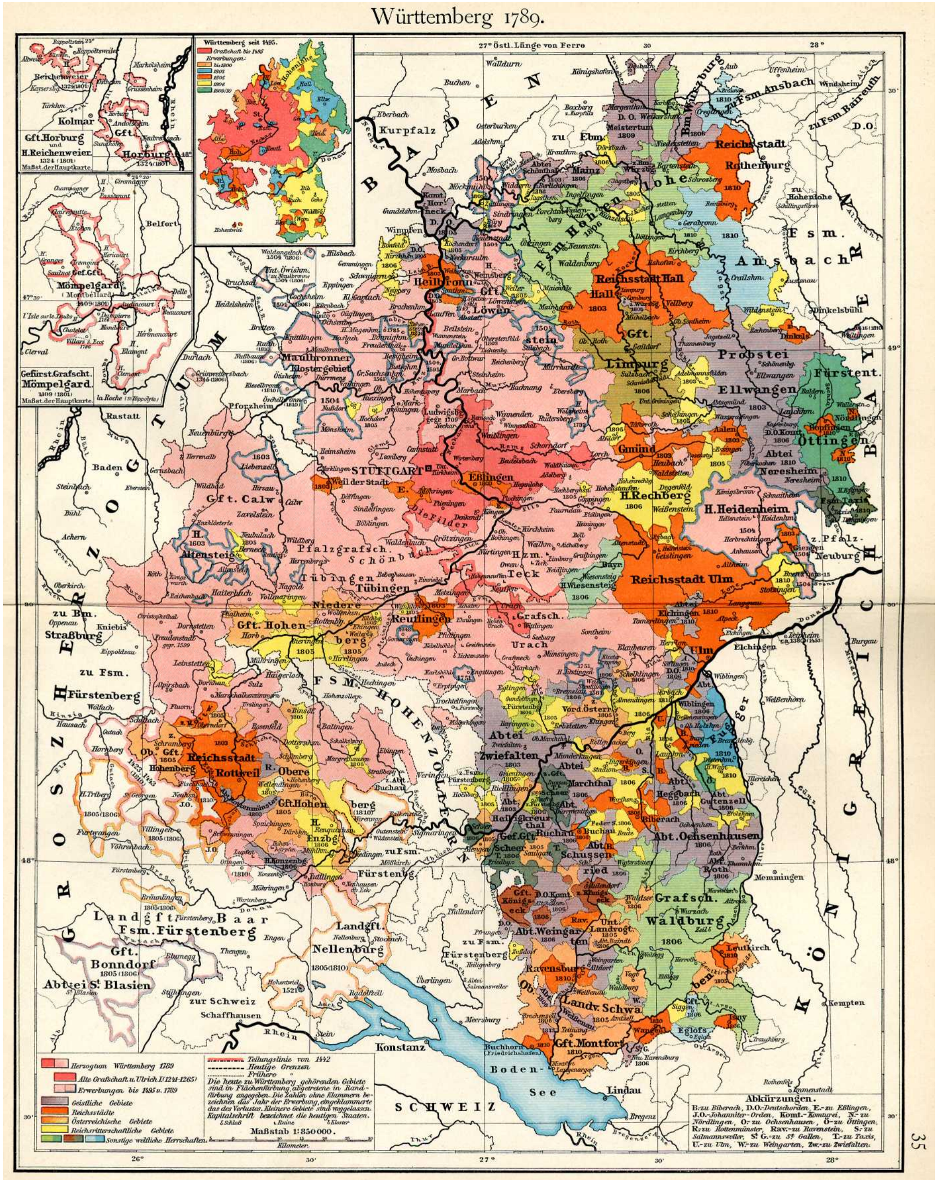
Soil quality: Soil quality indicator from Bundesanstalt für Geowissenschaften und Rohstoffe (2013) soil map (*Bodenübersichtskarte*) used by the 6-class soil coding.

Pre-Reformation city rights: Indicator for parishes with city rights as of 1518 (Kommission für geschichtliche Landeskunde in Baden-Württemberg, 1973).

Pre-Reformation market rights: Indicator for parishes with market rights as of 1518 (Kommission für geschichtliche Landeskunde in Baden-Württemberg, 1982).

A.2 Additional Maps

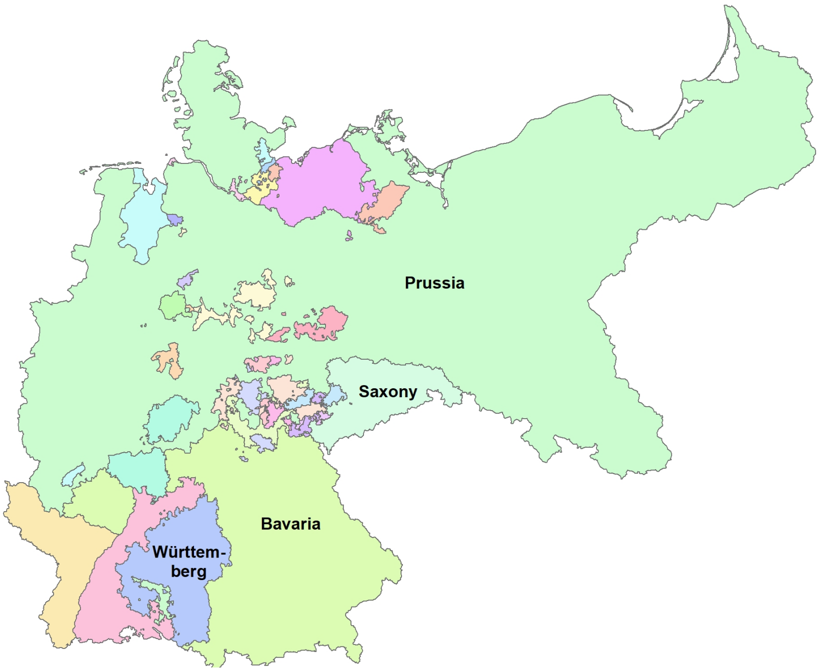
Figure A.1: Württemberg in 1789



Notes: The map shows the territorial division, as of 1789, of the lands that later became part of the Kingdom of Württemberg.

Source: Putzger (1905)

Figure A.2: The German Empire in 1871



Notes: The figure shows the German Empire in its 1871 borders. Labels mark the four Kingdoms that were part of the German Empire (namely, the Kingdoms of Bavaria, Prussia, Saxony and Württemberg).

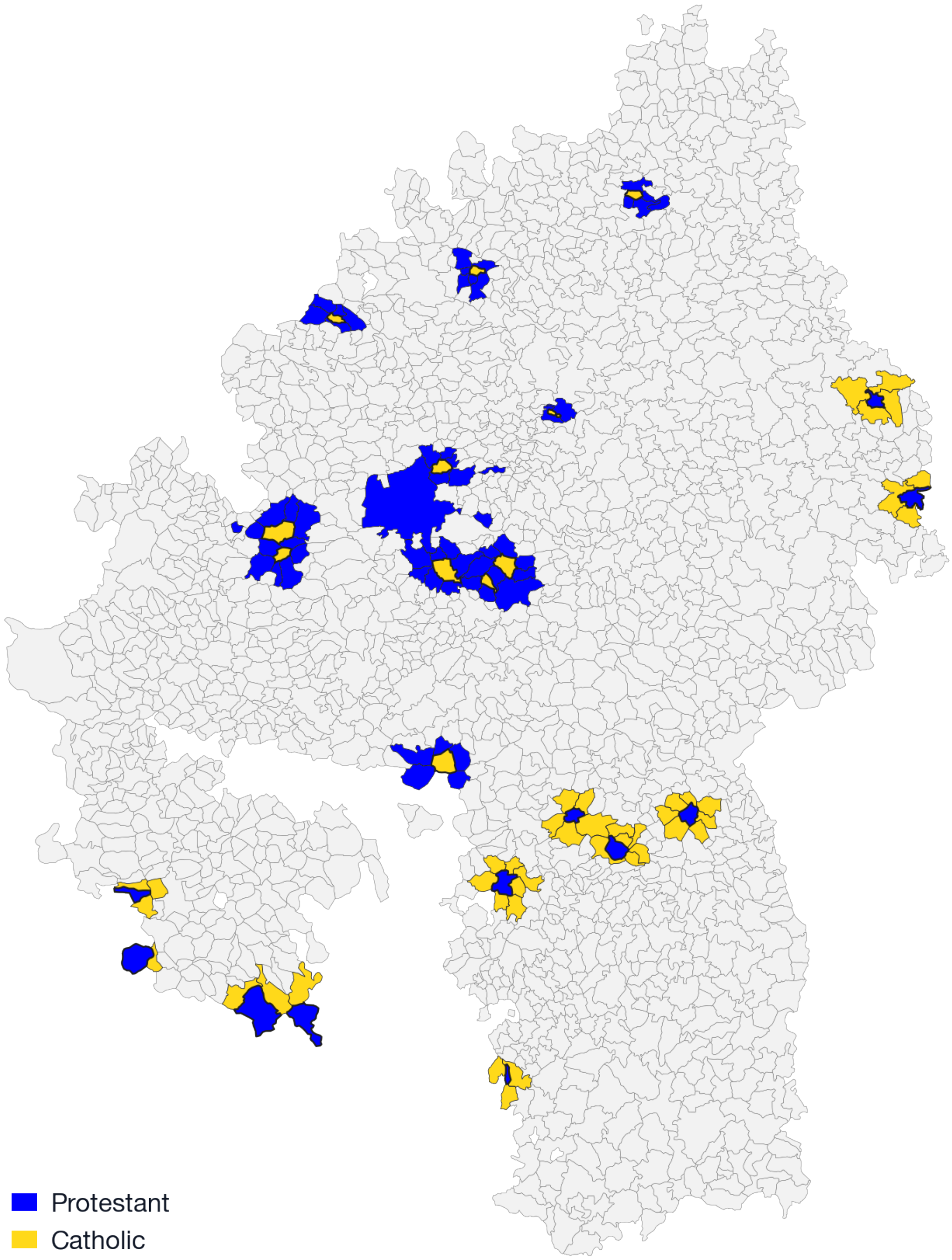
Source: Max Planck Institute for Demographic Research (MPIDR) and Chair for Geodesy and Geoinformatics, University of Rostock (CGG) (2011). Authors' design.

A.3 Additional Tables

Table A.1: Enclaves and surrounding parishes

Enclave	County	Surrounding parishes
		A. Protestant enclaves
Pflummern	Riedlingen	Mörsingen, Daugendorf, Grüningen, Altheim, Andelfingen, Friedingen, Upflamör
Rottenacker	Ehingen	Munderkingen, Volkersheim, Kirchbierlingen, Berg, Herbertshofen, Dettingen, Kirchen
Mundingen	Ehingen	Kirchen, Lauterach, Erbstetten, Granheim, Dächingen
Ersingen	Ehingen	Oberdischingen, Donaurieden, Dellmensingen, Stetten, Achstetten, Rißtissen, Öpfingen
Tuttlingen	Tuttlingen	Wurmlingen, Nendingen
Neuhausen ob Eck	Tuttlingen	Nendingen, Fridingen an der Donau
Schwenningen am Neckar	Rottweil	Mühlhausen
Schweindorf	Neresheim	Kösingen, Ohmenheim, Utzmemmingen
Walxheim	Ellwangen	Unterschneidheim, Zöbingen, Pfahlheim, Tannhausen
Flözlingen	Rottweil	Horgen, Zimmern ob Rottweil, Stetten ob Rottweil
Wilhelmsdorf	Ravensburg	Pfrungen, Esenhausen, Zußdorf
		B. Catholic enclaves
Weil der Stadt	Leonberg	Schafhausen, Ostelsheim, Simmozheim, Merklingen, Malmsheim, Renningen
Dätzingen	Böblingen	Ostelsheim, Schafhausen, Döffingen, Aidlingen, Deufringen, Gechingen
Neuhausen auf den Fildern	Esslingen	Denkendorf, Köngen, Nellingen auf den Fildern, Unterensingen, Wolfschlugen, Sielmingen, Bernhausen, Plieningen, Scharnhausen
Unterboihingen	Nürtingen	Köngen, Wendlingen am Neckar, Kirchheim unter Teck, Oberboihingen, Unterensingen
Wernau (Neckar)	Esslingen	Plochingen, Hochdorf, Notzingen, Kirchheim unter Teck, Wendlingen am Neckar, Köngen, Deizisau
Oeffingen	Waiblingen	Stuttgart, Stadt, Aldingen am Neckar, Neckarrems, Hegnach, Waiblingen, Schmiden
Stockheim	Brackenheim	Frauenzimmern, Haberschlacht, Kleingartach, Niederhofen, Brackenheim
Ebersberg	Backnang	Lippoldsweiler, Sechselberg
Wimmental	Weinsberg	Dimbach, Willsbach, Sülzbach, Ellhofen, Grantschen, Eberstadt
Amrichshausen	Künzelsau	Steinbach, Laßbach, Kocherstetten, Morsbach
Großengstingen	Reutlingen	Meidelstetten, Kleinengstingen, Honau, Undingen, Erpfingen

Figure A.3: Religious enclaves in Württemberg



Notes: The plot shows the location of all 22 religious enclaves and their surrounding parishes in Württemberg. Yellow marks enclave parishes; blue marks surrounding parishes of the other Christian denomination. We define religious enclaves as parishes that have been completely surrounded by parishes of the other Christian denomination since the 16th century (excluding "mixed parishes" with less than 90% adherence to the majority denomination).

