ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Becker, Sascha O.; Cinnirella, Francesco; Hornung, Erik; Woessmann, Ludger

**Working Paper**

# iPEHD: The ifo Prussian Economic History Database

CESifo Working Paper, No. 3904

**Provided in Cooperation with:**
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Becker, Sascha O.; Cinnirella, Francesco; Hornung, Erik; Woessmann, Ludger (2012) : iPEHD: The ifo Prussian Economic History Database, CESifo Working Paper, No. 3904, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/62321

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo Working Papers
www.cesifo.org/wp

# iPEHD – The ifo Prussian Economic History Database

Sascha O. Becker
Francesco Cinnirella
Erik Hornung
Ludger Woessmann

CESIFO WORKING PAPER NO. 3904
CATEGORY 5: ECONOMICS OF EDUCATION
AUGUST 2012

*An electronic version of the paper may be downloaded*
- *from the SSRN website: www.SSRN.com*
- *from the RePEc website: www.RePEc.org*
- *from the CESifo website: www.CESifo-group.org/wp*

CESifo *Working Paper No. 3904*

# iPEHD – The ifo Prussian Economic History Database

## Abstract

This paper provides a documentation of the ifo Prussian Economic History Database (iPEHD), a county-level database covering a rich collection of variables for 19th-century Prussia. The Royal Prussian Statistical Office collected these data in several censuses over the years 1816-1901, with much county-level information surviving in archives. These data provide a unique source for micro-regional empirical research in economic history, enabling analyses of the importance of such factors as education, religion, fertility, and many others for Prussian economic development in the 19th century. The service of iPEHD is to provide the data in a digitized and structured way.

JEL-Code: N130, N330.

Keywords: economic history, Prussia, 19th century, database, county.

*Sascha O. Becker*
*Department of Economics*
*University of Warwick*
*UK – Coventry, CV4 7AL*
*United Kingdom*
*s.o.becker@warwick.ac.uk*

*Francesco Cinnirella*
*Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich*
*Poschingerstrasse 5*
*Germany – 81679 Munich*
*cinnirella@ifo.de*

*Erik Hornung*
*Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich*
*Poschingerstrasse 5*
*Germany – 81679 Munich*
*hornung@ifo.de*

*Ludger Woessmann*
*Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich*
*Poschingerstrasse 5*
*Germany – 81679 Munich*
*woessmann@ifo.de*

June 21, 2012
Over the years, a large number of research assistants have contributed to the digitization work underlying iPEHD. We are grateful for their contributions, especially to Christian Steibl, as well as to Rajesh Bhateja. Financial support by the Pact for Research and Innovation of the Leibniz Association is gratefully acknowledged.

# 1. Introduction

Prussian economic history during the 19th century proves a fascinating setting to study many of the most fundamental questions in economic history. A country of such high diversity, but with a rather uniform institutional setting, allows answering many important research questions by analyzing the micro-regional variation existing within one country. For example, the Prussian setting allows analyzing the importance of such factors as education, religion, fertility, and many others for industrialization and historical economic development.

What is more, starting with the first full-scale population census in 1816, the Royal Prussian Statistical Office collected a huge amount of data in a number of censuses over the 19th century. Many interesting county-level data have survived in archives. Thanks to the Prussian proverbial orderliness and thoroughness, we have high-quality data for the Prussian counties (*Kreise*) covering nearly the whole range of the 19th century. These data provide a unique source for empirical research in economic history, with the particular potential to study historical micro-regional data with modern microeconometric methods.

The service of the ifo Prussian Economic History Database (iPEHD) is to provide many of these data in a digitized and structured way. Thus, iPEHD is a county-level database covering a rich collection of variables for all counties of Prussia over the period 1816-1901. iPEHD is freely available to any interested researcher at www.cesifo-group.de/ipehd. The iPEHD website does not only provide the raw data, but also background information, definitions, and sources of variables. It also makes suggestions on how to merge data from different census waves with varying administrative boundaries into panel datasets. Finally, it provides a collection of thematic maps visualizing the data, ready-made datasets and codes to replicate tables from published research, and additional material.

Throughout, iPEHD covers all Prussian counties, whose number increased over the 19th century from 308 in 1816 to 574 in 1901. Drawing from a total of 15 original sources – many of which contain several volumes – iPEHD comprises more than 1,500 variables. The available data cover a wide range of topics including a host of indicators of economic development, education, demographics, and more. iPEHD organizes these data into eight content areas: education, occupation, wages and income tax, industry, agriculture, population, religion, and miscellaneous. In total, iPEHD contains more than half a million data points at the county level. While nowhere near being a complete collection of all available data, we think that iPEHD provides a comprehensive micro-regional database on 19th-century economic history in Prussia.

