CAMBRIDGE
UNIVERSITY PRESS

RESEARCH PAPER

Prussia disaggregated: the demography of its universe of localities in 1871

Sascha O. Becker[1,2,3,4]* and Francesco Cinnirella[3,4,5,6]

[1]SoDa Labs @ Monash University, Melbourne, VIC, Australia, [2]CAGE, University of Warwick, Coventry, UK, [3]CEPR, London, UK, [4]CESifo, Munich, Germany, [5]University of Bergamo, Bergamo, Italy and [6]Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark
*Corresponding author. E-mail: sascha.becker@monash.edu

Abstract

We provide, for the first time, a detailed and comprehensive overview of the demography of more than 50,000 towns, villages, and manors in 1871 Prussia. We study religion, literacy, fertility, and group segregation by location type (town, village, and manor). We find that Jews live predominantly in towns. Villages and manors are substantially segregated by denomination, whereas towns are less segregated. Yet, we find relatively lower levels of segregation by literacy. Regression analyses with county-fixed effects show that a larger share of Protestants is associated with higher literacy rates across all location types. A larger share of Jews relative to Catholics is not significantly associated with higher literacy in towns, but it is in villages and manors. Finally, a larger share of Jews is associated with lower fertility in towns, which is not explained by differences in literacy.

1. Introduction

Nineteenth-century Prussia, the dominant state in the German Empire, provides for a fascinating setting to study many of the most fundamental questions in 19th-century economic history. Its relatively uniform institutional setting and extremely rich sources of surviving data provide researchers with unique opportunities to study the link between religion, literacy, fertility, economic development, and many other topics.

Prussian census data have long been recognized as a source of high-quality disaggregated data. Data for more than 400 Prussian counties (*Kreise*) have been used for demographic research for at least 25 years [see e.g., Galloway *et al.* (1994) and Lee *et al.* (1994) for research into the fertility decline in Prussia]. Compared to earlier research which either looked at national aggregates or used district-level data for the roughly three dozen districts (*Regierungsbezirke*), such as the work of the

Princeton fertility project [e.g., Knodel (1974)], county-level data allowed for richer regression analyses drawing on hundreds of geographic units. Over the last few decades, researchers have digitized large amounts of Prussian county-level data and made them available for download [e.g., Galloway (2007); Becker *et al.* (2014)]. This led to a further burst of research, some of which has been summarized in Becker and Woessmann (2019).

At the same time, researchers of Prussian and German economic history have used city-level data [e.g., Cantoni (2015); Hornung (2015)] to produce fascinating insights into drivers of economic development. City-level data have certain advantages over county-level data. In the case of Prussia, census data have been available for approximately 1,000 towns, over several decades [see e.g., Hornung (2015)]. To the extent that cities have been more extensively researched, urban historians have amassed large amounts of data on German cities going back even before the beginnings of universal censuses. This has permitted panel analyses that stretch over several centuries, going back into the middle ages [e.g., Cantoni (2015); Becker and Pascali (2019)]. City-level data, however, have the major shortcoming that it ignores rural populations. In the Prussian context, less than one-third of inhabitants lived in towns, as our analysis shows. As a consequence, city-level analyses are silent about more than two-thirds of the Prussian population.

Although county-level data cover all of the Prussian population, they do not always separate out urban and rural populations. Many fascinating questions cannot be answered with either county-level or city-level data. For instance, how are the three main religious denominations of Prussia distributed across towns, villages, and manors? Do Protestants reside more in urban areas than Catholics? Do Jews live mostly in towns? In religiously mixed areas, do Protestants and Catholics live side by side in all localities? Or is there segregation along denominational groups? If so, is segregation more pronounced in rural areas? Does segregation have implications for relevant economic outcomes such as literacy, one of the ingredients of economic success in 19th-century Europe? Is the religion–literacy link homogeneous across location types? Does fertility differ between towns, villages, and manors, and does its link with religion vary across location types?

In this paper, we provide an answer to all of these questions. For the first time, we can provide a detailed and comprehensive overview of the number of inhabitants, household structure (number of inhabitants below age 10, and older), religious denomination, and literacy for the universe of towns, rural villages, and manors in 1871 Prussia.

In particular, our contributions are as follows: first, at the descriptive level, our newly digitized data on more than 50,000 administrative units help to understand (i) whether and to what extent religious denominations were segregated across locations; (ii) the level of literacy across different types of administrative units; and (iii) the relationship between religious denomination, literacy, and fertility across different types and sizes of administrative units.

Second, from a methodological perspective, the extremely high level of disaggregation of the data allows us to study whether, at least in the Prussian case, official statistics at a more aggregated level are prone to an ecological fallacy. For example, we can test whether high levels of religious diversity observed at the county level "hide" relatively high levels of religious segregation across rural districts or manors. Closely related, locality-level regression analyses permit us to include county-fixed effects, thus exploiting within-county variation.

Finally, we make publicly available digitized locality-level data which, to our knowledge, have never been available to researchers before. Many smaller locations have since been incorporated into larger villages and towns and their distinctive features have thus "disappeared" in those larger administrative units. Our new data will give the opportunity to economic and social historians to reconstruct, at the time of the German unification in 1871, the demographic, religious, and educational structure of each location, giving renewed impetus to research on rural economic history.

Our main results are multiple. We find that as much as 79% of Jews in Prussia lived in towns. For comparison, only 35% of Protestants and 26% of Catholics lived in towns. For the first time, we can shed light on the level of literacy in rural administrative units. We find that about 73% of people living in manors were able to read and write. This is an astonishingly high level, considering the rural nature of these administrative units with an average size of 126 people. Such a literacy level compares favorably to the whole of France (62%) and it was much higher than Italy or Spain in the same period. The literacy rate in Italy at the time of unification in 1861 was only 27% and around 35% in 1871; the literacy rate in Spain was around 30% by 1870 [Cipolla (1969)]. In fact, the average literacy rate in Prussian manors in 1871 was much higher than in large Spanish cities such as Barcelona (0.50), Madrid (0.66), or Valencia (0.40) three decades later, in 1900 [Cinnirella *et al.* (2020)]. Compared to Italy, the literacy level in the Prussian manors was considerably higher than the literacy rate of Turin (0.58), the city with the highest level of literacy in 1871 Italy [Federico *et al.* (2019)].

We construct a dissimilarity index to document group segregation along religious lines and for literate and illiterate people. Our results unveil a significant level of denominational segregation within counties, suggesting that Protestants and Catholics tended to cluster in localities. In particular, we find that denominational segregation was more pronounced in villages and manors than in towns. This is an important result for those working with county-level data as county averages seem to hide a significant amount of variation across localities. On the contrary, we find relatively low levels of segregation between literate and illiterate people.

Our new data set allows us to perform regression analyses accounting for county-fixed effects. We find that the relationship between literacy and population size decreases for manors, whereas it is positive in the case of towns and villages.

The regression analysis shows that in towns, villages, and manors, a larger share of Protestants is associated with higher literacy. Although a larger share of Jews is associated with higher literacy in villages and manors, this is not the case in towns.

Finally, we find a large and significant negative association between the share of Jews in towns and fertility, measured as the number of children below age 10 over the total number of women (child–woman ratio). Importantly, this negative association is not explained by differences in literacy, consistent with a possible cultural explanation for the lower fertility.

Our regression results should not be interpreted causally. In fact, it is not the objective of this paper to identify any causal relationship between the socio-economic variables at our disposal. The aim is to provide a comprehensive cross-sectional description of some religious and socio-economic patterns of Prussian demography in 1871. Our hope is that the analysis provided in this paper will spark new research ideas.

The rest of the paper is structured as follows: we discuss the related literature in section 2 and introduce the data in section 3. We perform a descriptive analysis in

section 4 and document the extent of group segregation in section 5. In section 6, we present the results from the regression analysis and section 7 concludes providing some hints for future research.

2. Literature

The previous literature on demographic and religious aspects of 19th century Germany had a strong focus on the determinants of the fertility transition. In the 1960s and 1970s, the so-called European Fertility Project was carried out and Knodel analyzed the demographic patterns for Germany [Knodel (1974); Coale and Watkins, (1986)]. Yet, the studies on Germany mainly relied either on individual-level data for a very small sample of villages, or on region-level aggregates.

Knodel and Maynes (1976) review evidence on urban–rural and regional differences in marriage patterns in Germany around 1880. They use data from the 1880 census of the Imperial Statistical Bureau (*Statistisches Reichsamt*) at the district level (*Regierungsbezirk*, n = 74). They find that the proportions of people never marrying, and age at marriage were higher in urban areas and lower in the countryside. Knodel (1977) reviews results on urban–rural differences in nuptiality, fertility, illegitimacy, and infant mortality across German states. He finds that urban–rural residence patterns account for relatively small differences in the proportion of single women, overall fertility, and infant mortality. On the contrary, there is no clear urban–rural differential in illegitimacy, male nuptiality, and marital fertility. As in the previous case, these results are based on census data aggregated at the district level. Knodel (1979) uses a sample of 12 German village genealogies to document changes in reproductive patterns in Germany during the 19th century. He finds large differences in the timing of the emergence of fertility limitation. Religious aspects of fertility limitation behavior are not addressed.

Brown and Guinnane (2002) use district level data (138 rural districts and 38 urban districts) from Bavaria to study the determinants of the fertility transition in the 19th century. They find that Bavaria's fertility transition occurred late with respect to neighboring Prussia. Catholicism was strongly positively related to fertility, underlining the potential importance of cultural norms. Yet, the authors find also an important role for economic motives: areas with higher economic opportunities for women experienced the most rapid decline in fertility.

Their results differ from those put forward by Galloway *et al.* (1994). Using data for 407 counties in Prussia in the period 1875–1910, they find that religion is by far the most important determinant of fertility levels, followed by ethnicity. Galloway *et al.* (1998) analyze the fertility decline in cities and rural counties in Prussia for the same period. They find that urban fertility was much lower than rural fertility because of changes in female labor force participation, communications, improvements in education, and a reduction in infant mortality.[1] Galloway has given a substantial contribution to the study of Prussian economic history also with the creation of the Galloway Prussia Database which provides, at the county level, many socioeconomic variables from different Prussian censuses for the period 1861–1914.[2]

---

[1]More recently, Klüsener and Goldstein (2016) reintroduced geography in the debate on the determinants of the fertility decline. By applying spatial econometrics to county-level data in Prussia, they find a strong degree of spatial clustering in the fertility decline.