Table A.2: The effect of religion on alternative fertility transition dates

	Within transformation					
	County		Neighborhood		Enclave	
	(1)	(2)	(3)	(4)	(5)	(6)
Protestant (0/1)	-4.860***	-5.211***	-4.036***	-3.973***	-11.683***	-11.207***
	(1.579)	(1.537)	(1.198)	(1.143)	(3.523)	(3.677)
Standardized coefficient	-0.389	-0.417	-0.323	-0.318	-0.909	-0.872
Observations	1601	1601	1601	1601	125	125

Notes: The dependent variable is the alternative fertility transition date, based on parish-level pre-transition fertility in 1871 and 1878–1881. Columns (1) and (2) include county fixed effects. Columns (3) and (4) transform all variables into deviations from their means among parishes within a 10 km neighborhood. Columns (5) and (6) compare religious enclaves with their neighboring parishes and include enclave-cluster fixed effects and an enclave indicator. All regressions include a quadratic in longitude and latitude. Columns (2), (4), and (6) additionally control for geographic characteristics and indicators of pre-Reformation development listed in Table 1.1. Standard errors are clustered by county in Columns (1)–(4) and by enclave cluster in Columns (5)–(6). Standardized coefficients are standardized by the dependent variable's sample standard deviation. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

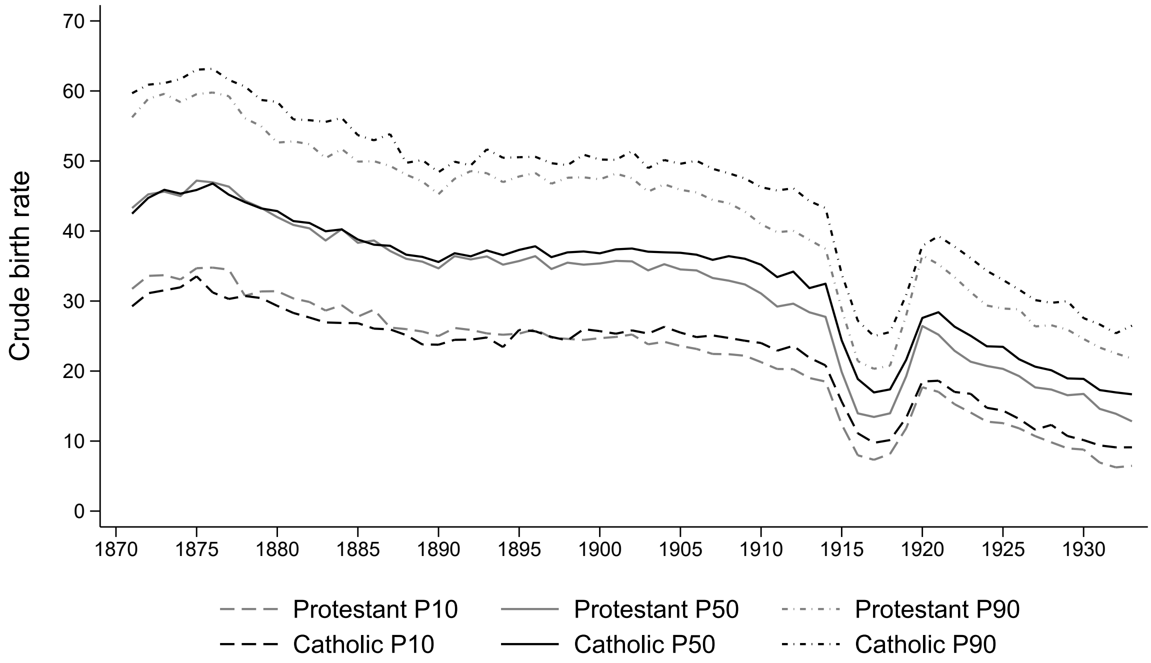
Table A.3: Robustness to denomination census year and large-town exclusion

	All parishes		Excl. towns $> 5{,}000$	
	1843	1871	1843	1871
A. Fertility transition date				
County FE + controls	-7.220***	-7.849***	-7.262***	-7.915***
	(1.194)	(1.232)	(1.196)	(1.224)
10 km neighborhood + controls	-6.245***	-6.875***	-6.350***	-6.956***
	(1.248)	(1.290)	(1.244)	(1.273)
Enclave comparison + controls	-9.762***	-11.397***	-10.318***	-12.046***
	(3.082)	(3.133)	(3.124)	(3.200)
B. Economic modernization, enclave comparison + controls				
Industrial employment, 1907	2.212	2.333	1.872	2.085
	(4.014)	(4.000)	(4.271)	(4.191)
Non-agricultural employment, 1907	1.566	1.638	1.251	1.512
	(4.314)	(4.212)	(4.599)	(4.422)
Steam engine, 1869	-0.030	0.035	-0.060	0.015
	(0.069)	(0.034)	(0.070)	(0.026)
Taxable income per capita, 1907	-1.221	19.190	-6.373	15.855
	(28.736)	(18.614)	(30.172)	(19.406)
Spatial trend	yes	yes	yes	yes
Controls	yes	yes	yes	yes
Observations (full/enclave)	1635/127	1606/123	1619/123	1595/120

Notes: The table reports Protestant-Catholic differences under alternative denomination definitions and sample restrictions. Protestant and Catholic parishes are defined using either the 1843 or 1871 religion census. Protestant parishes have a Protestant population share above 90 percent; Catholic parishes are defined analogously, and mixed parishes are excluded. Column (3) is the baseline specification in the paper. Columns (1)–(2) use all parishes. Columns (3)–(4) exclude towns with more than 5,000 inhabitants in 1871. Panel A reports controlled specifications for the fertility transition year; negative coefficients indicate an earlier transition in Protestant parishes. Panel B reports selected economic-modernization outcomes using the controlled enclave comparison. All regressions include a quadratic in longitude and latitude and the controls used in the corresponding main specifications. Standard errors are clustered by county in the county and neighborhood specifications and by enclave case in the enclave specifications. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

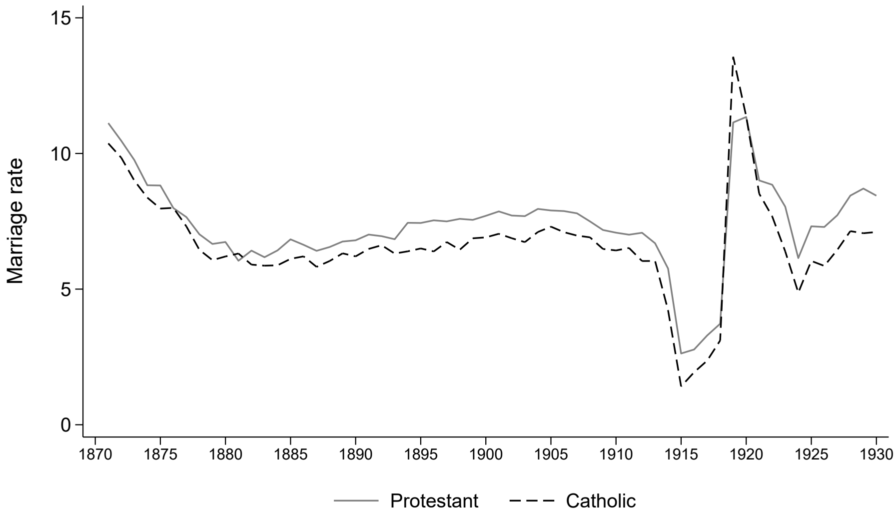
A.4 Additional Figures

Figure A.4: Crude birth rates at 10th, 50th, and 90th percentile



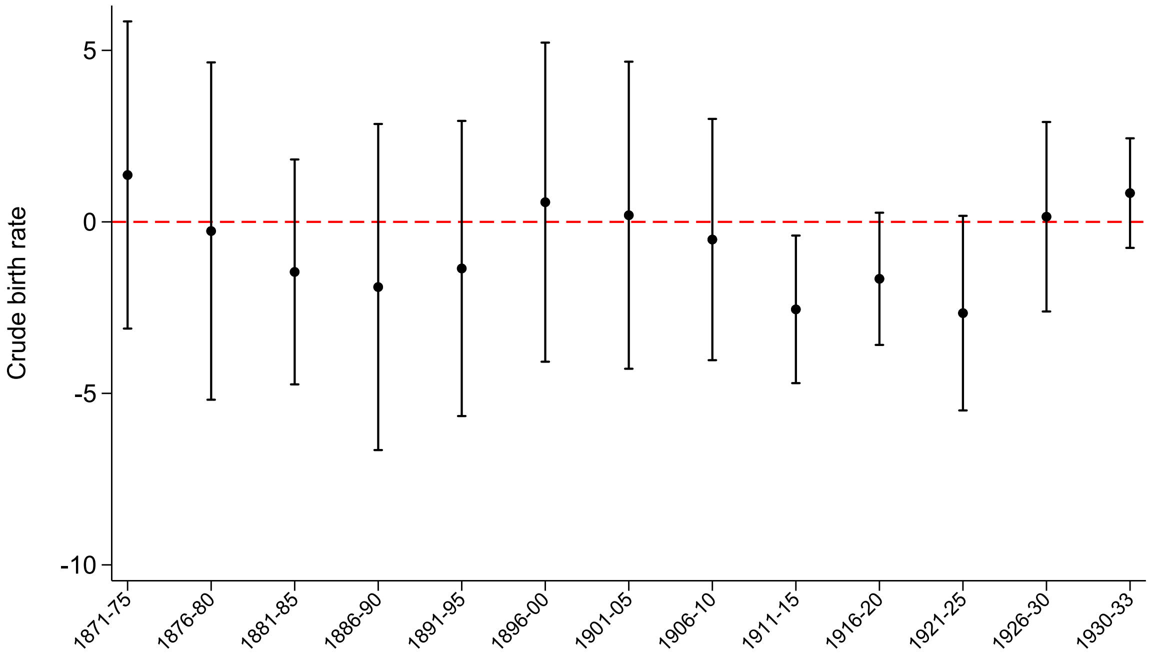
Notes: The plot shows crude birth rates at the 10th, 50th, and 90th percentile separately for Catholic and Protestant parishes.

Figure A.5: Marriage-rate gap between religions



Notes: The figure shows the time-series difference in marriage rates between Protestant and Catholic parishes.

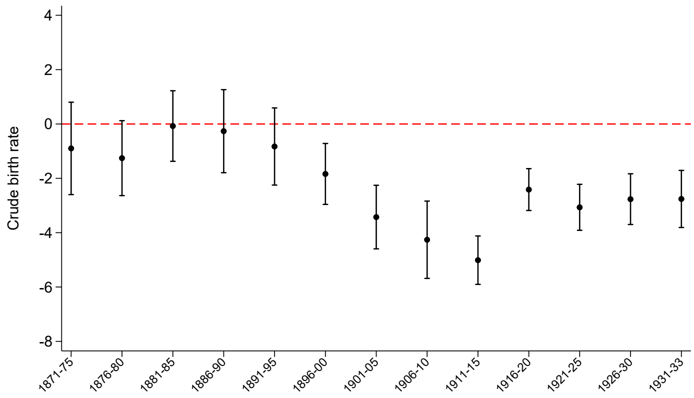
Figure A.6: Effect of being an enclave on fertility



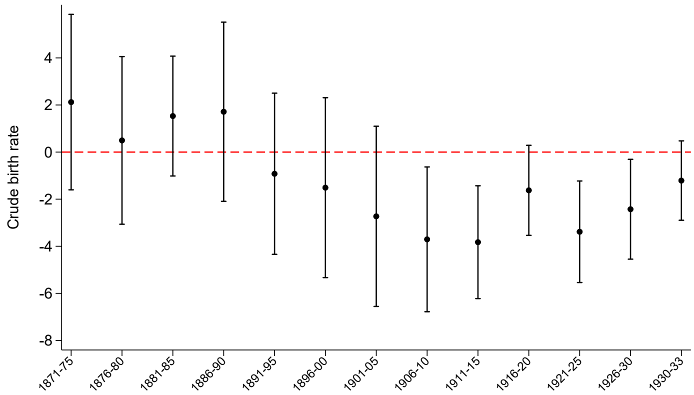
Notes: The figure reports period-specific OLS estimates of enclave effect. Regressions use the full 5-year period sample, include OA fixed effects, and control for elevation, river dummy, soil-type, and pre-Reformation city/market indicators. Coefficients are shown with 95% confidence intervals.

Figure A.7: Protestantism and crude birth rates with controls

(a) Full sample



(b) Enclave sample

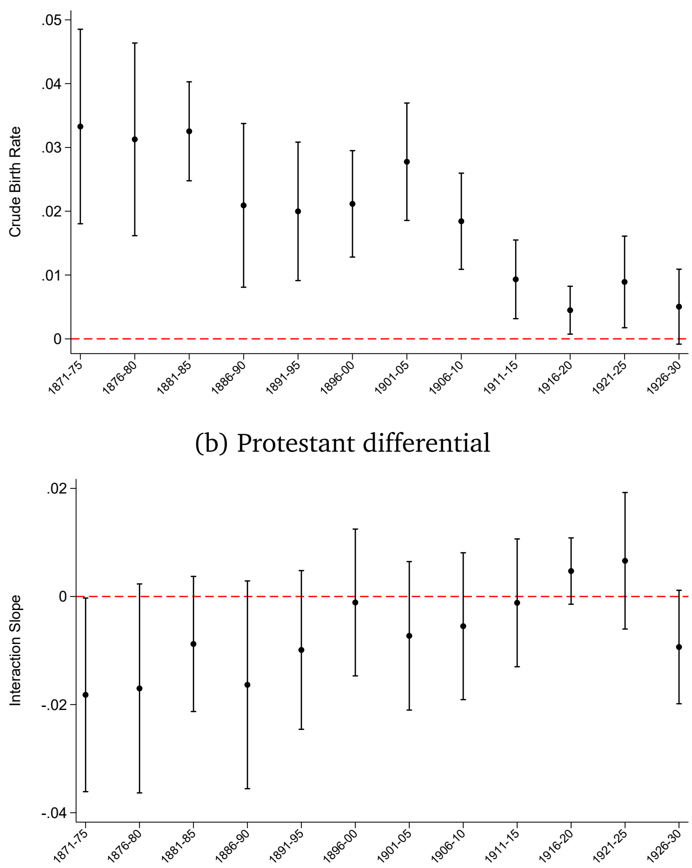


Notes: The figure shows the effect of Protestantism on crude birth rates for periods from 1871-75 to 1931-33, based on regression model 1.1. Panel (a) uses the full sample and includes county fixed effects. Panel (b) restricts the sample to religious enclaves and their surrounding parishes and includes enclave-cluster fixed effects and an enclave indicator. Both regressions include a quadratic in longitude and latitude, geographic characteristics, and indicators of pre-Reformation development, as listed in Table 1.1. Robust standard errors are clustered at the county level in panel (a) and at the enclave-cluster level in panel (b). Point estimates are indicated by a dot. Vertical bars indicate the 90 percent confidence interval around each estimate.

Figure A.8: Infant mortality and fertility

(a) Main effect

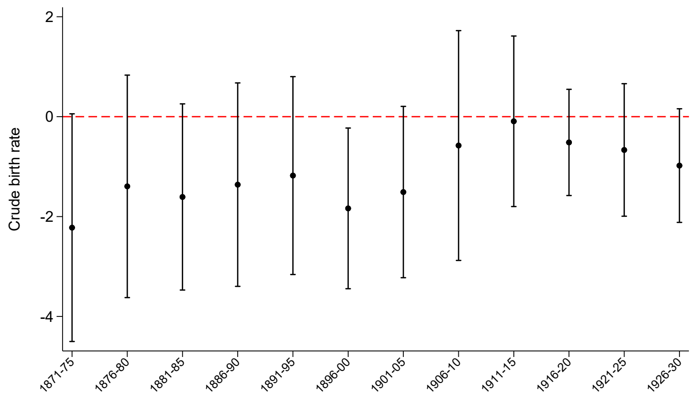
(b) Protestant differential



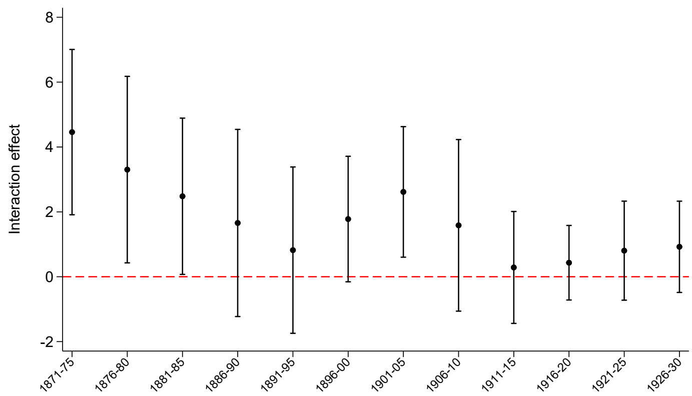
Notes: The figure shows the effect of infant mortality on crude birth rates in each five-year interval from 1871-75 to 1931-33. Panel (a) reports the main effect of infant mortality on crude birth rates. Panel (b) reports the Protestant differential in the infant-mortality fertility slope (interaction with the Protestant indicator). All regressions use the full sample, apply spatial transformation (radius 10 km), and include a quadratic in longitude and latitude. The regression controls include elevation, distance to river (1845), soil-type indicators, indicators for pre-Reformation city and pre-Reformation market, and Protestant status. Standard errors are clustered at the county level; vertical bars indicate 95 percent confidence intervals.