This paper documents iPEHD and provides guidance on how to use the data contained in it. The next section starts with some brief background on how iPEHD emerged. Section 3 provides an overview of the data contained in iPEHD. Section 4 describes the data structure and suggests a procedure to combine data from different census years. Section 5 lists the original sources, published by the Royal Prussian Statistical Bureau or its employees, from which the iPEHD data stem. Section 6 gives a brief overview of research that has been conducted using iPEHD data so far. Section 7 presents a few additional features of the iPEHD website, and Section 8 concludes.

## 2. A Brief History of the Birth of iPEHD

In 2006, when looking for data to analyze the relationship of religion and literacy with economic outcomes in German history, we stumbled upon the rich county-level data available from the Prussian census of 1871. After thorough studies of the data, we were fascinated by the depth and breadth of the historical information that the Royal Prussian Statistical Office had collected and documented. Prussian thoroughness had produced high-quality data at the county level in the 19th century documenting everything from education over religion and demographics to economic development (see Figure 1 for an example).

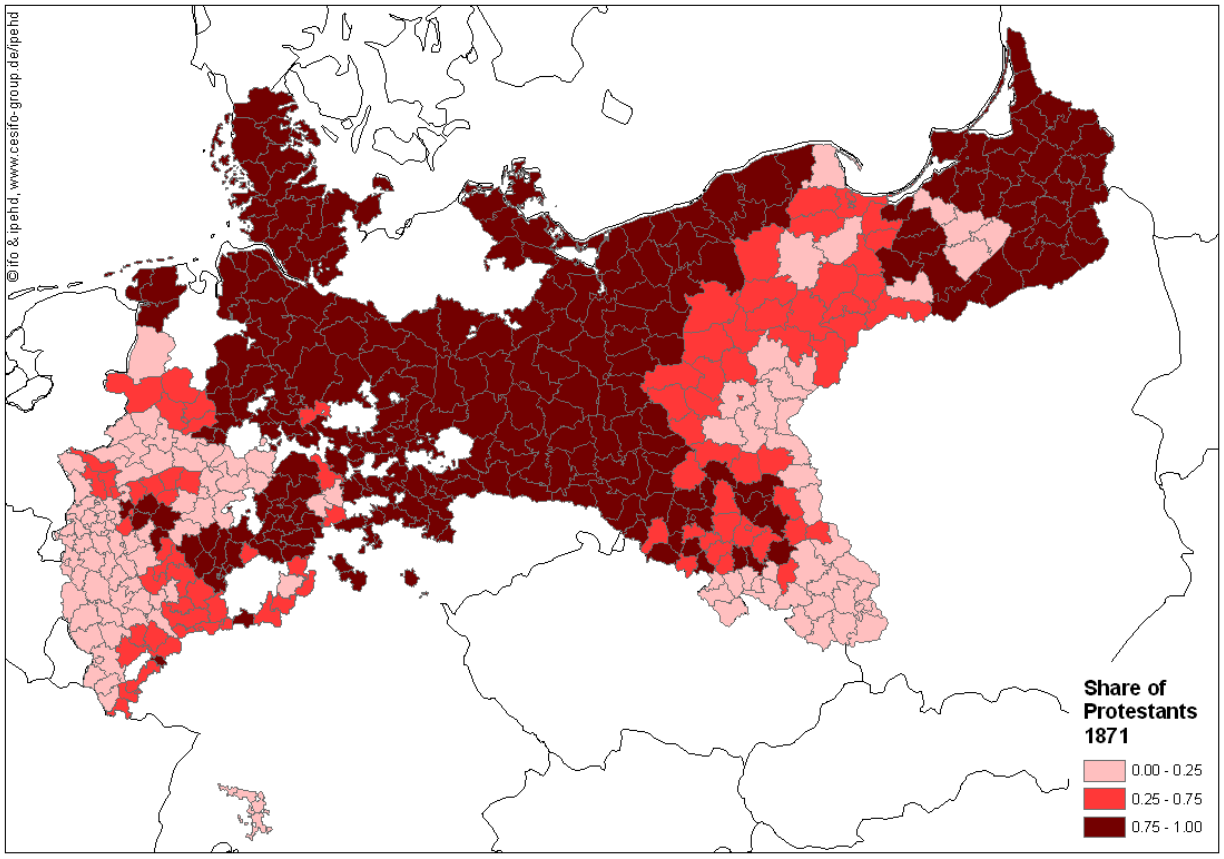
Soon, we recognized the sheer amount of data that were just sitting around in the statistical annals at German state libraries. The quality of this impressive collection of information, remarkable for the 19th century, has generally been regarded as excellent by historians and demographers (cf., e.g., Galloway, Hammel, and Lee (1994)).[1] And compared to the selective samples which a lot of historical research is restricted to, the full censuses covering the whole population provide a much more reliable picture of the historical setting.