[2]The Galloway database can be accessed at http://www.patrickgalloway.com/prussia.htm.

Becker *et al.* (2014) created a complementary database which goes further back in time and provides a rich collection of variables covering the period 1816–1901.

Recently, there has been an increased interest in the relationship between Protestantism and the accumulation of human capital.[3] Becker and Woessmann (2009) use county-level data for 1871 Prussia to take a fresh look at Weber's hypothesis of a Protestant work ethic. Their main finding is that, although Protestantism affects literacy, after controlling for literacy, Protestantism has no residual impact on economic outcomes. Cantoni (2015) addresses a similar question using city-level data for German territories for the period 1300–1900. He finds no evidence of a Protestant advantage in terms of urbanization. The difference in the results is likely explained by the different unit of observation, namely the county and the city. With our newly collected data, we can shed light on the relationship between literacy and religious denominations by locality.

Goldstein and Klüsener (2014) use county-level census data to document how different modes of production in agriculture between west and east Elbian territories might explain differences in non-marital fertility toward the end of the 19th-century. Related to our story, they argue that the rural population of the west lived more concentrated in rural villages (with comparatively more social control) whereas the rural population in East Elbia was more dispersed on rural estates. Different degrees of social control and more seasonal work migration might have generated persistent differences in non-marital fertility patterns. With our data at the highly disaggregated level of more than 50,000 localities, one can shed more light on different demographic patterns in rural areas in eastern and western Prussian territories. Relatedly, Cinnirella and Hornung (2017) document the relationship between the power of landed elites and marriage patterns. They find no systematic evidence in favor of the hypothesis that noble landowners directly interfered with marriage decisions. Yet, they find a robust negative relationship between education and the share of married women.

Ogilvie and Küpker (2015) study human capital levels in Wuerttemberg between 1610 and 1899. The analysis based on individual-level data from two selected locations finds that literacy (measured by the ability to sign their marriage inventory) declined significantly with age, controlling for family wealth at marriage. The authors interpret this finding as evidence of a progressive decay of writing skills after leaving school, consistent with the idea that literacy was not relevant for economic life.

3. Data

Starting with the first full-scale population census in 1816, the Royal Prussian Statistical Office collected a large amount of economic and demographic data. Although individual-level census data have not been preserved to this day, tens of thousands of pages of county-level data have survived in archives. Thanks to the Prussian proverbial orderliness and thoroughness, high quality data for the Prussian counties (*Kreise*) covering nearly the whole range of the 19th century are at researchers' disposal. These data provide a unique source for empirical research in economic history, with the potential to study historical micro-regional data with modern microeconometric methods.

---

[3]See Becker *et al.* (2016) for an extensive literature review.

For this paper, we collected data for the universe of localities in 1871 Prussia provided by the official population census (*Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung nach den Urmaterialen der allgemeinen Volkszählung vom 1. December 1871*). We collected information on the population (by gender), religious denomination (Protestant, Catholic, Jewish, other Christian religion, and other religion), number of people below age 10, and number of literate and illiterate people aged 10 or older. All these variables are provided for every single locality, i.e., towns (*Stadt*), villages (*Landgemeinde*), and manors (*Gutsbezirk*).[4] In total, we have data for 54,270 localities. The results presented below are based on 1,285 towns, 37,783 villages, and 15,202 manors. Descriptive statistics are reported in Table 1.

The average size of towns was 6,204 inhabitants, 386 inhabitants for villages, and only 126 inhabitants for manors (see Table B1 in the Appendix).[5] The urbanization rate in 1871 Prussia, i.e., the share of people living in towns was around 30%. Therefore, although the number of towns is only ca. 2% of the total number of localities (Table 1), they represent close to one-third of the whole population.

4. Descriptive analyses

4.1 Religion

In Table 2, we show the share of religious groups by type of locality. In towns, 67% are Protestants, 29% are Catholics, and 3% Jews. In villages, we find about 71% of Protestants and 29% of Catholics and a very small share of Jews. We find a similar distribution across manors. The most interesting result regards the Jewish population, 79% of which lived in towns (see Table 3).

For each type of locality (town, village, and manor), we can compute the Herfindahl index (also known as the index of diversity) which indicates the probability of sampling two individuals of different denominations.[6] In Figure 1, we display the distribution of the Herfindahl index for the entire Prussia and by locality type. As expected, we find that there is much more religious diversity in towns (0.21) than in villages (0.08) or in manors (0.12) (see Table B1 in the Appendix).

---

[4]There are few cases (*n* = 213) of localities "not incorporated" (*nicht incommunalisierte Wohnplätze*) which are not analyzed separately but included in the category of manors. For the county of Namslau (in the province of Silesia), data on religion and literacy are missing for manors. We therefore drop this county from the dataset (*n* = 132). We drop 191 observations where either demographic, religious, or literacy status is completely missing. There are—very few—typos in the printed originals for which there was no obvious correction. As described in the Appendix, the Prussian Statistical Office provided corrections in the appendix of each volume, which we took on board, but in a very low number of cases, there was no correction provided. As a result, we drop few observations in which the share of males, females, or literacy rates is, respectively, larger than 1.01 (*n* = 12). We do the same if the sum of the religious shares in a locality is above 1.01 (*n* = 18). We chose the value of 1.01 to allow for rounding. Regarding the data on literacy, it is important to note that, when constructing the literacy rate, we count individuals whose literacy status is missing as illiterate. For more details about the data sources see the Data Appendix A.

[5]It is important to note that there is no formal population threshold which defines towns and villages. In general, the legal status of a locality was granted by the local lord and/or political authority.

[6]We refer to this as the Herfindahl index for simplicity. In reality, we use (1-HHI), i.e. one minus the Herfindahl-Hirschman Index. HHI is an index of concentration, while we look at (1-HHI) as an index of diversity.

Table 1. Descriptive statistics

Variables	Obs.	Mean	Perc. 10	Median	Perc. 90	Std. dev.	Min.	Max.
Population size (thousands)	54,270	0.451	0.045	0.195	0.758	4.063	0.001	825.975
Share—males	54,270	0.488	0.441	0.487	0.536	0.047	0.000	1.000
Share—females	54,270	0.512	0.464	0.513	0.559	0.047	0.000	1.000
Sex ratio	54,266	0.976	0.788	0.950	1.156	0.381	0.000	49.667
Child–woman ratio	54,266	0.496	0.355	0.500	0.633	0.140	0.000	7.500
Share of Protestants	54,270	0.713	0.006	0.977	1.000	0.398	0.000	1.000
Share of Catholics	54,270	0.282	0.000	0.014	0.990	0.397	0.000	1.000
Share of other Christians	54,270	0.002	0.000	0.000	0.000	0.021	0.000	0.933
Share of Jews	54,270	0.003	0.000	0.000	0.003	0.014	0.000	0.524
Share of other religion	54,270	0.000	0.000	0.000	0.000	0.001	0.000	0.237
Share literate	54,270	0.795	0.474	0.877	0.979	0.207	0.000	1.000
Herfindahl index	54,270	0.096	0.000	0.012	0.372	0.150	0.000	0.977
Town	54,270	0.024	0.000	0.000	0.000	0.152	0.000	1.000
Village	54,270	0.696	0.000	1.000	1.000	0.460	0.000	1.000
Manor	54,270	0.276	0.000	0.000	1.000	0.447	0.000	1.000

Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Table 2. Share of denominations by locality

	Obs.	Mean	Std. dev.
Town			
Share of Protestants	1,285	0.672	0.363
Share of Catholics	1,285	0.292	0.349
Share of Jews	1,285	0.035	0.051
Village			
Share of Protestants	37,783	0.707	0.409
Share of Catholics	37,783	0.287	0.409
Share of Jews	37,783	0.003	0.012
Manor			
Share of Protestants	15,202	0.730	0.371
Share of Catholics	15,202	0.269	0.370
Share of Jews	15,202	0.001	0.006

Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Table 3. Share of denominations living in towns

Share of denominations living in towns	Obs.	Mean
Share of Protestants	1,285	0.347
Share of Catholics	1,285	0.265
Share of Jews	1,285	0.791

Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

4.2 Literacy

The literature has used various ways to measure historical literacy rates, e.g., signatures in marriage registers or age heaping, using the fact that self-reported age is often rounded [see A'Hearn *et al.* (2009)]. The census reports the number of people aged 10 or older who are able to read and write, or illiterate. Using this information, we construct the literacy rate for each locality.[7] The distribution of the literacy rate for the whole of Prussia and by the different type of localities are shown in Figure 2. It is well known that the average level of literacy in 19th-century Prussia, around 80%, was high in both absolute and relative terms. What, so far, was less known is the distribution of literacy in villages and manors. Although the distribution of literacy is left skewed in the case of villages, it appears to be more uniform in the case of