Figure A.9: Border parishes and fertility

(a) Border effect

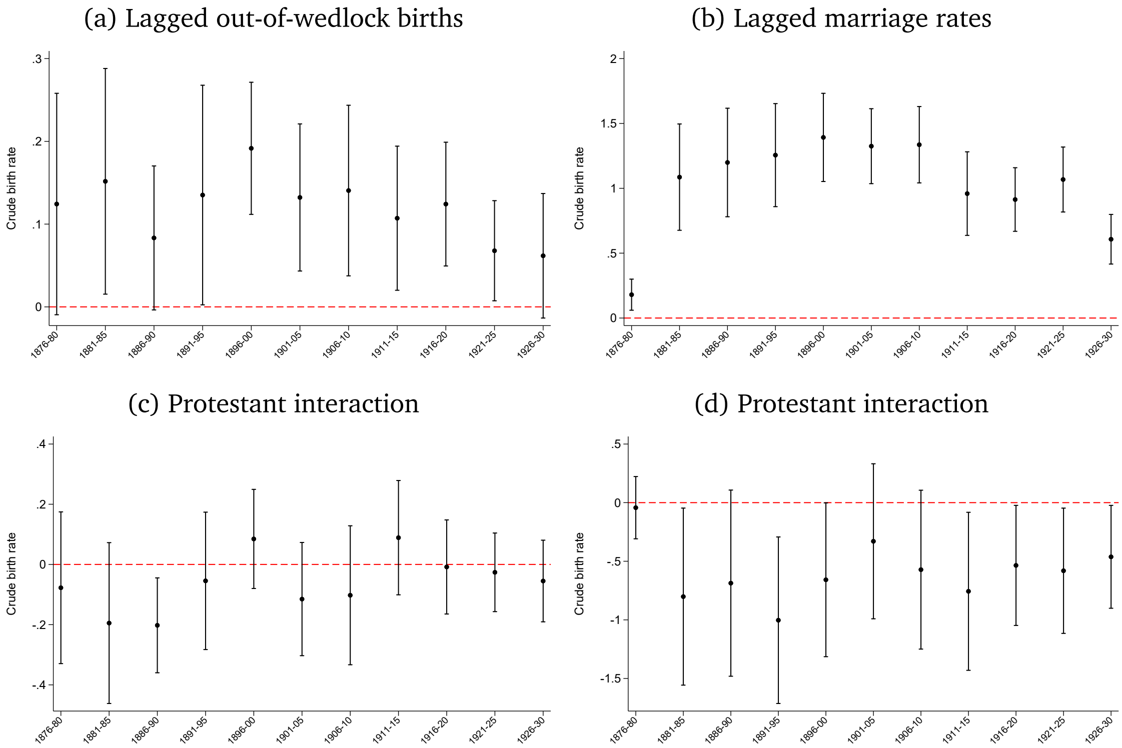


(b) Protestant interaction



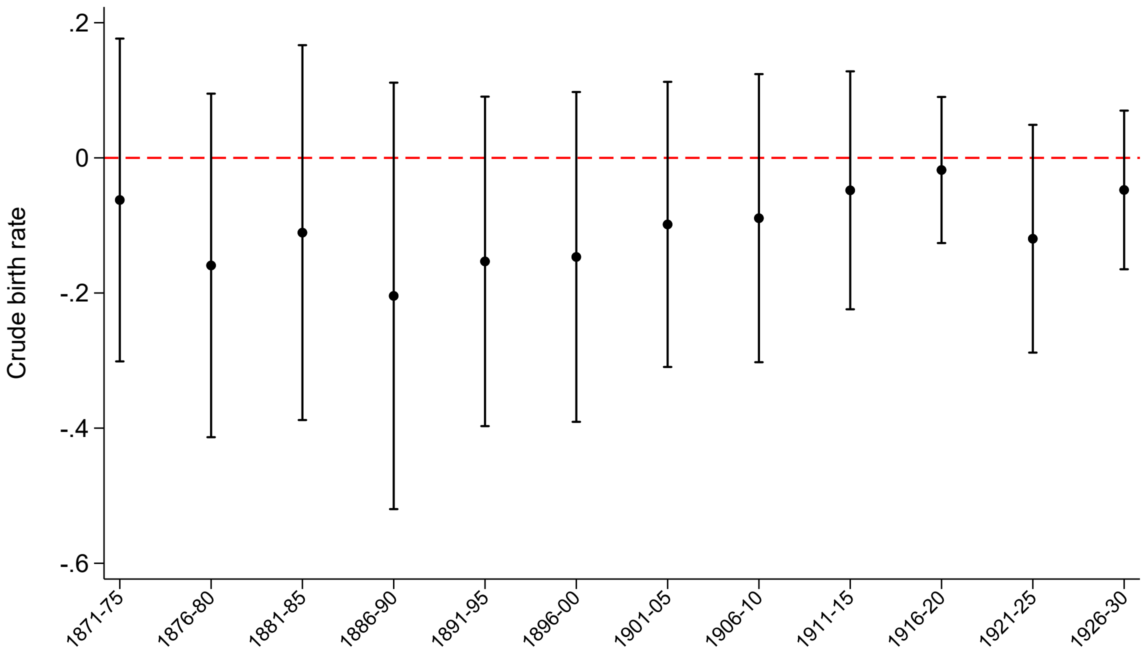
Notes: The figure reports the full-sample spatially transformed regressions for crude birth rates from 1871-75 to 1931-33 using the complete sample. Panel (a) shows the border effect for Catholics, and panel (b) shows the Protestant Border interaction. All regressions use all controls and spatial transformation; standard errors are clustered at the county level and 95% confidence intervals are shown.

Figure A.10: Marriage rates and out-of-wedlock births



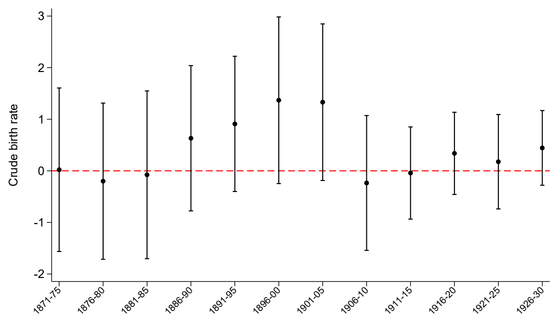
Notes: The figure reports lagged-outcome regressions by five-year interval (first year of each period only). Panel (a) shows the effect of lagged out-of-wedlock births on crude birth rates. Panel (c) shows the Protestant differential in the lagged out-of-wedlock effect. Panel (b) shows the effect of lagged marriage rates. Panel (d) shows the Protestant differential in the lagged marriage-rate effect. All regressions apply the 10-km spatial transformation and include full controls used in the main appendix gap decompositions. 95% confidence intervals are shown.

Figure A.11: Distance to the nearest monastery and fertility

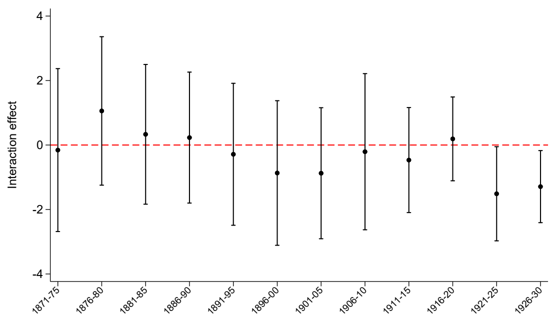


Notes: The figure shows the full-sample effect of distance to the nearest monastery on crude birth rates, estimated separately for each five-year period from 1871-75 to 1931-33. Regressions include county fixed effects and the full set of control variables. Standard errors are clustered at the county level and 95% confidence intervals are shown.

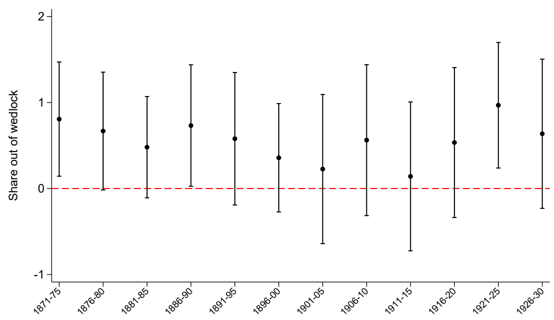
Figure A.12: Effect of Closed Inheritance Regime on Outcomes



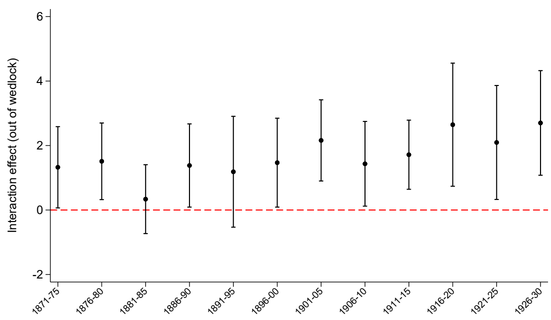
(a) Fertility: inheritance



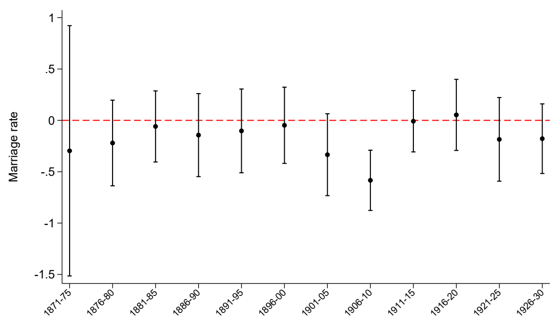
(b) Fertility: interaction



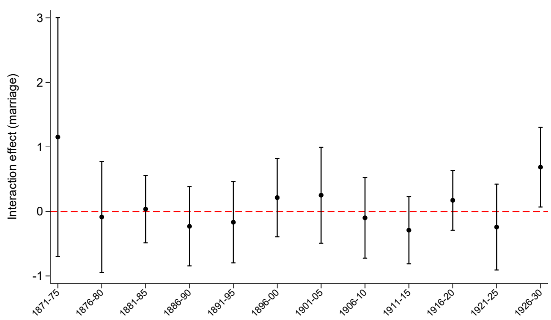
(c) Out-of-wedlock: inheritance



(d) Out-of-wedlock: interaction



(e) Marriage: inheritance



(f) Marriage: interaction

Notes: The figure reports period-specific estimates for the relationship between a closed-inheritance regime and demographic outcomes. Panels (a), (c), and (e) show the main effect of closed inheritance on crude birth rates, out-of-wedlock births, and marriage rates, respectively. Panels (b), (d), and (f) show the corresponding interaction with the Protestant parish indicator. Vertical bars indicate 95 percent confidence intervals.

Appendix B

From Weimar to West Germany: Confessional Politics, Political Fragmentation, and the Extreme Right

B.1 Data Sources

Table B-1A: Data sources (Part A)

Year	Variable	State	Source
A. Election Outcomes			
1928, 1932	Vote Shares, Reichstag Elections	Baden	Falter and Hänisch (1990b)
1932	Vote Shares, State Elections	Württemberg	Württembergischen Statistischen Landesamt (1932)
1949, 1965, 1969	Vote Shares, Bundestag Elections	B & W	Landesarchiv Baden-Württemberg (2026)
1968	Vote Shares, State Elections	B & W	Landesarchiv Baden-Württemberg (2026)
B. Population, Demography, and Religion			
1925	Total Population	Baden	Falter and Hänisch (1990b)
1925	Female Population	Baden	Badischen Statistischen Landesamt (1927)
1925	Total Population	Württemberg	Statistisches Landesamt Baden-Württemberg (2008)
1933	Total Population	Württemberg	Württembergischen Statistischen Landesamt (1935)
1925	Religious Composition	Baden	Falter and Hänisch (1990b)
1925	Religious Composition	Württemberg	Statistisches Landesamt Baden-Württemberg (2008)
1950, 1956, 1961, 1970	Population and Female Population	B & W	Landesarchiv Baden-Württemberg (2026)
1950	Religious Composition	B & W	Landesarchiv Baden-Württemberg (2026)

Table B-1B: Data sources (Part B)

Year	Variable	State	Source
C. Economic Structure			
1925	Population in Industry	Baden	Falter and Hänisch (1990b)
1933	Population in Industry	Württemberg	Württembergischen Statistischen Landesamt (1935)
1932	Taxation per capita	Württemberg	Württembergischen Statistischen Landesamt (1935)
1933	Unemployment	Württemberg	Württembergischen Statistischen Landesamt (1935)
1933	Sectoral Workforce Composition	Württemberg	Württembergischen Statistischen Landesamt (1935)
1950, 1961, 1970	Sectoral Workforce Composition	B & W	Landesarchiv Baden-Württemberg (2026)
D. Geography			
–	Elevation	B & W	Bundesamt für Kartographie und Geodäsie (2017)
–	Ruggedness	B & W	Bundesamt für Kartographie und Geodäsie (2017)
–	Soil Type (6 Classifications)	B & W	Bundesanstalt für Geowissenschaften und Rohstoffe (2013)
–	Latitude	B & W	Authors' own
–	Longitude	B & W	Authors' own
–	Area	B & W	Authors' own
–	Navigable River Access	B & W	Authors' own
E. Miscellaneous			
–	Reformation Status in 1619	B & W	Bundesamt für Kartographie und Geodäsie (2017)
–	City before Reformation	Württemberg	Kommission für geschichtliche Landeskunde in Baden-Württemberg (1988)
–	Enclave Specification	B & W	Kommission für geschichtliche Landeskunde in Baden-Württemberg (1988)
F. Individual-Level Survey Data			
1968	Individual-Level Political Attitudes	B & W	Klingemann (2012)

B.2 Enclave Cases

Table B-2A: Enclave cases: Enclave(s) and surrounding parishes (Cases 1–22)

Case ID	Enclave(s)	Surrounding Parishes
1	Pflummern	Altheim, Andelfingen, Daugendorf, Friedingen, Grüningen, Mörsingen, Upflamör
2	Ebersberg	Lippoldsweiler, Sechselberg
3	Dätzingen	Aidlingen, Deufringen, Döffingen, Gechingen, Ostelsheim, Schafhausen
4	Biberach, Kirchhausen, Massenbachhausen	Bonfeld, Frankenbach, Fürfeld, Großgartach, Heilbronn, Massenbach, Obereisesheim, Schwaigern
5	Stockheim	Brackenheim, Frauenzimmern, Haberschlacht, Kleingartach, Niederhofen
6	Neuhausen auf den Fildern	Bernhausen, Denkendorf, Nellingen auf den Fildern, Scharnhausen, Sielmingen, Stuttgart, Stadt, Unterensingen, Wolfschlugen
7	Wernau (Neckar)	Deizisau, Hochdorf, Kirchheim unter Teck, Köngen, Notzingen, Plochingen, Reichenbach an der Fils, Wendlingen am Neckar
8	Talheim	Flein, Heilbronn, Horkheim, Ilsfeld, Lauffen am Neckar, Schozach, Untergruppenbach
9	Weil der Stadt	Malmsheim, Merklingen, Ostelsheim, Renningen, Schafhausen, Simmozheim
10	Oeffingen	Aldingen am Neckar, Hegnach, Neckarrems, Schmiden, Stuttgart, Stadt, Waiblingen
11	Wimmental	Dimbach, Eberstadt, Ellhofen, Grantschen, Sülzbach, Willsbach
12	Lautlingen, Margrethausen	Burgfelden, Ebingen, Hossingen, Laufen an der Eyach, Meßstetten, Pfeffingen, Tailfingen
13	Unterboihingen	Kirchheim unter Teck, Köngen, Oberboihingen, Unterensingen, Wendlingen am Neckar
14	Großengstingen	Erpfingen, Honau, Kleinengstingen, Meidelstetten, Undingen

Case ID	Enclave(s)	Surrounding Parishes
15	Brigach, Buchenberg, Burgberg, Erdmannsweiler, Flözlingen, Gutach (Schwarzwaldbahn), Hornberg, Kirnbach, Königsfeld im Schwarzwald, Langenschiltach, Lehengericht, Mönchweiler, Peterzell, Reichenbach, Schiltach, St. Georgen, Stockburg, Tennenbronn, Weiler	Aichhalden, Fischbach, Gremmelsbach, Hardt, Hausach, Horgen, Kinzigtal, Lackendorf, Langenbach, Lauterbach, Mühlenbach, Neuhausen, Niedereschach, Niederwasser, Nußbach, Obereschach, Prechtal, Rohrbach im Schwarzwald, Schenkenzell, Schramberg, Stetten ob Rottweil, Unterkirnach, Villingen im Schwarzwald, Wolfach, Zimmern ob Rottweil
16	Schwenningen am Neckar	Bad Dürrheim, Dauchingen, Hochemmingen, Mühlhausen, Villingen im Schwarzwald, Weilersbach
17	Tuttlingen	Emmingen ab Egg, Hattingen, Liptingen, Möhringen, Nendingen, Wurmlingen
18	Hausen ob Verena, Rietheim	Balgheim, Dürbheim, Gunningen, Seitingen, Spaichingen, Weilheim, Wurmlingen
19	Neuhausen ob Eck	Buchheim, Fridingen an der Donau, Gallmannsweil, Heudorf im Hegau, Liptingen, Nendingen, Schwandorf, Worndorf
20	Biesingen, Oberbaldingen, Schura, Sunthausen, Talheim, Trossingen, Tuningen, Öfingen	Aasen, Aixheim, Aldingen, Deißlingen, Durchhausen, Eßlingen, Gunningen, Heidenhofen, Hochemmingen, Ippingen, Mühlhausen, Oberflacht, Pfohren, Unterbaldingen, Weigheim, Zimmern
21	Laubach, Neubronn	Abtsgmünd, Dewangen, Heuchlingen, Hohenstadt, Schechingen
22	Walxheim	Pfahlheim, Tannhausen, Unterschneidheim, Zöbingen

Table B-2B: Enclave cases: Enclave(s) and surrounding parishes (Cases 23–44)

Case ID	Enclave(s)	Surrounding Parishes
23	Hausen an der Rot	Fichtenberg (Vichberg), Fornsbach, Oberrot, Ottendorf (Oedendorf)
24	Amrichshausen	Kocherstetten, Künzelsau, Laßbach, Morsbach, Steinbach
25	Haagen, Honsbronn, Laudenbach	Elpersheim, Herrenzimmern, Neubronn, Pfitzingen, Queckbronn, Rüsselhausen, Vorbachzimmern, Weikersheim, Wermutshausen
26	Schweindorf	Kösingen, Ohmenheim, Utzmemmingen
27	Ahlen, Attenweiler, Biberach an der Riß	Aßmannshardt, Birkenhard, Laupertshausen, Mettenberg, Mittelbiberach, Moosbeuren, Reute, Rißegg, Rupertshofen, Seekirch, Stafflangen, Tiefenbach, Ummendorf, Uttenweiler, Warthausen
28	Ersingen	Achstetten, Dellmensingen, Donaurieden, Oberdischingen, Rißtissen, Stetten, Öpfingen
29	Ennahofen, Grötzingen, Weilersteußlingen	Allmendingen, Altsteußlingen, Dächingen, Ehingen (Donau), Frankenhofen, Hütten, Justingen, Schmiechen
30	Mundingen	Anhausen, Dächingen, Erbstetten, Granheim, Kirchen, Lauterach
31	Rottenacker	Berg, Dettingen, Herbertshofen, Kirchbierlingen, Kirchen, Munderkingen, Volkersheim
32	Bad Ditzenbach, Deggingen, Drackenstein, Gosbach, Hohenstadt, Mühlhausen im Täle, Reichenbach im Täle, Westerheim, Wiesensteig	Auendorf, Aufhausen, Donnstetten, Feldstetten, Gruibingen, Hausen an der Fils, Laichingen, Machtolsheim, Merklingen, Neidlingen, Nellingen, Schlat, Schopfloch, Unterböhringen, Zainingen
33	Oberholzheim	Achstetten, Bihlafingen, Bronnen, Burgrieden, Hüttisheim, Stetten

Case ID	Enclave(s)	Surrounding Parishes
34	Steinberg, Unterbalzheim, Wain	Dietenheim, Großschafhausen, Gutenzell, Kirchberg an der Iller, Orsenhausen, Schwendi, Sießen im Wald
35	Wilhelmsdorf	Esenhausen, Pfrungen, Zußdorf
36	Bissingen ob Lontal, Burgberg, Niederstotzingen, Oberstotzingen, Rammingen, Stetten ob Lontal	Asselfingen, Bergenweiler, Dettingen am Albuch, Hausen ob Lontal, Hermaringen, Hürben, Langenau, Setzingen, Sontheim an der Brenz, Öllingen
37	Gallenweiler	Eschbach, Grunern, Heitersheim, Tunsel, Wettelbrunn
38	Mengen, Opfingen, Schallstadt, Tiengen, Wolfenweiler	Biengen, Bollschweil, Ebringen, Freiburg, Hausen a. d. Möhlin, Kirchhofen, Merdingen, Munzingen, Niederrimsingen, Norsingen, Offnadingen, Pfaffenweiler, Scherzingen, Umkirch, Waltershofen
39	Bischoffingen, Königschaffhausen, Leiselheim	Amoltern, Burkheim, Endingen, Forchheim, Jechtingen, Kiechlinsbergen, Oberbergen, Oberrotweil, Sasbach, Wyhl
40	Weisweil	Forchheim, Kenzingen, Oberhausen, Wyhl
41	Hamberg, Hohenwart, Lehningen, Mühlhausen, Neuhausen, Schellbronn, Steinegg, Tiefenbronn	Bad Liebenzell, Bieselsberg, Büchenbronn, Friolzheim, Hausen an der Würm, Heimsheim, Huchenfeld, Monakam, Möttlingen, Münklingen, Pforzheim, Unterhaugstett, Unterreichenbach, Wimsheim, Würm
42	Bilfingen, Ersingen	Dietlingen, Eisingen, Ispringen, Königsbach, Pforzheim, Stein, Wilferdingen
43	Neibsheim	Bretten, Büchig bei Bretten, Diedelsheim, Gondelsheim, Gölshausen, Heidelsheim, Helmsheim, Oberacker
44	Waibstadt	Adersbach, Daisbach, Helmstadt, Neckarbischofsheim, Neidenstein, Rohrbach, Sinsheim

B.3 Additional Tables

Table B.3: The impact of religious composition on fragmentation in Baden, May 1928

	(1)	(2)	(3)	(4)
Catholics (%)	0.060***	0.065***	0.073***	0.070***
	(0.02)	(0.01)	(0.01)	(0.01)
Catholics squared (% * 10^{-2})	-0.081*	-0.142***	-0.158***	-0.158***
	(0.05)	(0.03)	(0.03)	(0.03)
Catholics cubed (% * 10^{-4})	0.007	0.064***	0.072***	0.075***
	(0.03)	(0.02)	(0.02)	(0.02)
R2	0.402	0.528	0.587	0.412
Spatial trend	no	yes	yes	yes
Controls	no	yes	yes	yes
Within transformation	none	none	county	radius

Notes: The table reports results from regressions of political fragmentation in Baden's May 1928 Reichstag election on the 1925 share of Catholics in a parish, its square, and its cube. All specifications control for the 1925 shares of Jews and other religions. Column (2) adds control variables for local geographic and economic characteristics. Column (3) adds county fixed effects. Column (4) reports coefficients from regressions using variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. Robust standard errors, clustered at the county level, are reported in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B.4: The impact of religious composition on fragmentation in Baden, November 1932

	(1)	(2)	(3)	(4)
Catholics (%)	0.056***	0.052***	0.055***	0.053***
	(0.02)	(0.01)	(0.01)	(0.01)
Catholics squared (% * 10^{-2})	-0.028	-0.066***	-0.070***	-0.068***
	(0.03)	(0.02)	(0.02)	(0.02)
Catholics cubed (% * 10^{-4})	-0.026	0.017	0.017	0.017
	(0.02)	(0.01)	(0.01)	(0.01)
R2	0.299	0.559	0.610	0.492
Spatial trend	no	yes	yes	yes
Controls	no	yes	yes	yes
Within transformation	none	none	county	radius

Notes: The table reports results from regressions of political fragmentation in Baden's November 1932 Reichstag election on the 1925 share of Catholics in a parish, its square, and its cube. All specifications control for the 1925 shares of Jews and other religions. Column (2) adds control variables for local geographic and economic characteristics. Column (3) adds county fixed effects. Column (4) reports coefficients from regressions using variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. Robust standard errors, clustered at the county level, are reported in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B.5: Religious composition and NSDAP voting across parishes in Württemberg, April 1932

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Catholics (%)	-0.265***	-0.291***	-0.290***	-0.287***	-0.295***	-0.307***	-0.285***
	(0.016)	(0.018)	(0.018)	(0.018)	(0.023)	(0.021)	(0.025)
Jews (%)	-0.006	-0.050	-0.020	-0.361	-0.531	-0.484*	-0.270
	(0.261)	(0.247)	(0.239)	(0.236)	(0.320)	(0.264)	(1.350)
Others (%)	-0.560**	-0.427*	-0.413*	-0.506**	-0.302	-0.363	-0.583
	(0.255)	(0.229)	(0.224)	(0.225)	(0.216)	(0.221)	(0.454)
Obs.	1744	1744	1744	1744	1744	1744	304
R2	0.437	0.485	0.493	0.504	0.589	0.245	0.619
Spatial trend	no	yes	yes	yes	yes	yes	yes
Geo. controls	no	no	yes	yes	yes	yes	yes
Sociodem. controls	no	no	no	yes	yes	yes	yes
Within transformation	none	none	none	none	county	radius	enclave

Notes: The table shows the relationship between religious composition—as measured by the population share of Catholics, Jews, and other religious groups, with the residual category being Protestants—and the Nazi vote share across parishes in Württemberg's state election on 24 April 1932. All columns present OLS estimates. Column (1) is for a parsimonious model without further controls. The remaining columns successively add a cubic polynomial in longitude and latitude to control for spatial trends as well as geographic and sociodemographic control variables, respectively. Column (5) adds county fixed effects. Column (6) reports coefficients from regressions using variables that have been transformed by subtracting the mean of all observations within a 10 km radius of each parish centroid. Column (7) restricts the sample to enclaves and surrounding parishes, adding fixed effects for each enclave case. Robust standard errors, clustered at the enclave-case level in Column (7) and at the county level in all other columns, are reported in parentheses. ***,**,* denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B.6: Protestant-Catholic differences in party support, individual-level evidence

	NPD support:								CDU support:	
	State election 1968		Any election		Next national election		Family members		State election 1968	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Catholic (0/1)	-0.020	-0.022	-0.018	-0.022*	-0.031*	-0.035*	-0.029**	-0.035**	0.344***	0.332***
	(0.018)	(0.017)	(0.013)	(0.013)	(0.018)	(0.018)	(0.015)	(0.015)	(0.039)	(0.040)
Obs.	570	570	852	852	641	641	583	583	570	570
Control mean	0.057	0.057	0.070	0.070	0.047	0.047	0.047	0.047	0.330	0.330
Controls	no	yes	no	yes	no	yes	no	yes	no	yes

Notes: The table shows the results of OLS regressions of indicators for NPD and CDU support on a dummy variable for Catholic respondents, with results for NPD support in Columns (1)–(8) and CDU support in Columns (9)–(10). The control group comprises Protestant respondents. The NPD support indicators capture whether respondents voted for the NPD in the 1968 state election (Columns (1)–(2)), in any previous election (Columns (3)–(4)), intend to vote NPD in the next federal election (Columns (5)–(6)), or have family members who generally support the NPD (Columns (7)–(8)). Non-voters are excluded in Columns (1), (2), (5), (6), (9), and (10). Conditional regressions control for age, gender, education, marital status, and community size. Robust standard errors in round brackets. ***, **, * denote significance at the 1%, 5%, and 10% levels.