---

[7]We compute the literacy rate as the ratio of the number of people aged 10 or above who are able to read and write over the total number of people aged 10 or above. As mentioned above, we do *not* consider the individuals whose literacy status is missing as literate.
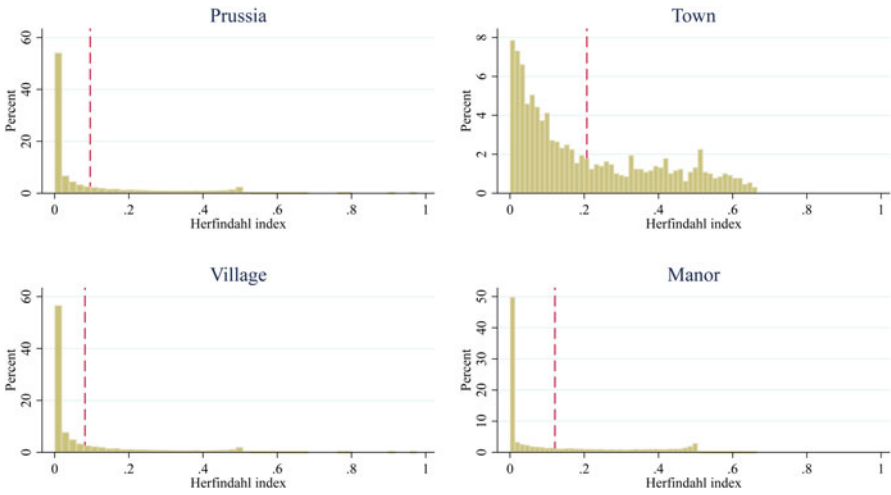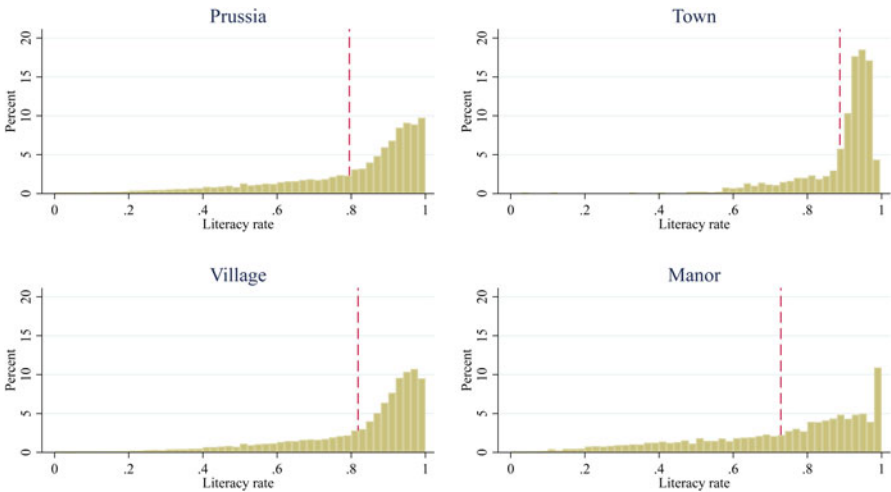
Figure 1. Religious diversity–Herfindahl index by locality.
Note: The Herfindahl index is computed across five denominations: Protestant, Catholic, other Christian, Jew, and other religion. The vertical dashed line indicates the sample mean. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Figure 2. Distribution of literacy by locality.
Note: The literacy rate is computed as the ratio of the number of people above age 10 able to read and write over the total number of people above age 10. The vertical dotted lines indicate the sample mean. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

manors. In particular, we find that even in small manors, on average, about 73% of the population was able to read and write (compared to 89% in towns and 82% in villages). Therefore, the literacy rate in manors in Prussia was higher than the country-wide
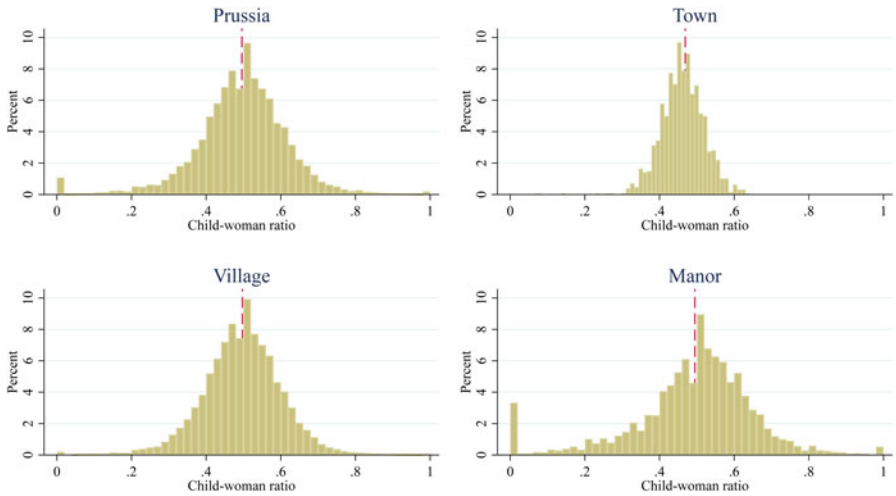
Figure 3. Distribution of fertility by locality.
Note: The child–woman ratio is computed as the number of children under age ten over the number of women. Only for display purposes, we omit from the graph 104 localities in which the child–woman ratio is larger than one.
Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

literacy rate in many European countries. For example, it compares with the overall literacy rate of 69% in France in 1870 and it is much higher than the literacy rate in Italy in 1870 (32%) or in Spain (26%) in 1860 [Cipolla (1969)].

4.3 Fertility

To study differences in fertility across localities, we compute the ratio of children below age 10 over the total number of women in the population (child–woman ratio) as our measure of fertility.[8] It is important to stress that we standardize by the number of women because including men could introduce a significant bias due to male deployment and casualties in the Franco–Prussian war of 1870/71, and because of gender-biased internal migration to larger settlements such as towns.[9]

This index for fertility is 0.5 for the whole of Prussia, i.e., there are 0.5 children under the age of 10 for each woman (of any age) in Prussia in 1871. The average by locality provides interesting insights: whereas fertility was comparatively lower in towns (0.469), we do not find a significant difference in fertility levels between villages (0.497) and manors (0.495) (Figure 3).[10]

---

[8]Note that the data do not allow us to measure the number of children ever born, i.e., we cannot explicitly account for child mortality.

[9]It should be noted that the child–woman ratio can take values over one when the number of children below age 10 is larger than the number of women. Indeed, we have 104 cases where this is the case. For display purposes we omit these observations in Figure 3 but we use them in the regressions.

[10]It should be noted that if we standardized the number of children below age 10 *per capita*, thus including also men in the denominator, we would find that fertility levels in manors are slightly but significantly higher than in villages.

5. Religious segregation

One of the advantages of our data is the possibility to study how two or more groups (e.g., religious denominations) are distributed across localities (towns, villages, and manors). In fact, obtaining a clearer understanding of how religious groups clustered locally is important as it sheds light on potential (i) interactions between the groups and (ii) spillover effects. The segregation analysis can also inform us about the potential ecological fallacy of analyzing data at a more aggregate level such as the county.

To study group segregation we use the dissimilarity index D which is the most common measure of group segregation. For the case of two groups (e.g., Protestants and Catholics), the dissimilarity index is defined as follows:

$$D = \frac{1}{2} \sum_i^N \left| \frac{p_i}{P} - \frac{c_i}{C} \right|$$

where p_i is the number of Protestants in the ith locality; P is the total number of Protestants in the large geographic entity for which the index of dissimilarity is calculated (in our case the county); c_i and C refer, respectively, to the Catholic population in the locality and in the county.

The index can be interpreted as the share of a group that would have to move to a different locality (town, village, or manor) to produce a distribution that matches that of the larger area, e.g., the county.[11] The index can take on values between zero and one. A value of the index close to one indicates that the groups are highly segregated, for example nearly all Protestants live in towns whereas nearly all Catholics live in villages (or manors); a value of the index close to zero indicates that the groups are not segregated and that the distribution in each locality closely matches the distribution at the county level. It is important to note that the index D is "scale free," i.e., it does not depend on the population size of the county or on the relative shares of the groups. We provide a simple example to explain how the index works. Let's assume that there are two counties, A and B. Within each county there are two localities, 1 and 2. In county A, there are 100 people in total, 90% Catholics and 10% Protestants. All Catholics ($n = 90$) live in locality 1 and all Protestants ($n = 10$) in locality 2. In this county, the groups are perfectly segregated, i.e., the dissimilarity index D for county A is 1. In county B there are also 100 people, 80% Catholics and 20% Protestants. In locality 1, there are 20 Catholics and 5 Protestants; in locality 2, there are 60 Catholics and 15 Protestants. The dissimilarity index D for county B is zero as the proportion of Catholics over Protestants is always 4:1, for both locations, and for the county as a whole.

Figure 4 shows the distribution of the dissimilarity index D for Protestants and Catholics across counties. The average is about 0.55, i.e., we would need to move 55% of a denomination between localities to match the distribution at the county level. This value for the dissimilarity index suggests that there is a substantial level of religious segregation in 1871 Prussia.[12] This important result is a warning for those,

---

[11]The index is "symmetric" in the sense that it does not matter which group is moved to match the distribution at the larger level.

[12]To put this value in context, gender occupational segregation in the USA is equal to 0.25 whereas multi-group occupational segregation by ethnicity is equal to 0.12 [Alonso-Villar et al. (2012)].
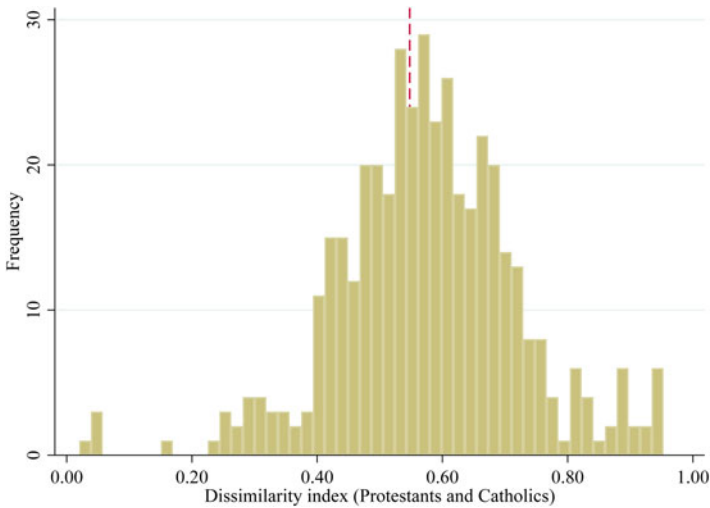
Figure 4. Distribution of dissimilarity index (Protestants and Catholics) across counties.
Note: The mean of the dissimilarity index across counties is 0.55. Only for display purposes, we remove the bar at zero that indicates the city-counties (Stadtkreis) for which, by definition, locality and county coincide. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

including the authors, who have been working with county-level data. In fact, in many cases the distribution of denominations at county level deviates significantly from the distribution at lower levels of administrative units. We show in Figure 5 how the segregation index is distributed geographically. The darkest regions, that is those with the higher levels of religious segregation, are in the Rhineland and Hessen-Nassau province (in the west) and in East Prussia (in the east).