Table B.7: Protestant-Catholic differences in evaluating policy topics as important

	Catholics:		Protestants:		Differences:	
	mean	st.d.	mean	st.d.	mean	s.e.
	(1)	(2)	(3)	(4)	(5)	(6)
Inflation	0.516	0.500	0.463	0.499	0.053	0.034
Confessional schools	0.267	0.443	0.187	0.391	0.079***	0.029
Agriculture	0.487	0.500	0.483	0.500	0.003	0.034
Grand coalition	0.487	0.500	0.481	0.500	0.006	0.034
Antisemitism	0.445	0.498	0.449	0.498	-0.004	0.034
Development aid	0.487	0.500	0.503	0.501	-0.017	0.034
Student riots	0.577	0.495	0.564	0.496	0.013	0.034
Defense spending	0.516	0.500	0.537	0.499	-0.021	0.034
Right to strike	0.279	0.449	0.300	0.459	-0.021	0.031
Major industries	0.191	0.393	0.156	0.363	0.035	0.026
Basic rights	0.357	0.480	0.293	0.456	0.064**	0.032
Eastern border	0.359	0.480	0.350	0.477	0.010	0.033
GDR relations	0.484	0.500	0.474	0.500	0.010	0.034
USA relations	0.306	0.461	0.248	0.433	0.057*	0.031
KPD ban	0.325	0.469	0.255	0.436	0.070**	0.031
Nazi politicians	0.200	0.401	0.181	0.385	0.020	0.027

Notes: The table shows sample means and standard deviations for the proportion of Catholics (Columns 1–2) and Protestants (Columns 3–4) who view a policy topic as important. Columns (5)–(6) show the mean difference and its standard error from a two-sided t-test. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table B.8: Protestant-Catholic differences in agreeing with policy statements

	Catholics:		Protestants:		Differences:	
	mean	st.d.	mean	st.d.	mean	s.e.
	(1)	(2)	(3)	(4)	(5)	(6)
Fight inflation through unempl.	0.544	0.499	0.588	0.493	-0.044	0.043
Support confessional schools	0.077	0.267	0.169	0.375	-0.092***	0.025
Support agriculture	0.697	0.461	0.701	0.459	-0.005	0.039
Grand coalition continues	0.655	0.476	0.757	0.430	-0.102***	0.039
Fight antisemitic statements	0.761	0.427	0.824	0.381	-0.064*	0.035
Reduce development aid	0.781	0.414	0.780	0.415	0.001	0.035
Fight student riots	0.735	0.442	0.787	0.410	-0.052	0.034
Reduce defense spending	0.778	0.416	0.782	0.413	-0.004	0.035
Limit right to strike	0.278	0.449	0.293	0.456	-0.015	0.039
Nationalize major industries	0.203	0.403	0.207	0.406	-0.004	0.040
Limit basic rights	0.209	0.407	0.184	0.389	0.025	0.034
Recognize Eastern border	0.385	0.488	0.404	0.492	-0.019	0.042
Establish relations to GDR	0.682	0.466	0.696	0.461	-0.014	0.039
Closer ties with USA	0.336	0.473	0.399	0.491	-0.063	0.042
Lift KPD ban	0.402	0.491	0.362	0.481	0.040	0.042
Allow ex-Nazi into politics	0.189	0.392	0.179	0.384	0.010	0.034

Notes: The table shows sample means and standard deviations for the proportion of Catholics (Columns (1)–(2)) and Protestants (Columns (3)–(4)) who agree with a particular policy goal. Columns (5)–(6) show the mean difference and its standard error from a two-sided t-test. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table B.9: Individual-level determinants of CDU and NPD voting

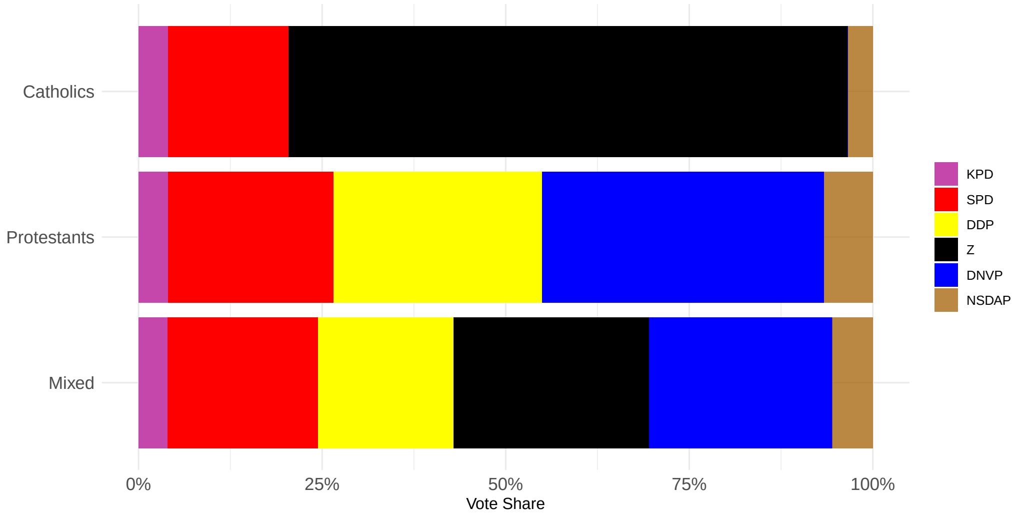
	CDU (1)	CDU (2)	CDU (3)	CDU (4)	NPD (5)	NPD (6)	NPD (7)	NPD (8)
Catholic (0/1)	0.332***	0.293***	0.292***	0.100	-0.022	-0.012	-0.012	-0.024
	(0.040)	(0.040)	(0.040)	(0.075)	(0.017)	(0.018)	(0.018)	(0.043)
Church very important (0/1)		0.349***	0.355***	0.255***		-0.084***	-0.084***	-0.092***
		(0.055)	(0.055)	(0.078)		(0.022)	(0.023)	(0.028)
Church important (0/1)		0.184***	0.189***	0.058		-0.057**	-0.057**	-0.067**
		(0.047)	(0.047)	(0.059)		(0.025)	(0.025)	(0.033)
Member in local club (0/1)			-0.069	-0.069			-0.002	0.003
			(0.047)	(0.061)			(0.022)	(0.032)
Catholic × church very important (0/1)				0.232**				0.019
				(0.101)				(0.046)
Catholic × church important (0/1)				0.318***				0.025
				(0.092)				(0.050)
Catholic × membership in local club (0/1)				0.001				-0.012
				(0.087)				(0.043)
R squared	0.188	0.251	0.254	0.272	0.048	0.069	0.070	0.070
Observations	570	564	564	564	570	564	564	564

Notes: The table shows the effects of religious affiliation, the importance of the church, and membership in a local club on whether respondents voted for the CDU (Columns (1) to (4)) or the NPD (Columns (5) to (8)) in the 1968 state election in Baden-Württemberg. All regressions control for respondents' age, gender, education, marital status, and the size of their place of residence. All models (1) to (8) also include a dummy variable for Catholic respondents. The baseline category is Protestant respondents; respondents who are neither Catholic nor Protestant are excluded. Columns (2) to (4) and (6) to (8) also include indicators for respondents who consider their church very or somewhat important. The baseline category is respondents who view their church as less important or are indifferent. Columns (3) to (4) and (7) to (8) add an indicator variable for whether respondents are members of a local club (sports, singing, or heritage). Robust standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

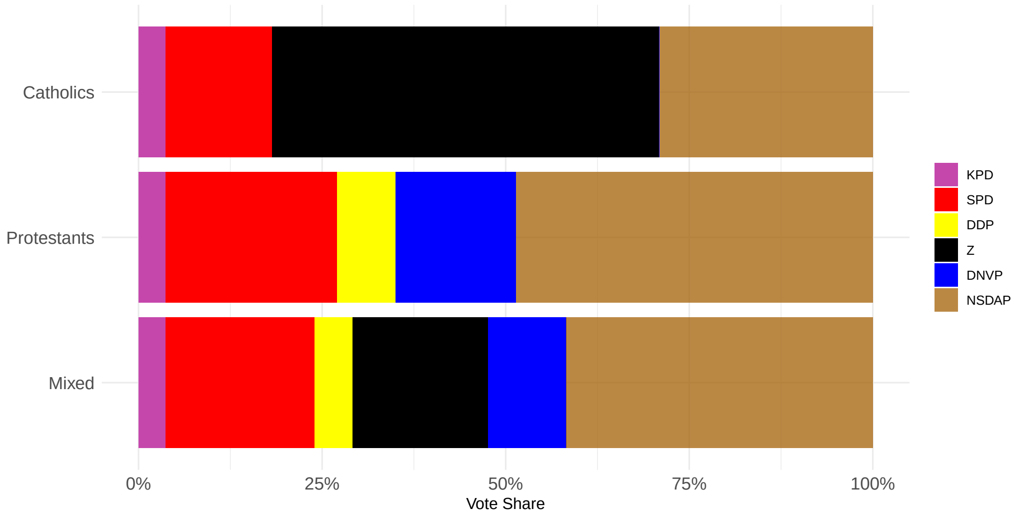
B.4 Additional Figures

Figure B.1: Confessional voting in the Weimar Republic: Vote shares

(a) 1928: Moderate electorate

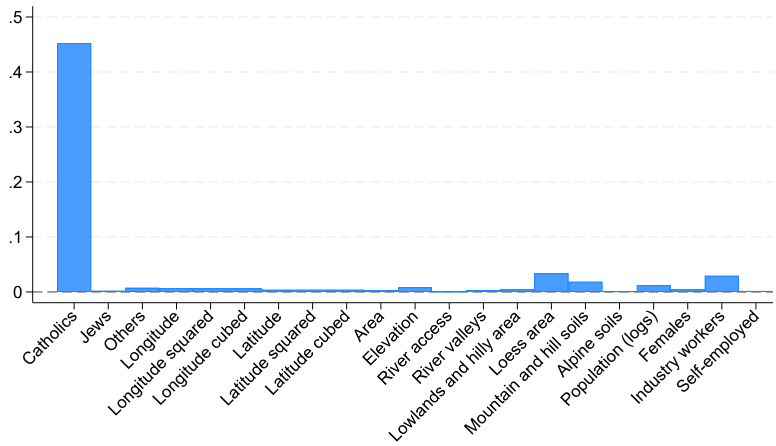


(b) 1932: Radicalized electorate



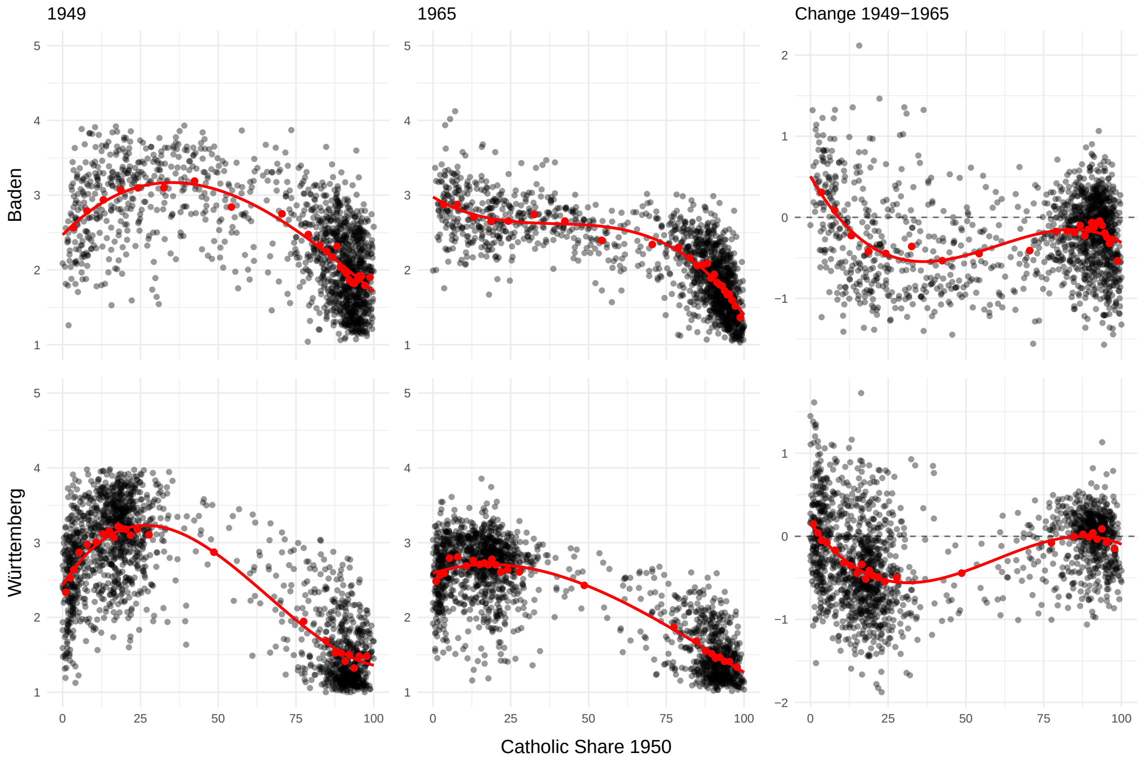
Notes: The figure shows the vote shares that result from the model predictions in Figure 2.1. We distinguish between an all-Catholic electorate, an all-Protestant electorate, and an electorate that is one-third Catholic and two-thirds Protestant. Panel (a) illustrates the predictions for a moderate electorate, as in the 1928 election, and panel (b) illustrates the predictions for a radicalized electorate, as in the 1932 election.

Figure B.2: Relative importance of predictors of NSDAP vote share in Baden, November 1932



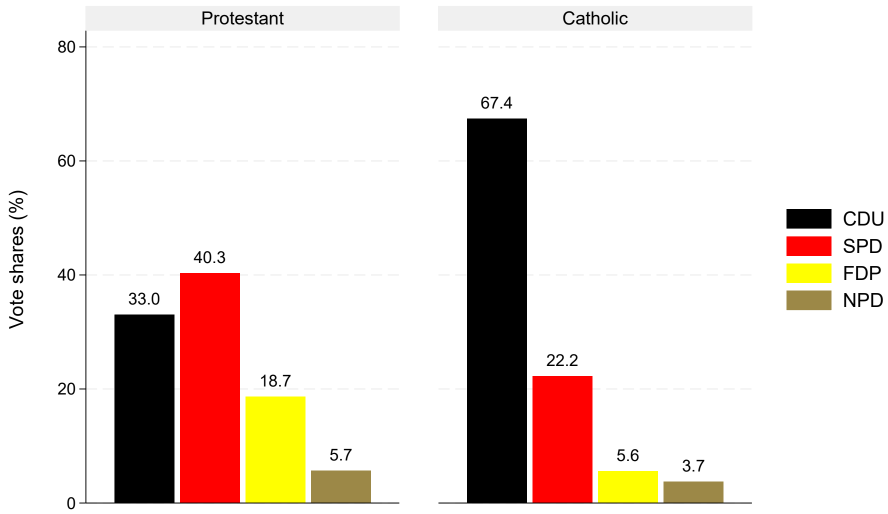
Notes: The figure illustrates the relative importance of each independent variable in predicting the Nazi vote share in the Reichstag election on 6 November 1932 in Baden, as determined by general dominance statistics calculated using Stata's domin command (Luchman, 2021). These statistics are Shapley values that decompose the overall R-squared of the underlying regression model into components attributable to each explanatory variable. Similar to Column (4) of Table 2.2, the underlying model regresses the local Nazi vote share against local religious composition, a cubic polynomial in longitude and latitude, and geographic and sociodemographic control variables.

Figure B.3: Share Catholic and political fragmentation in Baden and Württemberg, 8/1949 and 9/1965 elections



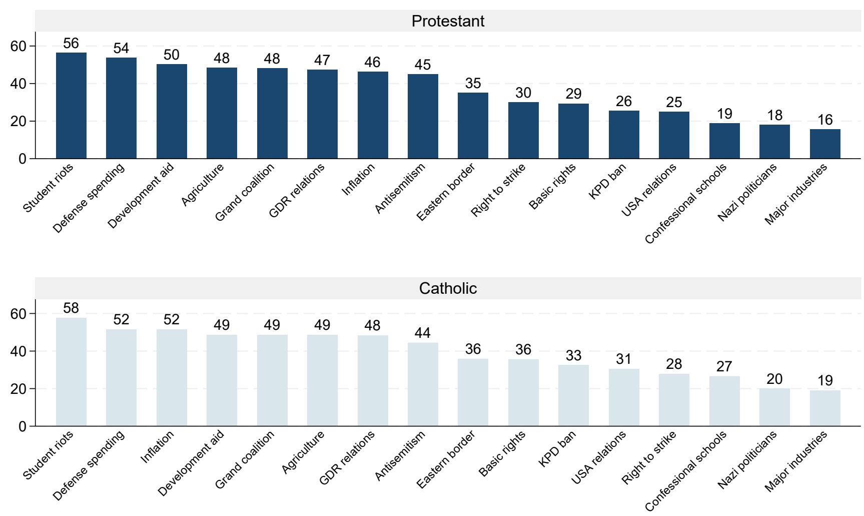
Notes: The figure presents scatter plots of the political fragmentation index against the 1950 Catholic population share, separately for Baden (upper row) and Württemberg (lower row). Column panels correspond to the Bundestag elections of 14 August 1949 and 19 September 1965, and the change in fragmentation between the two elections. The reference line at zero in the change panels distinguishes parishes where fragmentation increased from those where it decreased. A cubic polynomial regression fit is overlaid in red, together with red markers indicating mean fragmentation values within 25 equally sized quantile bins.

Figure B.4: Voting in the 1968 state election, by confession



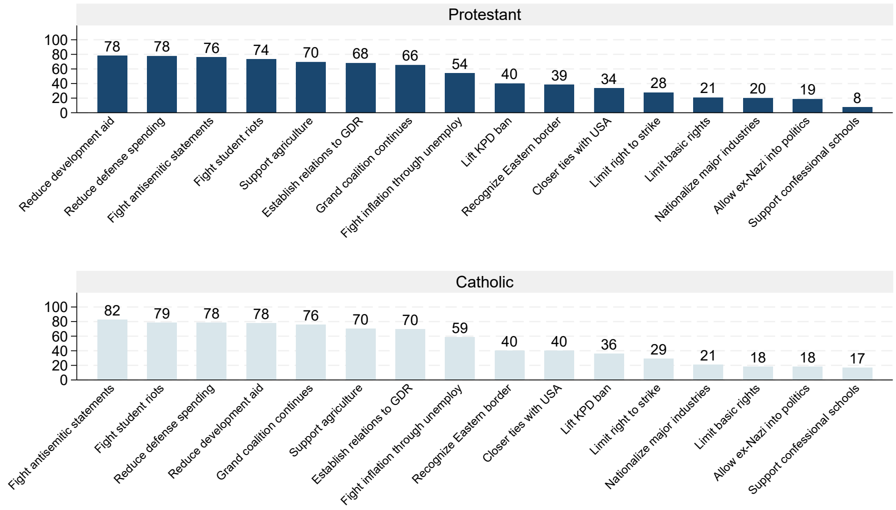
Notes: The figure shows the percentage of respondents who indicated that they voted for one of the four main parties in Baden-Württemberg's 1968 state election, broken down by denomination. The left panel shows the results for Protestant respondents and the right panel shows the results for Catholic respondents. Respondents who did not vote are excluded.

Figure B.5: Percentage of respondents evaluating a policy topic as important, by confession



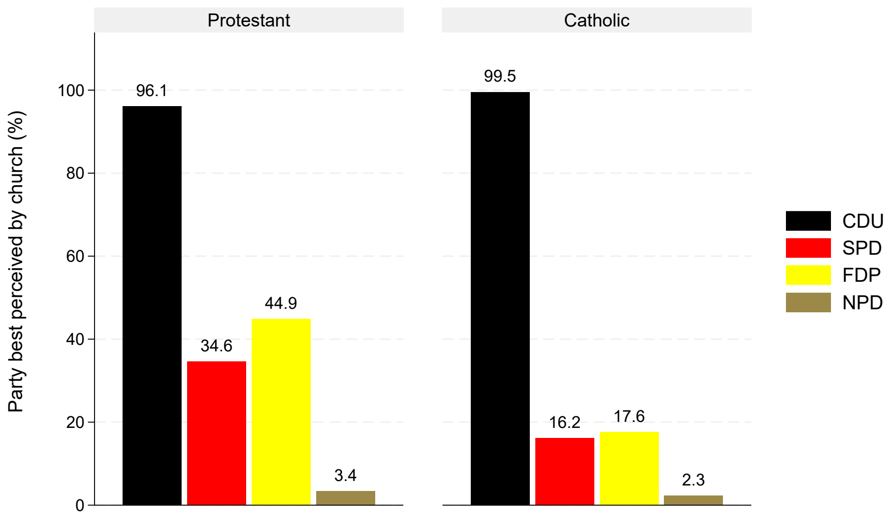
Notes: The figure shows the percentage of respondents who identified a specific policy issue as important. The top panel shows the results for Protestant respondents and the bottom panel shows the results for Catholic respondents. Topics are ordered by the proportion of respondents naming them as important.

Figure B.6: Percentage of respondents agreeing with a policy statement, by confession



Notes: The figure shows the percentage of respondents who agreed with a particular policy (goal). The top panel shows the results for Protestant respondents, while the bottom panel shows the results for Catholic respondents. Topics are ordered by the proportion of respondents who agreed with them.

Figure B.7: Party perceived as most favored by the Protestant and Catholic churches



Notes: The figure shows the percentage of respondents who rated the CDU, SPD, FDP, and NPD as the party most favored by their church. If two or more parties received the highest score, both were counted as most favored choices (i.e., multiple best choices are possible). The left panel shows the results for Protestant respondents and the right panel shows the results for Catholic respondents.

Appendix C

Reproducible LLM Workflows in Economic History: Building a Dataset from Degener's Wer Ist's?

C.1 Reproducibility Manifest

This appendix is to show whether the AI-generated data can be inspected, rerun, and checked against alternative runs or external benchmarks. This section is to display the key elements of the workflow. It should be the relevant standard for the stochastic reproducibility discussed in the main text: exact model responses may vary, but the inputs, instructions, validation rules, and comparison artifacts are documented.

First, I show that deterministic steps, such as file parsing, JSONL merging, and table generation, can be rerun from the archived code and inputs. Second, stochastic API calls can be rerun using the same prompts and settings, but exact equality is not guaranteed because proprietary models may change and because some calls do not use an explicit seed. Third, stochastic outputs are evaluated through versioned runs, automatic comparison, and adjudication logs.

Table C.1 shows the files needed for rerunning the workflow, while the following table records the artifacts needed for checking whether reruns produce materially similar results.

Table C.1: Core Reproducibility Manifest

Item	Location or setting
Repository entry point	`main.py`. All major workflow actions are exposed through this script.
Central configuration	`config.py`. This file stores default paths, model names, prompts, schemas, concurrency settings, and logging paths.
Command reference	`README.md` and `COMMANDS.md`. These files document the intended rerun commands.
Python environment	`requirements.txt`. The recorded dependency set includes `openai`, `pydantic`, `pandas`, `Pillow`, `pdf2image`, `opencv-contrib-python`, `openpyxl`, and `google-genai`.
Raw biography artifact	`reproduce/mistral/biographies.jsonl`. This raw segmented artifact contains 16,115 JSONL records before final paper-level filtering and cleaning.[a]
Image audit artifacts	`archive/random/` contains 200 page images, and `archive/processed/` contains 400 split-column images. These local artifacts document the image preprocessing logic for the sampled pages; they should not be read as the complete source-image archive for the full book.
Main extraction sample	`reproduce/mistral/random-biographies.jsonl`. This sample contains 1,000 records and is the default input for the personal, education, career, and family extraction tasks.
Social-standing sample	`reproduce/mistral/random-biographies-social-standing.jsonl`. This sample contains 200 records. Its metadata are stored in `reproduce/mistral/random-biographies-social-standing.meta.json`.
Sampling seed	1871 for the recorded random samples and adjudication pilot settings.
Main extraction model	`gpt-5-mini` for personal, education, career, and family extraction in the current configuration.
Social-standing models	`gpt-5-mini` as the default proprietary model; open-weight comparison runs are stored separately under `reproduce/social-standing/open-source/`.
Adjudication model	`gpt-5-mini`. The adjudication settings use a similarity threshold of 0.9, majority threshold of 9, and default seed 1871.
API credentials	Not archived. The workflow expects API keys through environment variables, especially `OPENAI_API_KEY`, `MISTRAL_API_KEY_PRODUCTION`, or `MISTRAL_API_KEY`.

[a] The main chapter reports the final cleaned dataset used for analysis. The raw JSONL artifact is retained because it is useful for audit and rerun checks, even where later cleaning removes duplicates, redirects, or other non-analysis records.

Because stochastic outputs are stored, model variation becomes observable and outputs can be compared across runs. Table C.2 records the versioned outputs that make run-to-run comparison possible.

Together, Tables C.1 and C.2 define what can be rerun, what can be compared, and which metadata connect model calls to final outputs. The prompts for personal,

Table C.2: Versioned outputs and audit artifacts

Artifact family	Location and use
OpenAI extraction versions	`reproduce/openai/version_*/`. Each versioned folder contains the structured outputs `personal.jsonl`, `education.jsonl`, `career.jsonl`, and `family.jsonl`. The checked version folders contain 1,000 records per extraction file.
Merged OpenAI datasets	`reproduce/openai/version_*/dataset_*.csv` when the dataset merge command is run with a version argument.
Social-standing versions	`reproduce/social-standing/version_*/`. Each folder contains `social_standing.jsonl` and, where available, `social_standing.meta.json`. The metadata record the input path, model, temperature, prompt label, prompt hash, service tier, creation time, and seed when specified.
Open-weight social-standing runs	`reproduce/social-standing/open-source/version_*/`. These runs document local or open-weight model comparisons separately from proprietary API runs.
Adjudication candidates	`reproduce/adjust/candidates.jsonl`. This file contains disagreement cases selected for adjudication across extraction versions.
Adjudicated output	`reproduce/adjust/adjust_adjudicated.jsonl`. This file records the final resolution for disagreement cases.
Adjudication report	`reproduce/adjust/adjust_report.txt`. The report summarizes 2,153 candidate cases, of which 2,130 were resolved: 432 automatically and 1,698 through LLM adjudication; 23 cases returned errors.
Human-readable tables	`paper/tables/`. These are the table fragments used in presentations and paper drafts for extraction stability, score robustness, human-AI comparison, open-weight comparison, and HISCO experiments.
Occupation experiments	`occupation/outputs/version_*/` and `occupation/stats/version_*/`. The recorded HISCO runs include five classification runs, one supervisor output, 200 pairwise prestige comparisons per run, and 20 ranking samples per run.

education, career, and family extraction are stored in `config.py`. Model settings are also recorded in configuration files and run metadata. In the latest configuration, the main extraction tasks use `gpt-5-mini` with a maximum concurrency of 200. Adjudication uses `gpt-5-mini`, a maximum concurrency of 100, and the thresholds reported in Table C.1.

The workflow saves JSONL logs for each step to `logs/`. The logs should be kept with the output data because they make failed calls, skipped existing outputs, runtime differences, and token costs visible. They also make clear which parts of the pipeline were rerun rather than inferred from final files alone. The principal log files are:

- `logs/openai_personal.jsonl`
- `logs/openai_education.jsonl`
- `logs/openai_career.jsonl`
- `logs/openai_family.jsonl`
- `logs/openai_social_standing.jsonl`
- `logs/openai_social_standing_oss.jsonl`
- `logs/openai_adjust.jsonl`

This section addresses the main reproducibility limitation of proprietary LLM APIs: if a model alias is updated, retired, or served with different backend behavior, exact replication of an individual response may no longer be possible. The best way to address this limitation by separating tasks, keeping prompts and schemas explicit, logging calls, saving intermediate outputs, rerunning selected tasks across versions, and making disagreement cases transparent.

C.2 Versioned Extraction Stability

This section documents how repeated LLM extractions behave when the same biographies are processed more than once, with the same parameters. The purpose is to make the stochastic part of the workflow transparent. In the experiment below, the results are presented in two steps: the first table reports aggregate field-level agreement, and the second table shows one case where a stable controlled field coexists with unstable free-text extraction.

Table C.3 compares two independent OpenAI extraction runs over the same 1,000 biographies. The similarity column uses the same normalized strings but counts fields whose string similarity is at least 75 percent. Cases in which both versions returned `unknown` are reported separately and are excluded from the exact and similarity denominators, so the rates are not inflated by shared missingness.

Table C.3: Similarity statistics for versioned biography extractions

Scope	Exact	>=75%	Unknown
Overall	75.67%	89.38%	42.45%
personal.gender	99.38%	99.48%	3.50%
personal.birth_date	99.52%	100.00%	16.00%
personal.birth_place	97.41%	99.63%	19.00%
personal.address	74.39%	98.72%	6.30%
education.education	57.27%	82.99%	23.00%
career.job	65.87%	76.11%	3.30%
career.career	6.51%	52.08%	6.30%
family.father	86.34%	99.81%	47.30%
family.mother	92.91%	99.76%	57.70%
family.ancestors	86.81%	89.01%	90.90%
family.ancestors_paternal	64.71%	64.71%	98.30%
family.ancestors_maternal	68.75%	81.25%	98.40%
family.spouse	95.96%	99.55%	55.50%
family.children	96.47%	99.36%	68.80%

Notes: The table compares `reproduce/openai/version_2912/` and `reproduce/openai/version_3012/`, two 1,000-biography runs produced on 29 December 2025 and 30 December 2025. Values are normalized by lowercasing and whitespace normalization before comparison. The `>=75%` column reports the share of paired fields whose similarity ratio is at least 0.75 after excluding cases where both runs returned `unknown`. The `Unknown` column reports the share of records where both runs returned `unknown`; it should be read alongside the agreement columns, especially for sparse family fields.