In Figure 6, we report the dissimilarity index computed over all five denominations (Protestants, Catholics, Jews, other Christians, and other religion). As one can see the distribution of D does not change by much, and the mean is 0.54. For this reason, and because Protestants and Catholics are by far the largest denominations in Prussia, the following segregation analysis will be based only on these two main groups.

5.1 Religious segregation in urban and rural localities

To assess the contribution of urban and rural localities to religious segregation, we perform the segregation analysis separately for towns and rural localities, i.e., villages and manors.[13] In Figure 7, we report the kernel density estimates of the distributions of D separately for towns and rural localities. As one might expect, most of the religious segregation takes place across rural localities, i.e., villages and manors tend to be either predominantly Protestant or Catholic. On the contrary, there is much less segregation across towns, although the right tail of the distribution indicates the presence of counties with highly segregated towns. For example, the county of Zell, near Koblenz, in the Rhineland province, has a dissimilarity index equal to 0.87. In this county there are two towns, Zell and Trarbach, whose populations are

---

[13]We consider only counties in which there are at least two towns for the urban sample and two villages or manors for the rural sample.
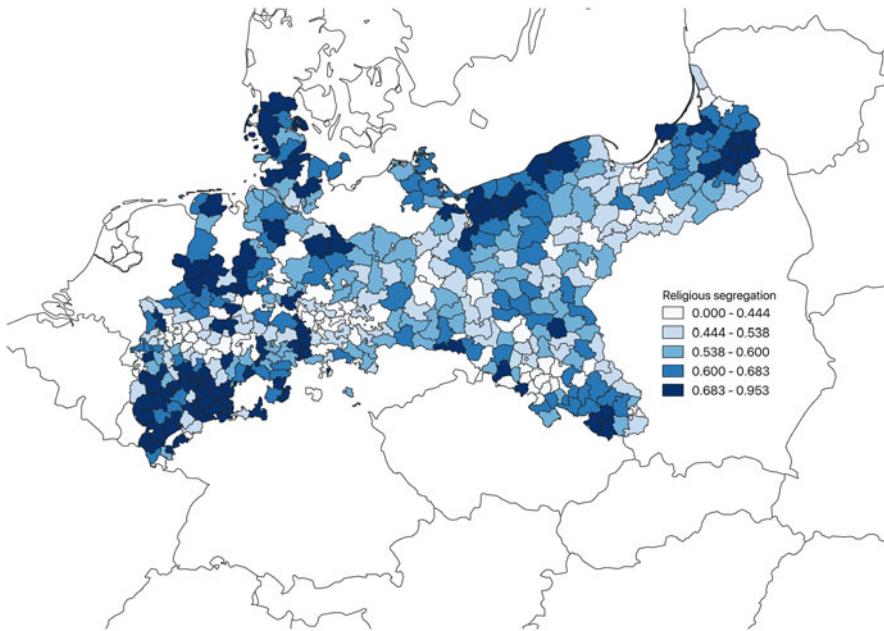
Figure 5. Geographic distribution of religious segregation.
Note: Dissimilarity index for Protestants and Catholics. The color graduation is based on five quantiles. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
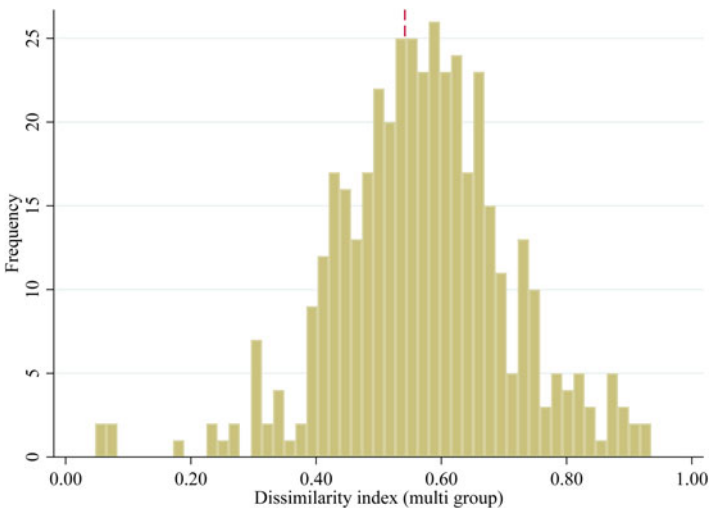

Figure 6. Distribution of dissimilarity index (multi-group) across counties.
Note: The mean of the dissimilarity index across counties is 0.54. Only for display purposes, we remove the bar at zero that indicates the city-counties (Stadtkreis) for which, by definition, locality and county coincide. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
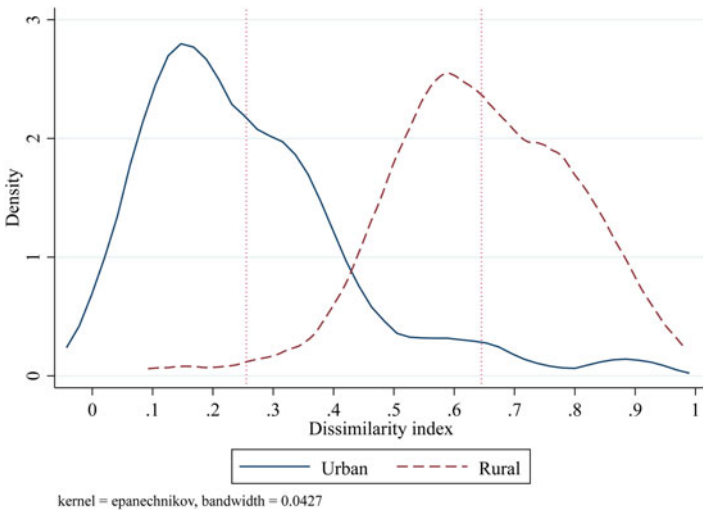
Figure 7. Distribution of dissimilarity index across counties for urban and rural localities.
Note: Kernel density estimates for urban (towns) and rural (villages and manors) localities. When computing the dissimilarity index for towns we discard 63 city-counties where the dissimilarity index is 0 by definition. The mean of the urban dissimilarity index is 0.25 whereas for rural one is 0.64. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

denominationally highly clustered. The town of Zell is predominantly Catholic with 2,223 Catholics, 60 Protestants, and 40 Jews; Trarbach has 1,475 Protestants, 227 Catholics, 1 Jew, and 1 from another Christian denomination. The two towns are only about 10 km distant from each other.

5.2 Human capital segregation

Since we have data on the number of literates and illiterates for each locality, we can compute the dissimilarity index for these two groups to study if people in Prussia clustered in high vs. low human capital groups.[14] The histogram of the human capital dissimilarity index is shown in Figure 8. The first thing to note is the relatively low level of human capital segregation with an average of 0.23. This means that the distribution of literates and illiterates across localities does not deviate much from the distribution of the two groups at the county level. We notice also the relatively low level of variance compared to religious segregation.

We show in Figure 9 the geographic distribution of human capital segregation in 1871 Prussia. We find relatively high levels of human capital segregation in Hessen-Nassau, Hannover, and in the western part of the province of Prussia (regions of Danzig and Marienwerder).

Also in this case, we can try to disentangle the urban and rural contributions to the human capital segregation of the county. Thus, we compute the dissimilarity index for the group of literates and illiterates in, respectively, urban and rural localities. The

---

[14]Note that we do not consider the individuals whose information on their literacy status is missing.
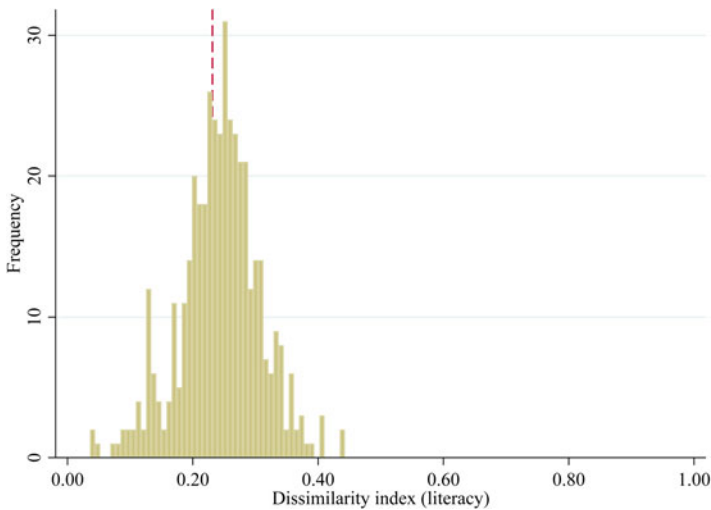
Figure 8. Distribution of dissimilarity index for literacy across counties.
Note: The mean of the dissimilarity index for literacy is 0.23. Only for display purposes, we remove the bar at zero that indicates the city-counties (Stadtkreis) for which, by definition, locality and county coincide. *Source:* Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

kernel density estimates of the respective distribution (with the relative mean) are shown in Figure 10. Consistent with the findings on religious segregation, we find a relatively higher level of human capital segregation in rural localities than in urban centers. However, in both cases the level of human capital segregation is lower than the religious segregation.

At this point one could ask: How much of the human capital segregation is "explained" by religious segregation? A higher level of religious segregation could mechanically result in a high level of human capital segregation.[15]

To address this question, we first compute the dissimilarity index for literacy for localities with above-average and below-average Protestant shares. The distributions of the respective indexes are displayed in Figure 11. We find that, in a given county, localities with above-average and below-average Protestant shares have similarly low levels of human capital segregation. This result suggests that, whereas Protestants and Catholics tended to be segregated across (mainly rural) localities (see Figure 7), we do not observe differential human capital segregation by religious majorities.

Second, we can plot the human capital dissimilarity index against the religious dissimilarity index (Figure 12). As one can see, almost all of the religious segregation indexes lie below the 45 degree line suggesting that increasing levels of religious segregation are less than proportionally related to increments in human capital segregation. Referring to the previous example of the county of Zell which has a level of religious segregation equal to 0.87 (!), the dissimilarity index for literacy is only 0.19.