The table is consistent with two narrower claims used in the main text. First, fields with limited admissible values are comparatively stable in this audit: gender, birth date, and birth place have very high agreement after shared `unknown` values are excluded. The family fields require more caution. They often have high agreement conditional on both runs extracting a value, but several of them also have high shared-`unknown` rates. Their agreement therefore reflects both extraction stability and the sparsity of family information in the source. Second, open-ended fields are much less stable: `career.career` has 6.51 percent exact agreement and 52.08 percent similarity agreement. This should be interpreted as sensitivity in span selection and paraphrasing, rather than automatically as extraction failure. In many cases, the versions capture overlapping information but disagree about whether to expand abbreviations, preserve the original shorthand, include offices from the header, or reorder phrases. The aggregate pattern is easier to interpret in a concrete biography where the short job label is stable but the longer career span varies across versions. Table C.4 shows one such case. Across ten extraction versions, the short `career.job` field is stable: all versions identify the person as *Oberbürgermeister*. The longer `career.career` field, however, varies across runs. Some versions start from the office, others from the supervisory-board membership; some expand *Schles.* to *Schlesischen* or *Schlesien*; and some include the place as part of the career span while others treat it as residence-like information.

This example illustrates why versioning is useful. If only one extraction were kept, the researcher would see a plausible answer but would not know whether the answer was robust. With multiple versions, the unstable portion of the field becomes measurable and can be routed to an adjudication step. In this case, the adjudicator resolves the disagreement by extracting the literal span *Ob.-Bürgermstr. s. 83. M. d. Aufs. R. d. Schles. Kleinbahn A.-G.* from the biography text. This preserves the source wording instead of selecting one paraphrased version as if it were uniquely correct.

The field-level statistics show that the extraction pipeline is not uniformly unstable: most short fields converge to the same values across repeated runs. The example shows why the remaining instability is concentrated in longer text fields. These fields have a larger interpretation space, especially in abbreviated biographical prose where

Table C.4: Example of career-field variation across extraction versions

Version	Extracted `career.career` value
0101	Mitglied des Aufsichtsrats der Schlesischen Kleinbahn-Aktiengesellschaft, Beuthen (Schlesien)
0201	Mitglied des Aufsichtsrats der Schlesischen Kleinbahn-Aktiengesellschaft, Beuthen (Schles.)
2312	Ob.-Bürgermstr.; Mitglied des Aufsichtsrates der Schles. Kleinbahn A.-G. — Beuthen (Schles.)
2412	Oberbürgermeister in Beuthen (Schles.), Mitglied des Aufsichtsrats der Schlesischen Kleinbahn A.-G.
2512	Oberbürgermeister; Mitglied des Aufsichtsrats der Schlesischen Kleinbahn A.-G., Beuthen (Schles.)
2612	Mitglied des Aufsichtsrats der Schlesischen Kleinbahn-Aktien-Gesellschaft; Beuthen (Schles.)
2712	Oberbürgermeister; Mitglied des Aufsichtsrats der Schlesischen Kleinbahn-Aktiengesellschaft in Beuthen (Schlesien)
2812	Mitglied des Aufsichtsrats der Schlesischen Kleinbahn-Aktiengesellschaft; Beuthen (Schlesien)
2912	Mitglied des Aufsichtsrates der Schlesischen Kleinbahn-Aktiengesellschaft; Oberbürgermeister (s. 83.); Beuthen (Schles.)
3012	Oberbürgermeister (siehe S. 83); Mitglied des Aufsichtsrates der Schlesischen Kleinbahn A.-G.; Beuthen (Schles.)

Notes: The biography ID is `Image00344_right:7`. The source text is: *BRÜNING, Gg., Dr., Ob.-Bürgermstr. s. 83. M. d. Aufs. R. d. Schles. Kleinbahn A.-G. — Beuthen (Schles.)*. In the same ten versions, the `career.job` field is stable as *Oberbürgermeister*.

the boundary between job title, career statement, membership, honor, and address is not always explicit.

C.3 Score Robustness Across LLM Settings

This section tests whether generated occupation and social-standing scores are stable enough to use as variables. The exercise uses a fixed, 200-biography sample to generate two scores. The two generated fields are `occupation_score`, which asks for the prestige of the occupation described in the biography, and `social_standing`, which asks for the person's social status using titles, offices, honors, institutional positions, and other status markers in the biography. Both fields are integer scores on a 0–100 scale.

The prompt and model tables compare base, anchored, and empty prompts. The base prompt gives the task definition without scale anchors. The anchored prompt adds verbal anchors for the 0–100 scale. The empty prompt gives almost no task-specific guidance beyond the biography and requested output format. Most rows use 10 repeated versions on the fixed sample; where the table uses a different number of versions, the table note states this explicitly. The temperature and seed tables compare otherwise similar configurations with low and high temperature settings and, where supported, an explicit seed. A reported `None` seed means that no explicit seed was supplied; it does not imply deterministic behavior.

The tables report three stability measures. *Unique* is the average number of distinct scores assigned to the same biography across versions. Δ is the average absolute difference between all version pairs for the same biography, averaged over biographies. *Change rate* is the share of biographies whose score changes at least once. Lower values indicate greater within-configuration stability. The mean, standard deviation, median, and interquartile range describe the level and dispersion of the resulting score distribution.

The first pair of tables varies prompt and model choice for the two generated score fields. Tables C.5 and C.6 compare models and prompt variants. The base prompt asks the model to assign scores using only the biography text. The anchored prompt adds verbal anchors to the 0–100 scale. The empty prompt gives the model almost no task-specific guidance beyond the biography and requested output format.

The anchored prompt usually reduces dispersion across repeated versions. In the occupation-score exercise, for example, the `gpt-5-mini` base prompt has 3.69 distinct scores per biography on average and an average pairwise difference of 3.11 points, while the anchored prompt reduces these values to 2.54 and 2.79. The same pattern appears in the social-standing exercise: for `gpt-5-mini`, the base prompt has 4.93 distinct scores and an average pairwise difference of 4.31 points, while the anchored prompt reduces these values to 3.29 and 3.70. This does not mean that anchored prompts are always more correct. It means that they make the task definition sharper,

Table C.5: Occupation-score robustness by prompt and model

Prompt	Model	Mean	SD	Median	IQR	Unique	Δ	Change rate
Base	gpt-5	82.46	9.61	83.80	75.3–89.8	5.37	2.70	100.00%
Anchored	gpt-5	79.13	9.24	79.40	74.9–84.4	3.88	2.36	94.00%
Base	gpt-5-mini	81.22	10.02	83.70	75.4–88.9	3.69	3.11	100.00%
Anchored	gpt-5-mini	75.70	9.51	77.50	72.5–80.3	2.54	2.79	92.00%
Empty	gpt-5-mini	83.37	8.41	86.10	79.3–89.4	4.24	4.37	100.00%
Base	gpt-5-nano	86.53	8.46	90.20	81.0–92.6	5.46	3.65	100.00%
Anchored	gpt-5-nano	75.21	11.35	76.50	74.0–81.0	3.22	5.53	87.00%
Empty	gpt-5-nano	50.72	15.61	50.40	38.5–62.8	6.79	29.01	100.00%

Notes: Constant sample of $n = 200$ biographies; each row uses 10 repeated versions. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography. Δ is the mean absolute pairwise difference across versions within biography. Change rate is the share of biographies with any score change. Lower Unique, Δ, and change-rate values indicate greater within-configuration stability.
Source: Author's calculations from versioned score-generation runs under `reproduce/social-standing/version_*/`.

Table C.6: Social-standing robustness by prompt and model

Prompt	Model	Temp	Seed	Mean	SD	Median	IQR	Unique	Δ	Change rate
Base	gpt-5	None	None	83.94	10.35	85.10	77.7–91.5	5.61	2.70	100.00%
Anchored	gpt-5	None	None	83.54	10.17	84.90	78.5–90.8	4.62	2.29	95.50%
Base	gpt-5-mini	None	None	80.18	11.84	83.40	73.1–88.9	4.93	4.31	100.00%
Anchored	gpt-5-mini	None	None	76.68	11.73	78.50	69.5–84.5	3.29	3.70	97.00%
Empty	gpt-5-mini	None	None	79.19	9.06	80.90	73.4–85.4	4.88	4.87	100.00%
Base	gpt-5-nano	None	None	83.78	11.63	87.80	78.8–91.5	6.15	5.05	100.00%
Anchored	gpt-5-nano	None	None	76.48	14.28	78.00	71.5–85.0	4.25	6.99	98.50%
Empty	gpt-5-nano	None	None	50.65	15.02	50.10	40.3–62.1	6.83	27.67	100.00%
Base	o4-mini	None	None	73.62	14.18	76.00	64.0–84.8	3.94	5.45	100.00%
Anchored	o4-mini	None	None	72.77	13.20	74.50	66.0–81.5	3.13	4.15	98.00%
Empty	o4-mini	None	None	73.12	11.84	76.00	66.0–81.0	4.63	7.83	100.00%
Base	gpt-oss:20b	None	None	78.86	12.55	82.40	73.9–87.0	6.80	6.81	100.00%
Anchored	gpt-oss:20b	None	None	76.30	13.38	79.50	70.5–85.6	4.95	7.14	100.00%

Notes: Rows use $n = 200$ biographies; repeated-version counts vary by configuration from 10 to 20. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography. Δ is the mean absolute pairwise difference across versions within biography. Change rate is the share of biographies with any score change. Lower values on these three stability columns indicate greater within-configuration stability.
Source: Author's calculations from versioned score-generation runs under `reproduce/social-standing/version_*/`.

and that sharper task definitions reduce stochastic variation.

The empty-prompt rows are useful because they show what happens when the model is asked to infer the scoring task with minimal guidance. The most visible case is the `gpt-5-nano` empty-prompt configuration. For the occupation score, the mean falls to 50.72 and the average pairwise difference rises to 29.01 points. For social standing, the corresponding mean is 50.65 and the average pairwise difference is 27.67 points. This is the kind of behavior that the workflow should expose before such variables are used in downstream analysis.

The second pair of tables separates task definition from sampling variation by comparing temperature and seed settings. Tables C.7 and C.8 compare runs that vary temperature and, where supported, seed settings. These tables are useful for separating two concepts that are often conflated. A low temperature can make outputs more stable, but it does not by itself make the task well defined. Conversely, a high temperature can increase variation even when the prompt is otherwise reasonable.

Table C.7: Occupation-score robustness by temperature and seed

Model	Temp	Seed	Mean	SD	Median	IQR	Unique	Δ	Change rate
gpt-4.1-mini	0	None	76.83	8.46	75.00	75.0–83.0	1.38	1.44	32.50%
gpt-4.1-mini	2	None	76.66	8.56	77.30	73.5–82.1	3.99	5.21	98.50%
gpt-4o-mini	0	None	79.38	10.68	75.00	75.0–82.5	1.18	1.38	16.00%
gpt-5.1	0	1871	79.29	7.95	78.00	77.2–82.6	1.82	1.13	61.00%

Notes: Constant sample of $n = 200$ biographies; each row uses 10 repeated versions. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography. Δ is the mean absolute pairwise difference across versions within biography. Change rate is the share of biographies with any score change. Lower Unique, Δ, and change-rate values indicate greater within-configuration stability.
Source: Author's calculations from versioned score-generation runs under `reproduce/social-standing/version_*/`.

The temperature comparisons show the expected pattern. For `gpt-4.1-mini` in the occupation-score exercise, the anchored prompt at temperature 0 produces 1.38 distinct scores per biography on average and an average pairwise difference of 1.44 points. At temperature 2, these increase to 3.99 and 5.21. In the social-standing exercise, the base prompt at temperature 0 produces 1.61 distinct scores and an average pairwise difference of 1.32 points, compared with 4.47 and 5.46 at

Table C.8: Social-standing robustness by temperature and seed

Prompt	Model	Temp	Seed	Mean	SD	Median	IQR	Unique	Δ	Change rate
Base	gpt-4.1-mini	0	None	77.61	9.40	80.00	72.5–83.5	1.61	1.32	51.00%
Base	gpt-4.1-mini	2	None	77.02	10.29	79.00	71.6–84.0	4.47	5.46	100.00%
Anchored	gpt-4.1-mini	0	None	74.72	11.81	76.50	70.0–82.0	1.77	1.74	63.00%
Anchored	gpt-4.1-mini	2	None	74.27	11.54	76.30	67.5–81.5	4.17	5.74	100.00%
Empty	gpt-4.1-mini	0	None	77.15	7.69	80.00	72.5–80.5	1.58	1.27	52.50%
Empty	gpt-4.1-mini	2	None	76.33	8.10	78.60	71.5–81.6	4.59	5.75	99.50%
Base	gpt-4o-mini	0	None	82.40	9.40	85.00	76.0–90.0	1.43	1.10	37.00%
Anchored	gpt-4o-mini	0	None	81.22	12.96	84.00	74.5–88.0	1.50	1.51	43.00%
Empty	gpt-4o-mini	0	None	82.69	7.49	85.00	80.0–90.0	1.34	1.05	34.00%
Base	gpt-5.1	0	1871	83.05	10.68	86.20	77.4–90.4	1.84	1.11	63.00%
Anchored	gpt-5.1	0	1871	80.38	9.57	80.20	75.2–87.2	2.21	1.60	74.50%

Notes: Constant sample of $n = 200$ biographies; each row uses 10 repeated versions. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography. Δ is the mean absolute pairwise difference across versions within biography. Change rate is the share of biographies with any score change. Lower Unique, Δ, and change-rate values indicate greater within-configuration stability.
Source: Author's calculations from versioned score-generation runs under `reproduce/social-standing/version_*/`.

temperature 2. The seeded low-temperature `gpt-5.1` runs are more stable than the high-temperature runs, but they still change for a non-trivial share of biographies. This is important: low temperature and seed settings reduce variation, but they should not be treated as substitutes for validation.

Across all four tables, three patterns matter for the workflow. First, prompts should define the measurement task as explicitly as possible. Anchors and examples are not cosmetic; they reduce the room for the model to invent its own scale. Second, temperature and seed settings should be documented with the same care as model names and prompts. Third, generated scores should be treated differently from extracted text fields. Extracted fields can be checked against the source biography; generated scores are interpretive summaries, so their stability across prompt, model, and parameter choices is part of their validation.

C.4 Open-Weight Model Comparison

This section compares open-weight model runs with the proprietary-model robustness checks reported in Appendix C.3. The comparison clarifies one limitation discussed in the main chapter's conclusion: open-weight models are useful for accessibility and local reproducibility, but they are not automatically equivalent substitutes for proprietary API models.

The open-weight runs use the same score-generation task as the runs before: each biography receives an `occupation_score` and a `social_standing` score on a 0–100 integer scale. The comparison summarizes repeated runs for `gpt-oss:20b`, `gpt-oss:120b`, and `llama3.1:70b`. As in Appendix C.3, *Unique* is the average number of distinct scores assigned to the same biography across versions, Δ is the mean absolute pairwise difference across versions within biography, and *Change rate* is the share of biographies whose score changes at least once. These are stability measures, not quality measures. A model can be stable because it has learned the task, but it can also be stable because it collapses onto coarse or repeated scores. Tables C.9 and C.10 report the corresponding open-weight results for occupation score and social standing.