---

[15]This is only possible if there is not a perfect correspondence between Protestantism (Catholicism) and literacy (illiteracy).
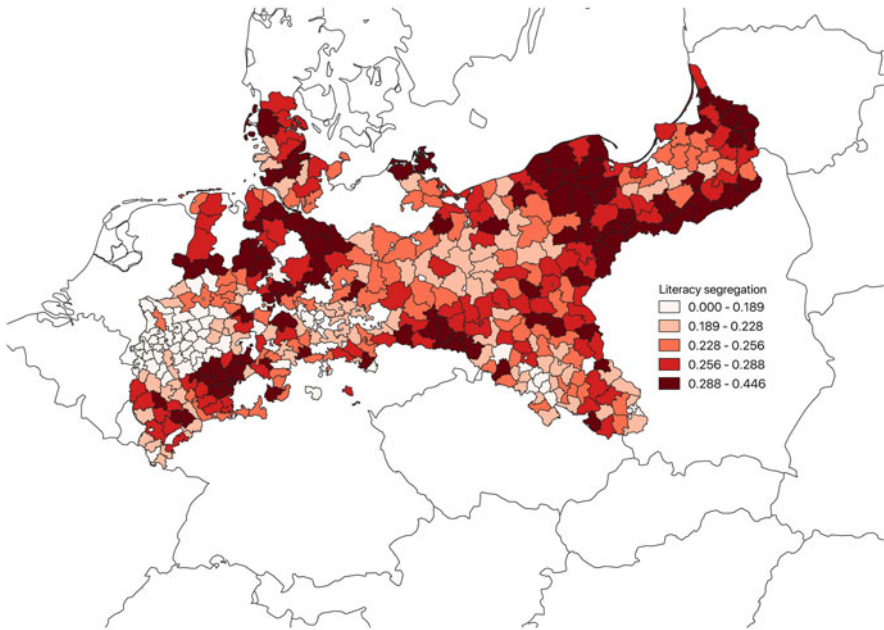
Figure 9. Geographic distribution of human capital segregation.
Note: Dissimilarity index for literates and illiterates. The color graduation is based on 5 quantiles. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

6. Regression analyses

So far, we have documented how religion, literacy, and fertility were distributed across localities. In this section, we study the relationship between these variables in a regression framework. In particular, we analyze, for the whole of Prussia and by type of locality (i) the relationship between literacy and population size; (ii) the relationship between literacy and religious denomination; and (iii) the relationship between fertility and religious denomination.

Some of these relationships have been studied before in the literature and with more sophisticated econometric analyses. For example, the relationship between religious denominations and literacy has been explored in Becker and Woessmann (2008). The relationship between fertility, human capital, and religion has been investigated in Becker *et al.* (2012). However, the advantage of our setting is that we can exploit fine-grained local variation, accounting for time-invariant county-specific characteristics that might have affected previous results based mainly on cross-county or cross-city estimates. Yet, it is important to bear in mind that the relationships reported and discussed below do not have a causal interpretation. In fact, earlier work by Becker and Woessmann (2009) used distance to Wittenberg, the center of Martin Luther's Reformation movement, as an instrumental variable for the spread of the Reformation throughout the regions of Prussia. Obviously, this instrument does not lend itself to exploiting *within-county* variation in Protestantism. Although distance to Wittenberg is expected to identify variation in Protestantism at a more aggregated level such as the county, this instrument has no power to explain variation within very short distances such as across localities within counties.
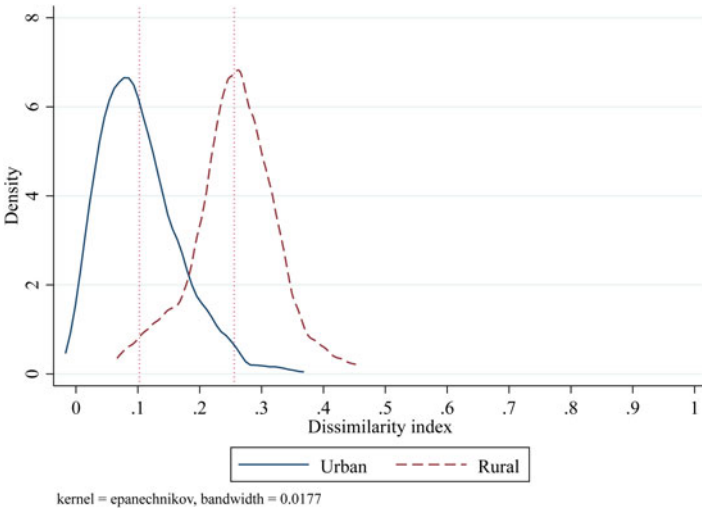
Figure 10. Distribution of dissimilarity index for literacy across counties by type of locality.
Note: The mean of the urban dissimilarity index is 0.10 whereas the rural one is 0.25. Kernel density estimates for urban (towns) and rural (villages and manors) localities. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Figure 11. The relationship between segregation of human capital and religious segregation.
Note: Kernel density estimates of dissimilarity index for literacy in localities with above-average and below-average Protestant shares. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
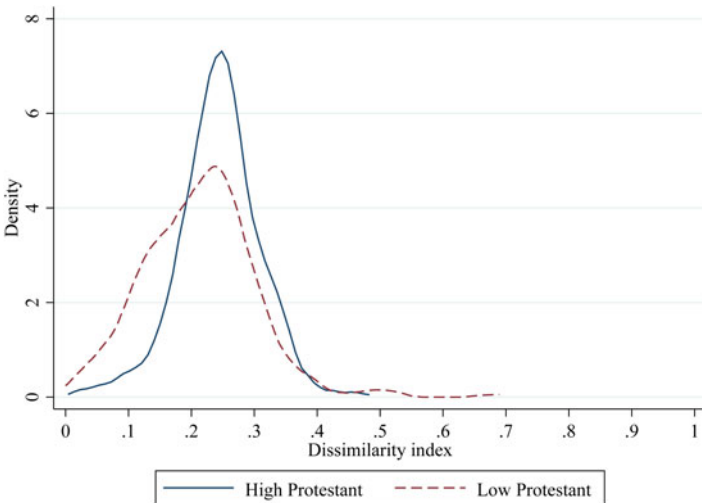
Figure 12. The relationship between human capital segregation and religious segregation.
Note: Scatter plot of dissimilarity index for literacy against dissimilarity indexes for religious groups. The dotted line shows the 45 degree line. The continuous line is the linear fit. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
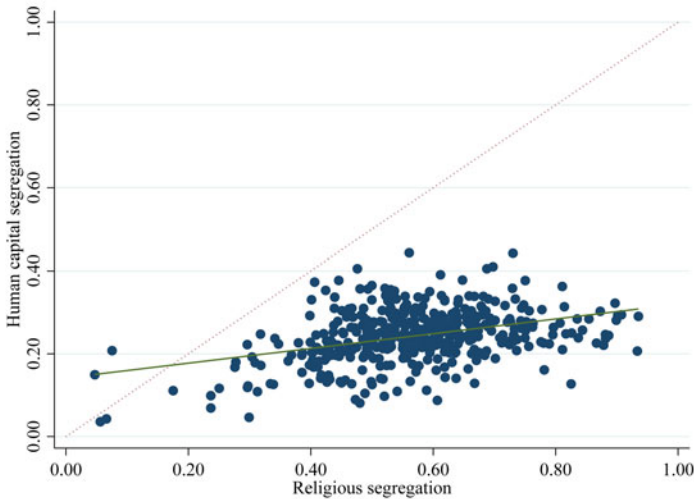
6.1 Literacy and population

In this section, we explore the possibility of a heterogeneous relationship between population size and literacy across locality type. This could be due to agglomeration economies in the provision of education.

We start by exploring graphically the relationship between literacy and log population size.[16] In Figure 13, we present scatter plots of literacy rates against log population size by locality. We unveil some interesting patterns. The relationship between literacy and log population size is increasing for towns and villages. On the contrary, we find that, for manors, the relationship between literacy and population size is negative. Therefore, the larger the manors, the lower the literacy rate. This important finding is substantiated by the regression analysis below.

In Table 4, we report the ordinary least squares (OLS) estimates. Results in column 1 indicate the existence of a positive relationship between literacy and log population size which holds even when including dummies for locality type (column 2). The coefficients for the locality dummies indicate that literacy levels in villages and manors were significantly lower than that in towns. The estimates by locality in columns 4–6 (with county-fixed effects) confirm the results shown in Figure 13, namely a positive relationship between literacy and log population size for towns and villages and a negative relationship for manors.[17] The result on the negative relationship between literacy and log population in manors is consistent with the

---

[16]We use log population size as the population distribution is highly skewed to the right. The main conclusions of the analysis hold if we use population levels instead of log population.

[17]Note that these results hold if we control for the share of people with missing information on their literacy status.
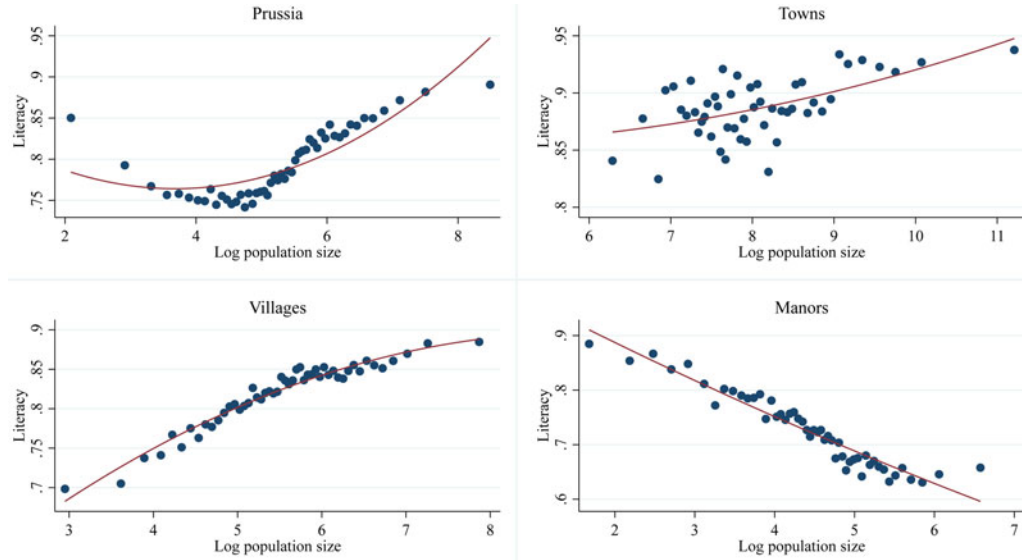
Figure 13. The relationship between literacy and log population size by locality.