Table C.9: Open-weight occupation-score robustness

Model	Temp	Seed	Mean	SD	Median	IQR	Unique	Δ	Change rate
gpt-oss:120b	None	None	75.16	9.62	77.70	71.6–80.7	4.24	4.17	99.48%
gpt-oss:20b	None	None	73.08	11.65	76.50	69.5–79.5	3.79	6.04	99.48%
llama3.1:70b	None	None	78.09	7.69	76.50	75.0–82.5	2.46	5.28	78.12%
gpt-oss:20b	0	None	73.35	12.99	75.00	75.0–80.0	1.02	0.06	1.56%
llama3.1:70b	0	None	78.10	9.89	75.00	75.0–85.0	1.00	0.00	0.00%
llama3.1:70b	0	1871	77.71	8.91	75.00	75.0–75.0	1.00	0.00	0.00%

Notes: Rows use the common overlap sample of $n = 192$ biographies. Repeated-version counts vary by configuration from 3 to 10. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography; Δ is the mean absolute pairwise difference across versions within biography; Change rate is the share of biographies with any score change. Rows are restricted to open-weight/local model runs.
Source: Author's calculations from versioned open-weight LLM runs.

The occupation-score results illustrate the distinction between stability and quality. The unseeded open-weight rows show substantial variation across versions. For example, `gpt-oss:20b` has an average pairwise difference of 6.04 points and changes for 99.48 percent of biographies. At temperature zero, however, the same model becomes almost perfectly stable ($\Delta = 0.06$, change rate 1.56 percent). This stability should not be read as sufficient evidence of quality. A deterministic or low-temperature model can produce repeated values without necessarily having quality results. The same diagnostic for social standing is reported in Table C.10.

Table C.10: Open-weight social-standing robustness

Prompt	Model	Temp	Seed	Mean	SD	Median	IQR	Unique	Δ	Change rate
Base	gpt-oss:120b	None	None	79.61	10.36	81.40	76.7–86.2	5.58	4.85	100.00%
Anchored	gpt-oss:120b	None	None	75.00	12.51	76.80	69.0–84.1	5.43	5.69	99.48%
Base	gpt-oss:20b	None	None	79.30	12.77	82.60	73.3–88.1	5.59	6.78	100.00%
Anchored	gpt-oss:20b	None	None	74.96	13.80	78.50	70.0–83.5	4.31	6.78	100.00%
Base	llama3.1:70b	None	None	84.79	14.37	91.10	81.0–94.3	3.15	5.53	98.96%
Anchored	llama3.1:70b	None	None	78.19	16.87	85.50	67.5–90.0	3.22	6.68	96.35%
Anchored	gpt-oss:20b	0	None	74.86	14.56	80.00	70.0–85.0	1.03	0.20	2.60%
Base	llama3.1:70b	0	None	84.57	15.84	90.00	80.0–95.0	1.00	0.00	0.00%
Anchored	llama3.1:70b	0	None	79.45	18.04	85.00	80.0–90.0	1.00	0.00	0.00%
Base	llama3.1:70b	0	1871	85.09	15.73	90.00	80.0–95.0	1.00	0.00	0.00%
Anchored	llama3.1:70b	0	1871	76.56	18.91	80.00	60.0–90.0	1.00	0.00	0.00%

Notes: Rows use the common overlap sample of $n = 192$ biographies. Repeated-version counts vary by configuration from 3 to 10. Mean, SD, median, and IQR summarize biography-level mean scores across versions. Unique is the mean number of distinct scores assigned to a biography; Δ is the mean absolute pairwise difference across versions within biography; Change rate is the share of biographies with any score change. Rows are restricted to open-weight/local model runs.
Source: Author's calculations from versioned open-weight LLM runs.

The social-standing results show the same issue more sharply. Some `llama3.1:70b` temperature-zero rows have zero measured change across versions, but their score distributions remain highly concentrated around round values. In contrast, the proprietary model results in Appendix C.3 often have more within-model movement but preserve more graded variation across biographies. This is why I argue that low-temperature stability is a necessary diagnostic.

As a subconclusion, open-weight models should be evaluated in the same versioned framework as proprietary models. Local availability improves accessibility and reduces dependence on remote APIs, but it does not remove the need for measurement checks. In this application, open-weight models can produce plausible score distributions, yet their behavior differs from the proprietary runs in level, dispersion, and stability.

C.5 HISCO and Occupation Validation

The previous checks show whether the model is consistent with itself. This section asks a stricter question: do its occupation classifications resemble external historical coding systems? Because no hand-coded benchmark exists for the Degener occupations, I use external HISCO-coded occupations and HISCAM prestige orderings to test whether the workflow reproduces historically meaningful occupational structure.

The first benchmark tests whether the workflow can reproduce externally assigned HISCO[1] codes in Graham Mooney's HISCO-coded 1866 London occupational data (Mooney, 2016). These data contain occupational titles from death certificates recorded in London during the 1866 cholera epidemic and are associated with Mooney's study of diagnostic spaces in mid-Victorian London (Mooney, 2009). For this appendix, I use a deterministic 100-row sample stored locally as `occupation/inputs/Mooney2009Random.csv`. The file supplies externally coded occupation strings and reference HISCO labels for testing the coding workflow; it is not drawn from the Degener biographies. The sample is drawn with seed 1871. Each occupation is classified into HISCO in stages. The baseline workflow first asks for the major group, then the minor group conditional on the major choice, then the unit group conditional on the minor choice, and finally the full code conditional on the unit choice. This staged structure keeps each LLM decision local to a smaller codebook. Each configuration is run five times. A supervisor step then sees the five candidate classifications and selects a final code.

The experiment labels refer to prompt and input variants, not to different occupation samples. *Baseline* is the workflow described above. *More* and *Full* are progressively more context-rich variants of the same classification task; *Full* is the most context-rich

[1]For HISCO, see Marco H. D. van Leeuwen, Ineke Maas, and Andrew Miles, *HISCO: Historical International Standard Classification of Occupations*, Leuven University Press, 2002. For HISCAM, see Paul S. Lambert, Richard L. Zijdeman, Marco H. D. van Leeuwen, Ineke Maas, and Kenneth Prandy, "The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research," *Historical Methods*, 46(2), 2013, 77–89.

version reported in Table C.11. All rows use the same 100-occupation sample and five repeated runs per version.

Table C.11: HISCO classification consistency

Model	Experiment	Major	Minor	Unit	Full	Distance	Error Agreement (%)
gpt-5.1	Baseline	71.6	49.0	44.6	31.8	4.24	74.1
		(1.4)	(1.3)	(1.9)	(2.2)	(0.12)	
gpt-5.1	More	66.0	46.6	44.0	30.4	4.26	78.5
		(1.4)	(1.4)	(1.8)	(1.7)	(0.10)	
gpt-5.1	Full	80.0	57.8	53.2	36.8	3.44	77.1
		(0.9)	(0.7)	(0.7)	(0.7)	(0.05)	
gpt-5-mini	Baseline	75.2	56.0	51.0	32.8	3.87	71.4
		(1.0)	(1.7)	(1.4)	(1.2)	(0.11)	
gpt-5-mini	More	68.2	57.8	55.6	37.4	3.62	74.3
		(1.0)	(1.6)	(1.4)	(1.0)	(0.08)	
gpt-5-mini	Full	75.2	63.0	58.6	40.0	3.26	77.1
		(1.5)	(1.5)	(1.9)	(1.1)	(0.11)	
gpt-5-nano	Baseline	64.2	45.8	42.6	27.4	4.02	61.5
		(1.9)	(1.7)	(1.6)	(1.9)	(0.16)	
gpt-4o-mini	Baseline	29.8	24.6	23.6	15.6	4.81	77.7
		(1.6)	(1.7)	(1.7)	(1.6)	(0.20)	
gpt-4o-mini	More	51.0	40.0	37.8	27.0	4.72	78.8
		(1.1)	(0.6)	(0.4)	(1.7)	(0.05)	

Notes: Cells report means with the SD shown below in smaller text; accuracy values are in percentage points. Distance is average HISCO tree-step distance (0 exact; 2 same unit; 4 same minor; 6 same major; 8 different major; mean $\pm$ SD). Error Agreement (%) is the agreement rate among incorrect full-code predictions (conditional on being wrong). Average across versions (omitted columns): All agree 56.5%, Majority 92.9%, Avg unique 1.57. Runs per version: 5. Sample size per run: 100 occupations.
Source: Author's calculations from `occupation/inputs/Mooney2009Random.csv`, the HISCO/HISCAM reference files, and versioned occupation outputs under `occupation/outputs/version_*/`.

The classification table reports accuracy at four levels of the HISCO hierarchy: major, minor, unit, and full code. It also reports the average tree distance to the correct code. The distance measure is useful because not all errors are equally severe. A wrong full code within the correct unit group is a smaller error than a prediction in the wrong major group. The error-agreement column asks a different question: when the model is wrong, does it tend to be wrong in the same way across repeated runs? High error agreement can indicate a systematic disagreement between the source label and the model's occupational interpretation rather than pure noise.

A second classification check varies the amount of prompt and codebook context while holding the occupation sample fixed. In Table C.12, *Pure* denotes the direct codebook-label variant with minimal additional context, while *More* denotes the corresponding variant with additional prompt context. Both variants use the same 100-occupation sample and five repeated runs per version. This is because labels can be historically or semantically ambiguous when mapped into HISCO. For example, *druggist* is classified as a retail-proprietor while LLMs tend to classify as pharmacist.

Table C.12: HISCO classification prompt variants

Model	Experiment	Major	Minor	Unit	Full	Distance	Error Agreement (%)
gpt-5-mini	Pure	91.4	79.4	68.8	68.8	1.83	85.7
		(1.0)	(1.2)	(1.8)	(1.8)	(0.11)	
gpt-5-mini	More	91.8	83.4	81.4	81.4	1.24	70.3
		(1.5)	(1.5)	(1.5)	(1.5)	(0.08)	
gpt-4o-mini	Pure	36.4	29.2	23.6	23.6	3.09	82.3
		(0.5)	(0.7)	(1.4)	(1.4)	(0.15)	
gpt-4o-mini	More	77.0	66.2	64.2	64.2	2.46	65.5
		(1.7)	(1.3)	(1.3)	(1.3)	(0.13)	

Notes: Cells report means with the SD shown below in smaller text; accuracy values are in percentage points. Distance is average HISCO tree-step distance (0 exact; 2 same unit; 4 same minor; 6 same major; 8 different major; mean ± SD). Error Agreement (%) is the agreement rate among incorrect full-code predictions (conditional on being wrong). Average across versions (omitted columns): All agree 75.7%, Majority 96.7%, Avg unique 1.32. Runs per version: 5. Sample size per run: 100 occupations.
Source: Author's calculations from `occupation/inputs/Mooney2009Random.csv`, the HISCO reference files, and versioned occupation outputs under `occupation/outputs/version_*/`.

Table C.12 shows that additional context improves accuracy for the stronger models, but accuracy remains below 100%. The agentic design is useful because it separates major-level errors, full-code errors, repeated wrong answers, and supervisor decisions for direct inspection.

The next benchmark turns from classification to occupational prestige by asking the model to choose the higher-status occupation within sampled pairs. This task is intentionally easier than full HISCO classification. It does not require selecting one code among the full codebook; it only asks for a binary ordering between two occupations. The table therefore measures whether the model's comparative judgments align with an external occupational-status scale and whether repeated judgments are internally consistent across runs. Overall, this section experiments with an LLM workflow with categorization and "scoring".

Table C.13: Pairwise occupational-prestige validation

Model	Experiment	Accuracy	All agree	Majority	Avg unique
gpt-5.1	Baseline	57.5 (1.4)	88.0	100.0	1.12
gpt-5-mini	Baseline	57.5 (1.0)	85.5	100.0	1.15
gpt-5-mini	More	57.3 (1.8)	85.0	100.0	1.15
gpt-5-mini	Full	60.1 (1.1)	84.0	100.0	1.16
gpt-5-nano	Baseline	58.4 (1.4)	77.0	100.0	1.23
gpt-4o-mini	Baseline	52.8 (0.7)	87.0	100.0	1.13
gpt-4o-mini	More	53.8 (1.4)	83.5	100.0	1.17

Notes: Cells report means with the SD shown below in smaller text; accuracy values are in percentage points. All agree and Majority report consensus rates across runs; Avg unique is the average number of distinct choices per pair. Runs per version: 5. Sample size per run: 200 pairs.
Source: Author's calculations from HISCO/HISCAM reference files and versioned occupation outputs under `occupation/outputs/version_*/`.

Table C.14: Occupational-prestige ranking validation

Model	Experiment	Spearman	Pairwise mean	Pairwise median
gpt-5.1	Baseline	0.489 (0.002)	0.963	0.979
gpt-5-mini	Baseline	0.481 (0.011)	0.916	0.930
gpt-5-mini	More	0.468 (0.009)	0.911	0.932
gpt-5-mini	Full	0.479 (0.021)	0.889	0.928
gpt-5-nano	Baseline	0.399 (0.015)	0.824	0.851
gpt-4o-mini	Baseline	0.385 (0.035)	0.876	0.905
gpt-4o-mini	More	0.385 (0.015)	0.858	0.874

Notes: Cells report means with the SD shown below in smaller text. Pairwise mean/median summarize Spearman correlations across run pairs. Runs per version: 5. Sample size per run: 17–20 rank sets.
Source: Author's calculations from HISCO/HISCAM reference files and versioned occupation outputs under `occupation/outputs/version_*/`.

C.6 Human-AI Benchmark for Generated Scores

The main chapter uses LLM-generated occupation and social-standing scores as empirical variables. This section compares those generated scores with human-coded scores for the same 200 biographies. The human benchmark is the average across five humans, while the AI benchmark is the average across 14 anchored AI settings.

Table C.15: Human-AI benchmark for generated scores

Subject	Mean	SD	Median	IQR
Panel A (Occupation Score)				
Humans	66.80	11.42	70.00	57.2–76.0
AI	76.31	8.69	77.00	71.8–81.2
Panel B (Social Standing)				
Humans	68.82	10.49	71.00	61.0–75.8
AI	76.93	12.64	79.21	69.0–86.1

Notes: The Humans rows report biography-level means across five human-coded files ($n = 200$ biographies). The AI rows report biography-level means across 14 anchored AI settings for the same biographies. Spearman rank correlations between human and AI biography-level means are $\rho_{\text{occ}} = 0.64$ for occupation score and $\rho_{\text{soc}} = 0.58$ for social standing.

Table C.15 shows that the AI scores are higher than the human scores in both dimensions. For occupation score, the human mean is 66.80 and the AI mean is 76.31; for social standing, the human mean is 68.82 and the AI mean is 76.93. The rank correlations are moderate, with Spearman correlations of 0.64 for occupation score and 0.58 for social standing, so the AI scores capture part of the human ordering but should not be treated as identical to human-coded scores.