Note: Binned scatterplots of the relationship between literacy and log population size with 50 equal-sized bins and quadratic fit. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Table 4. Literacy and log population size

Dep. var.: Literacy rate				By locality		
	(1)	(2)	(3)	(4)	(5)	(6)
				Town	Village	Manor
Log population size	0.025*** (0.003)	0.009** (0.004)	−0.010*** (0.001)	0.010*** (0.002)	0.004*** (0.001)	−0.027*** (0.003)
Village		−0.047*** (0.009)	−0.063*** (0.005)			
Manor		−0.127*** (0.016)	−0.113*** (0.008)			
County FE	No	No	Yes	Yes	Yes	Yes
Observations	54,270	54,270	54,248	1,180	37,783	15,187
Number of clusters	447	447	425	327	423	292
R^2	0.02	0.04	0.65	0.91	0.65	0.65
Mean literacy	0.79	0.79	0.79	0.88	0.82	0.73

Note: Standard errors clustered at the county level.
***Denotes statistical significance at the 1% level, ** at 5% level, and * at 10% level.

interpretation that a larger manorial population would be more spatially distributed and would thus have less access to churches and schools, impairing the accumulation of human capital.

It is important to bear in mind that this data source does not allow us to account for the Polish minority, which was predominantly present in eastern counties. If Poles were disproportionally living in large manors in the east, and if their level of literacy was comparatively lower, this could also explain the negative relationship between literacy and population size for manors [Kersting *et al.* (2020)]. One way to address this point is to control for the share of Catholics, because Poles were predominantly Catholic. If we include the share of Catholics in the specification of column 6 in Table 4 (not shown), the negative relationship between literacy and log population is virtually unaffected. This result should attenuate the concern that the Polish population affects the association between literacy and population size across localities.

6.2 Literacy and religion

In Table 5, we report OLS estimates of the relationship between literacy and religious denomination. The coefficients for the religious groups in column 1 confirm previous results of the literature, namely that literacy rates are larger when the share of Protestants and Jews is higher (the share of Catholics is the reference group in the regression). This result for the share of Protestants holds when accounting for locality fixed effects (column 2) and county-fixed effects (column 3).

The estimates by locality (columns 4–6) also unveil important heterogeneity. First, the Protestant "literacy premium" is much larger in manors. Second, we have previously seen that Jews lived predominantly in towns (see Table 3). The estimates in column 4 suggest that, in towns, a larger share of Jews is not associated with higher literacy.[18] There is, however, a significantly positive relationship between the share of Jews and literacy rates in rural villages and manors (column 5 and 6, respectively).[19]

Following Becker and Woessmann (2009), a large literature has studied the relationship between Protestantism and literacy.[20] Our new data allow us to investigate more in detail this relationship. In Figure 14, we plot the relationship between literacy and the share of Protestants by locality. The scatter plots reveal an interesting pattern, namely that literacy does not increase monotonically with the share of Protestants. Instead, we observe that the literacy rate first declines at very low levels of the share of Protestants, and then it starts to increase only at around 20% of Protestants. This pattern seems to be consistent across types of locality although it is more accentuated in villages and manors. A possible explanation for this U-shape pattern refers to the provision of public education. Religiously mixed

---

[18]See Abramitzky and Halaburda (this issue) on the literacy of urban and rural Jews in the context of interwar Poland.

[19]Note that when we talk about the Protestant "literacy premium," we are essentially associating the *share* of Protestants and the literacy *rate* in a location with an underlying individual-level association. This is subject to any potential caveat arising from ecological fallacy, which, however, is less pronounced in smaller locations, such as municipalities, than in larger geographic units, such as counties.

[20]See Becker *et al.* (2016) for a review of the literature on the Protestant Reformation.

Table 5. Literacy and denominations

Dep. var.: Literacy rate				By locality		
	(1)	(2)	(3)	(4)	(5)	(6)
				Town	Village	Manor
Share Protestants	0.179*** (0.024)	0.181*** (0.023)	0.141*** (0.015)	0.071*** (0.013)	0.106*** (0.013)	0.242*** (0.023)
Share Jews	0.432*** (0.107)	−0.144 (0.104)	0.356*** (0.057)	−0.012 (0.080)	0.114* (0.068)	0.394** (0.176)
Village		−0.080*** (0.008)	−0.030*** (0.004)			
Manor		−0.175*** (0.010)	−0.068*** (0.005)			
County FE	No	No	Yes	Yes	Yes	Yes
Observations	54,270	54,270	54,248	1,180	37,783	1,5187
Number of clusters	447	447	425	327	423	292
R^2	0.12	0.16	0.66	0.91	0.66	0.66
Mean literacy	0.79	0.79	0.79	0.88	0.82	0.73

Note: Standard errors clustered at the county level.
***Denotes statistical significance at the 1% level, ** at 5% level, and * at 10% level.
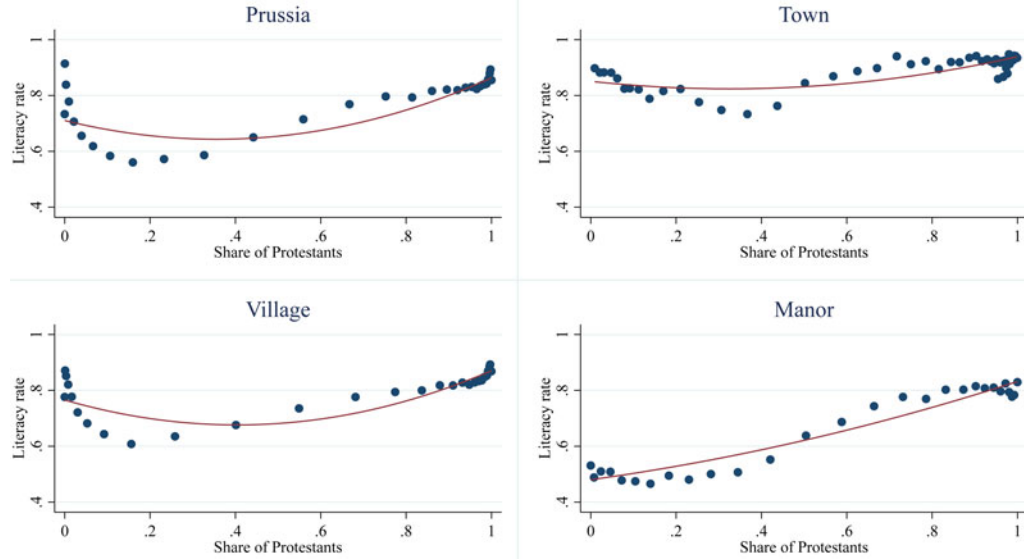
Figure 14. Literacy and Protestants by location.
Note: Binned scatterplots of the relationship between literacy rate and share of Protestants with 50 equal-sized bins and quadratic fit. *Source*: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Table 6. Fertility and denominations

Dep. var.: Child–woman ratio				By locality		
	(1)	(2)	(3)	(4)	(5)	(6)
				Town	Village	Manor
Share Protestants	−0.053*** (0.006)	−0.053*** (0.006)	−0.024*** (0.005)	−0.124 (0.094)	−0.019*** (0.005)	−0.050*** (0.014)
Share Jews	−0.036 (0.047)	0.084* (0.046)	−0.135*** (0.041)	−0.177*** (0.042)	−0.043 (0.053)	−0.097 (0.173)
Village		0.033*** (0.006)	0.025*** (0.006)			
Manor		0.031*** (0.007)	0.010 (0.007)			
County FE	No	No	Yes	Yes	Yes	Yes
Observations	54,266	54,266	54,244	1,180	37,783	15,183
Number of clusters	447	447	425	327	423	292
R^2	0.02	0.02	0.13	0.58	0.20	0.12
Mean fertility	0.50	0.50	0.50	0.47	0.50	0.49

Note: Standard errors clustered at the county level.
***Denotes statistical significance at the 1% level, ** at 5% level, and * at 10% level.

localities might have suffered from coordination problems in the provision of public education, thus generating lower levels of literacy [Cinnirella and Schueler (2016)].[21]

6.3 Fertility and religion

Finally, we analyze the relationship between fertility and the shares of religious denominations (Table 6). We proxy the fertility level by the ratio of the number of children below age 10 over the total number of women (child–woman ratio). The specification with county-fixed effects in column 3 shows that higher shares of Protestants and of Jews are both associated with lower fertility. The sample split by locality shows that the negative relationship between the share of Jews and fertility is particularly accentuated in towns (column 4). The negative coefficient for the share of Protestants is also large in towns but not statistically significant at standard levels of statistical significance.

As discussed in the literature and shown in Table 5, denominations are systematically different in terms of literacy levels. This begs the question to which extent differences in literacy can "explain" the relationship between fertility and denominations. In Table 7, we estimate, by locality, the relationship between fertility and the share of different denominations conditional on literacy. The estimates suggest that differences in literacy levels do not explain away the large negative relationship between the share of Jews and fertility in towns. Instead, controlling for literacy reduces, to some extent, the negative coefficient for the share of Protestants. These results are consistent with the notion that fertility of Jewish minorities living in urban centers might have also had a cultural explanation. Botticini *et al.* (2019), who analyze fertility patterns of Jews in central and eastern Europe from 1500 to 1930, argue that childcare practices such as breastfeeding led to a higher survival rate of newborns and that can explain the lower number of children born in the first place.

7. Conclusion

In this paper, we analyze, for the first time, demographic, religious, and educational patterns in 1871 Prussia at the level of locality (town, village, and manor) for more than 50,000 localities. This allows us not only to provide basic descriptive statistics at locality level but also to analyze more in depth whether and to what extent demographic groups were geographically segregated.

From the descriptive analysis, we find that Jews in Prussia lived predominantly in urban centers, as about 79% of them resided in towns. Prussia was already well-known for high levels of literacy in the 19th century. Our new data set shows that levels of literacy were quite high even in manors. With an average of 73% of people able to read and write, manors in Prussia had literacy levels higher than the whole of France and much higher than Italy and Spain.

---

[21]An alternative explanation would be that members of a religious minority are more fervent believers and attachment to their own identity can help to resist the influence of the majority religion [Nunziata and Rocco (2016, 2018); Bisin and Verdier (2000, 2001)]. In this context, literacy could have been an important element of Protestant identity. This could explain the relatively high levels of literacy at low shares of Protestantism, i.e., when Protestants are clearly a minority. What makes this alternative interpretation the less likely one is the fact that for very low shares of Protestants, their influence on the aggregate literacy rate in a municipality is small.

Table 7. Fertility, denominations, and human capital

Dep. var.: Child–woman ratio	By locality		
	(1)	(2)	(3)
	Town	Village	Manor
Share Protestants	−0.116 (0.101)	−0.009** (0.005)	−0.019 (0.014)
Share Jews	−0.178*** (0.042)	−0.032 (0.050)	−0.047 (0.174)
Literacy rate	−0.106 (0.125)	−0.094*** (0.009)	−0.128*** (0.016)
Constant	0.652*** (0.063)	0.581*** (0.007)	0.602*** (0.015)
County FE	Yes	Yes	Yes
Observations	1,180	37,783	15,183
Number of clusters	327	423	292
R^2	0.58	0.21	0.13
Mean fertility	0.47	0.50	0.49

Note: Standard errors clustered at the county level.
***Denotes statistical significance at the 1% level, ** at 5% level, and * at 10% level.

From the segregation analysis, we find that there was a substantial level of segregation by religious denomination within counties. That is, for a given distribution of denominations at the county level, Protestants and Catholics tended to cluster in villages and/or manors. As expected, religious segregation in towns was significantly smaller. Interestingly, we find a relatively low level of human capital segregation, i.e., towns, villages, and manors were substantially less clustered by literacy. We also provide some evidence that there is not a perfect correlation between denominational and human capital segregation, suggesting a potential mixing of more literate Protestants with less literate Catholics and *vice-versa* in the localities.

The results from the regression analysis, which accounts for county fixed effects, show that the relationship between literacy and population size differs by locality. In particular, we find that the relationship between literacy and log population is negative for manors. This result is consistent with the interpretation that a larger manorial population lived more sparsely distributed and had less access to schools and/or churches. An alternative explanation is selective migration of domestic migrants.

The regression analysis also shows that, in towns, a larger share of Jews is not associated with higher literacy. This result is consistent with the hypothesis that towns formed and possibly attracted high-literacy individuals, leveling out differences between Jews and other religious groups. In fact, in villages and manors, a larger share of Jews is associated with higher literacy rates, in line with the work of Botticini and Eckstein (2012).

Finally, we find a large negative association between fertility and the share of Jews in towns. This fertility pattern is not explained by differences in literacy, suggesting the existence of cultural explanations, possibly related to childcare practices.

The regression results presented in this paper should not be interpreted as causal. In fact, it was not the purpose of this paper to unveil any causal relationship between the

socio-economic variables at our disposal but to report "stylized facts" which, so far, had been investigated only at a more aggregated level. It is, however, our hope that the findings unveiled here will trigger more research on Prussian economic history. For example, it would be interesting to study whether localities which were predominantly Catholic had some spillover effects from neighboring Protestant localities which invested more in primary schools. Another "fact" that begs for an answer is how manors managed to achieve such high levels of literacy. Finally, although there is already substantial evidence on the trade-off between quantity and quality of children in Prussia [Becker *et al.* (2010, 2012)], it would be interesting to analyze the extent of the trade-off across localities and in conjunction with differences in the sex ratio, the latter possibly driven by gender-specific migration.

References

Abramitzky, Ran and Hanna Halaburda (this issue) Were Jews in interwar Poland more educated? *Journal of Demographic Economics.*

A'Hearn, Brian, Joerg Baten and Dorothee Baten (2009) Quantifying quantitative literacy: age heaping and the history of human capital. *The Journal of Economic History* 69(3), 783–808.

Alonso-Villar, Olga, Coral Del Rio and Carlos Gradin (2012) The extent of occupational segregation in the United States: differences by race, ethnicity, and gender. *Industrial Relations* 51(2), 179–212.

Becker, Sascha O., Francesco Cinnirella, Erik Hornung and Ludger Woessmann (2014) iPEHD – the ifo Prussian economic history database. *Historical Methods* 47, 57–66.

Becker, Sascha O., Francesco Cinnirella and Ludger Woessmann (2010) The trade-off between fertility and education: evidence from before the demographic transition. *Journal of Economic Growth* 15(3), 177–204.

Becker, Sascha O., Francesco Cinnirella and Ludger Woessmann (2012) The effect of investment in children's education on fertility in 1816 Prussia. *Cliometrica* 6, 29–44.

Becker, Sascha O. and Luigi Pascali (2019) Religion, division of labor and conflict: anti-semitism in German regions over 600 years. *American Economic Review* 109(5), 1764–1804.

Becker, Sascha O., Steven Pfaff and Jared Rubin (2016) Causes and consequences of the protestant reformation. *Explorations in Economic History* 62, 1–25.

Becker, Sascha O. and Ludger Woessmann (2008) Luther and the girls: religious denomination and the female education gap in nineteenth-century Prussia. *Scandinavian Journal of Economics* 110, 777–805.

Becker, Sascha O. and Ludger Woessmann (2009) Was weber wrong? A human capital theory of protestant economic history. *Quarterly Journal of Economics* 124(2), 531–596.

Becker, Sascha O. and Ludger Woessmann (2019) Education and socioeconomic development during the industrialization. In Claude Diebolt and Mike Haupert (eds.), *Handbook of Cliometrics*, pp. 253–273, 2nd ed. Springer Nature.

Bisin, Alberto and Thierry Verdier (2000) "Beyond the Melting Pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics* 115(3), 955–988.

Bisin, Alberto and Thierry Verdier (2001) The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory* 97(2), 298–319.

Botticini, Maristella and Zvi Eckstein (2012) *The Chosen Few: How Education Shaped Jewish History, 70–1492.* Princeton University Press.

Botticini, Maristella and Zvi Eckstein and Anat Vaturi (2019) Child care and human development: insights from Jewish history in central and Eastern Europe, 1500–1930. *The Economic Journal* 129(623), 2637–2690.

Brown, John C. and Timothy W. Guinnane (2002) Fertility transition in a rural, catholic population: Bavaria, 1880–1910. *Population Studies* 56(1), 35–49.

Cantoni, Davide (2015) 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.

Cinnirella, Francesco and Erik Hornung (2017) Land inequality, education, and marriage: empirical evidence from nineteenth-century Prussia. in Cervellati, Matteo and Sunde, Uwe (eds.), *Demographic Change and Long-Run Development*. The MIT Press

Cinnirella, Francesco, Alireza Naghavi and Giovanni Prarolo (2020) *Islam and Human Capital in Historical Spain*. CEPR, Discussion Paper 14561.

Cinnirella, Francesco and Ruth M. Schueler (2016) *The Cost of Decentralization: Linguistic Polarization and the Provision of Education*. CEPR, Discussion Paper 11274.

Cipolla, Carlo (1969) *Literacy and Development in the West*. Harmondsworth: Penguin.

Coale, Ansley Johnson and Susan Cotts Watkins (eds) (1986) *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Federico, Giovanni, Alessandro Nuvolari and Michelangelo Vasta (2019) The origins of the Italian regional divide: evidence from real wages, 1861–1913. *The Journal of Economic History* 79(1), 63–98.

Galloway, Patrick R. (2007) *Galloway Prussia Database 1861 to 1914*. www.patrickgalloway.com.

Galloway, Patrick R., Eugene A. Hammel and Ronald D. Lee (1994). Fertility decline in Prussia, 1875–1910: a pooled cross-section time series analysis. *Population Studies* 48(1), 135–158.

Galloway, Patrick R., Ronald D. Lee and Eugene A. Hammel (1998) Urban versus rural: fertility decline in the cities and rural districts of Prussia, 1875 to 1910. *European Journal of Population* 14, 209–264.

Goldstein, Joshua R. and Sebastian Klüsener (2014) Spatial analysis of the causes of fertility decline in Prussia. *Population and Development Review* 40(3), 497–525.

Hornung, Erik (2015) Railroads and growth in Prussia. *Journal of the European Economic Association* 13 (4), 699–736.

Kersting, Felix, Iris Wohnsiedler and Nikolaus Wolf (2020) Weber revisited: the protestant ethic and the spirit of nationalism. *Journal of Economic History* (forthcoming).

Klüsener, Sebastian and Joshua R. Goldstein (2016) A long-standing demographic east-west divide in Germany. *Population, Space and Place* 22(1), 5–22.

Knodel, John (1974) *The Decline of Fertility in Germany, 1871–1939*. Princeton University Press.

Knodel, John (1977) Town and country in nineteenth-century Germany: a review of urban–rural differentials in demographic behavior. *Social Science History* 1(3), 356–382.

Knodel, John (1979) From natural fertility to family limitation: the onset of fertility transition in a sample of German villages. *Demography* 16(4), 493–521.

Knodel, John and Mary Jo Maynes (1976) Urban and rural marriage patterns in imperial Germany. *Journal of Family History* 1(2), 129–161.

Lee, Ronald D., Patrick R. Galloway and Eugene A. Hammel (1994) Fertility decline in Prussia: estimating influences on supply, demand, and degree of control. *Demography* 31(2), 347–373.

Nunziata, Luca and Lorenzo Rocco (2016) A tale of minorities: evidence on religious ethics and entrepreneurship. *Journal of Economic Growth* 21(2), 189–224.

Nunziata, Luca and Lorenzo Rocco (2018) The protestant ethic and entrepreneurship: evidence from religious minorities in the former Holy Roman Empire. *European Journal of Political Economy* 51, 27–43.

Ogilvie, Sheilagh and Markus Küpker (2015) Human Capital Investment in a Late-Developing Economy: Evidence from Württemberg, c. 1600–c. 1900, Cambridge Working Papers in Economics 1528, Faculty of Economics, University of Cambridge.

Appendix A. Data sources

We digitized more than half a million entries from 11 volumes of municipality-level data from the 1871 Population Census: a dozen population figures for each of more than 50,000 locations. Specifically, we digitized male, female, and total population; the number of Protestants, Catholics, other Christians, Jews, and other religious affiliations; as well as the number of inhabitants under the age of 10, and, for those aged 10 and older, the number of literate and illiterate people, and those with missing literacy information.

We cross-checked data entry via sum checks: the total population is the sum of males and females. The sum of inhabitants across different religious denominations sums up to total population. The sum of those under age 10 and those aged 10 across different groups of literacy also add up to the total population.

The 11 volumes of municipality-level data, one for each of the eleven Prussian provinces, provided by the Prussian Statistical Office are described below. In the appendix of the various volumes, the Prussian Statistical Office offered corrections to the numbers provided in the main tables when errors were discovered after the main tables went into print. We took on board all those corrections.

We added the county identifier and county name used in the ifo Prussian Economic History Database [iPEHD] [see Becker *et al.* (2014)] to allow linkage with county-level variables from iPEHD for further analysis.

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 1: Die Gemeinden und Gutsbezirke der Provinz Preussen und ihre Bevölkerung. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 2: Die Gemeinden und Gutsbezirke der Provinz Brandenburg und ihre Bevölkerung. 1873.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 3: Die Gemeinden und Gutsbezirke der Provinz Pommern und ihre Bevölkerung. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 4: Die Gemeinden und Gutsbezirke der Provinz Posen und ihre Bevölkerung. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 5: Die Gemeinden und Gutsbezirke der Provinz Schlesien und ihre Bevölkerung. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 6: Die Gemeinden und Gutsbezirke der Provinz Sachsen und ihre Bevölkerung. 1873.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 7: Die Gemeinden und Gutsbezirke der Provinz Schleswig-Holstein und ihre Bevölkerung. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 8: Die Gemeinden und Gutsbezirke der Provinz Hannover. 1873.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 9: Die Gemeinden und Gutsbezirke der Provinz Westfalen und ihre Bevölkerung: nebst einem Anhange, betreffend die Fürstenthümer Waldeck und Pyrmont. 1874.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 10: Die Gemeinden und Gutsbezirke der Provinz Hessen-Nassau und ihre Bevölkerung. 1873.*

Königliches Statistisches Bureau. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Band 11: Die Gemeinden und Gutsbezirke der Rheinprovinz und ihre Bevölkerung: nebst einem Anhange, betreffend die Hohenzollerschen Lande. 1874.*

Appendix B. Additional figures and tables
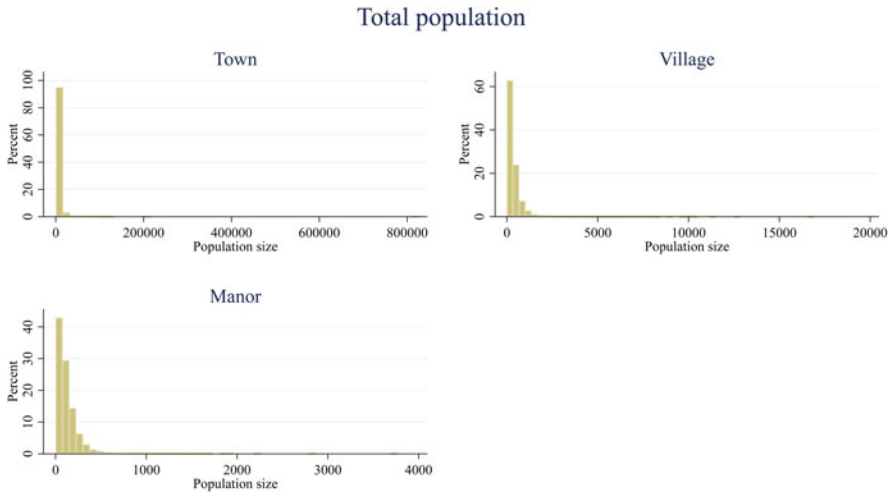
See Figures B1–B4 and Table B1.

Figure B1. Total population by locality.
Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
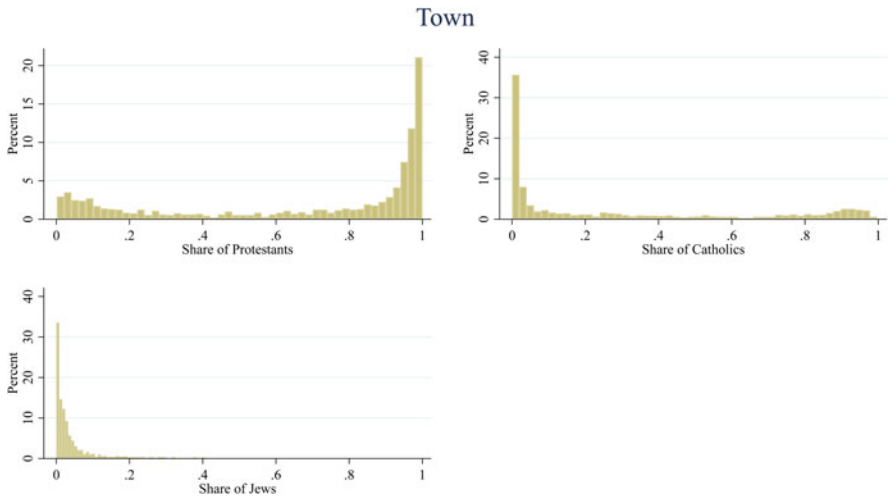

Figure B2. Shares of religious denominations.
Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.

Figure B3. Shares of religious denominations.
Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
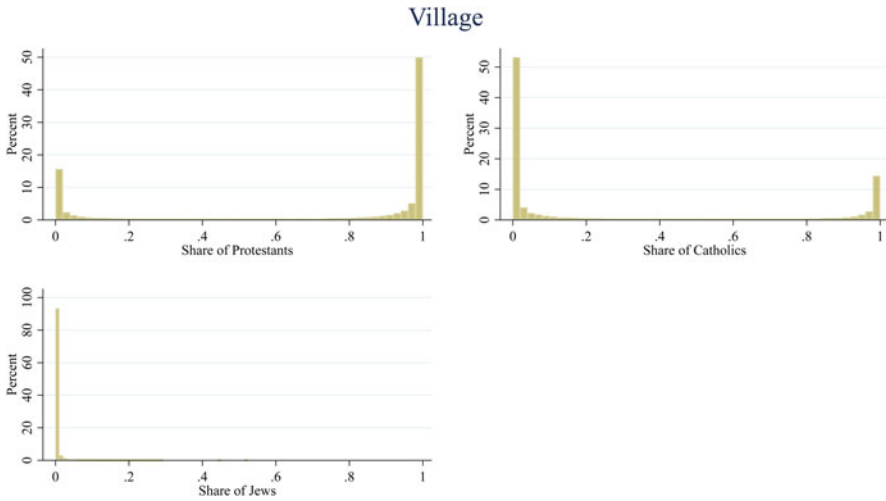

Figure B4. Shares of religious denominations.
Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.
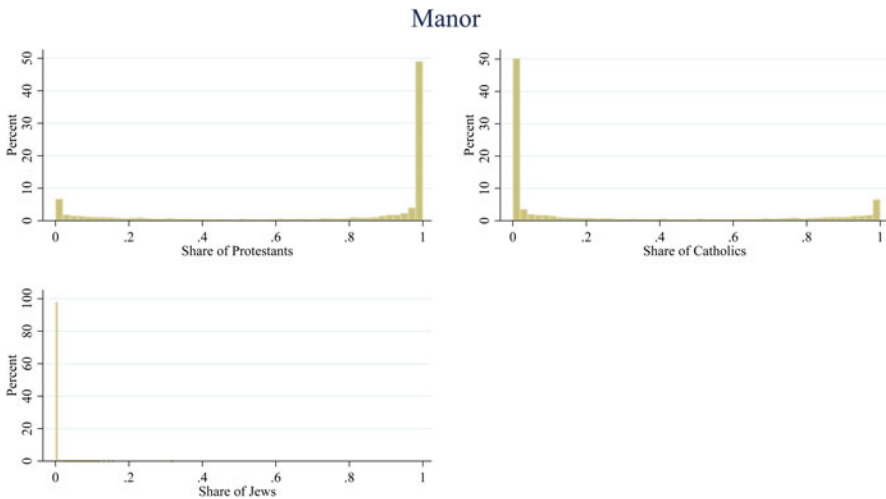
Table B1. Descriptive statistics by locality

	Obs.	Mean	Std. dev.
Town			
Population size	1,285	6,204.271	25,602.794
Literacy rate	1,285	0.888	0.108
Share males	1,285	0.486	0.031
Share females	1,285	0.514	0.031
Sex ratio	1,285	0.966	0.465
Share of stayers	1,285	0.229	0.291
Child–woman ratio	1,285	0.469	0.153
Herfindahl index	1,285	0.207	0.186
Village			
Population size	37,783	386.365	503.426
Literacy rate	37,783	0.818	0.190
Share males	37,783	0.489	0.037
Share females	37,783	0.511	0.037
Sex ratio	37,783	0.970	0.184
Share of stayers	37,783	0.268	0.301
Child–woman ratio	37,783	0.497	0.115
Herfindahl index	37,783	0.082	0.138
Manor			
Population size	15,202	125.749	138.390
Literacy rate	15,202	0.728	0.235
Share males	15,202	0.486	0.066
Share females	15,202	0.515	0.066
Sex ratio	15,198	0.991	0.645
Share of stayers	15,202	0.109	0.161
Child–woman ratio	15,198	0.495	0.186
Herfindahl index	15,202	0.121	0.167

Source: Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. Berlin: Verl. d. Königl. Statist. Bureaus, 1873–1874.