After the original "Was Weber Wrong?" paper (eventually published as Becker and Woessmann (2009)) which relied mainly on the 1871 census and subsequent data, we explored annals covering rather unknown census data from 1816 to 1821.[2] Although lots of effort had to be undertaken to make these data ready for research and to ensure their comparability, we soon found it to be very promising and equally reliable. A third large data digitization project involved the census of 1849. The sheer amount of information provided in the sources was overwhelming.

[1] After we had digitized the data used in Becker and Woessmann (2009), data from that project became available online at www.patrickgalloway.com.

[2] We are grateful to Davide Cantoni for pointing us to these data sources at the time.

**Figure 1: Protestantism in 19th-Century Prussia**




Note: County-level depiction based on the 1871 Prussian Population Census. For details, see Becker and Woessmann (2009).

The censuses of 1816, 1849, and 1871 became the foundation of iPEHD. But, as time went by, we also digitized data from different other censuses to fill in the gaps. Although far from complete, we find the data to provide a rather comprehensive overview of 19th-century economic history in Prussia. Thus, we are happy to be able to make the digitized data available to the scientific community and the interested public. iPEHD went online in the summer of 2012 to be freely used by anyone interested at www.cesifo-group.de/ipehd.

The collection of these data and their provision to the scientific community is part of the project "Establishment of a leading international center for empirical research on the importance of education for long-term economic development," generously funded by the Leibniz Association under the Pact for Research and Innovation. The project was carried out at the Human Capital and Innovation department at the Ifo Institute – Leibniz Institute for Economic Research at the University of Munich.

## 3. An Overview of the Data contained in iPEHD

This section provides an overview of iPEHD, discussing its scope, the structure of its data files, the areas of content covered, and the information contained in the codebooks.

### *3.1 The Scope and Data Files of iPEHD*

iPEHD starts with the population census in 1816, the first full-scale census released by the Royal Prussian Statistical Office, which had been founded in 1805. The 1816 census covers the 308 Prussian counties at the time. Further extensive census data are available in 1849, 1864, 1871, and 1882, but – as indicated in Table 1 – many more detailed data were collected in additional years. As the number of counties grew over time, by 1901 the data cover 574 Prussian counties. In total, iPEHD contains more than 1,500 variables and more than half a million data points, all at the county level.

**Table 1: The Scope of iPEHD**

| Year | No. of variables | No. of county observations | No. of data points |
|---|---|---|---|
| 1816 | 58 | 308 | 17,864 |
| 1819 | 5 | 344 | 1,720 |
| 1821 | 22 | 344 | 7,568 |
| 1816-1821 | 24 | 456 | 10,944 |
| 1829 | 6 | 59 | 354 |
| 1849 | 712 | 335 | 238,520 |
| 1858 | 6 | 342 | 2,052 |
| 1862 | 4 | 346 | 1,384 |
| 1864 | 53 | 347 | 18,391 |
| 1866a | 1 | 342 | 342 |
| 1866b | 11 | 334 | 3,674 |
| 1871a | 25 | 453 | 11,325 |
| 1871b | 14 | 458 | 6,412 |
| 1878 | 5 | 426 | 2,130 |
| 1882a | 269 | 464 | 124,816 |
| 1882b | 14 | 465 | 6,510 |
| 1886a | 156 | 544 | 84,864 |
| 1886b | 97 | 518 | 50,246 |
| 1892 | 8 | 550 | 4,400 |
| 1896 | 15 | 552 | 8,280 |
| 1901 | 8 | 574 | 4,592 |
| Sum | 1,513 | | 606,388 |

Note: Some of the data points may contain missing information.

iPEHD consists of county-level information gathered from these different censuses. The data are currently presented in 76 separate data files, organized by content area, specific topic, and census year. Each data file in iPEHD contains a unique county (*Kreis*) identifier (discussed in Section 4.2 below), the county name, the abbreviated district (*Regierungsbezirk*) name (rb), and a set of variables of census data. iPEHD stores its data in comma-separated values (csv) format, which is easily accessible from any statistical software. For example, to open the csv data files in Stata, one just has to type:

```
insheet using "xxxxxx.csv"
```

To give an example of a data file, Table 2 shows a brief extract of a few variables for the first few counties (by alphabet) from the data file "ipehd_1819_indu_fac.csv", which contains data on the number of factories in a county in 1819. E.g., the variable "fac1819_brick" documents the total number of brick manufactories in a county in 1819, and the variable "mill1819_water" the total number of water mills.

**Table 2: Extract from an Example Data File**

| kreiskey1800 | County | rb | fac1819_brick | fac1819_lime | fac1819_glass | mill1819_water |
|---|---|---|---|---|---|---|
| 277 | Achen | AAC | 5 | 10 | 2 | 26 |
| 33 | Adelnau | POS | 11 | 6 | 0 | 26 |
| 254 | Adenau | KOB | 0 | 1 | 0 | 71 |
| 196 | Ahaus | MUN | 11 | 15 | 0 | 20 |
| 255 | Ahrweiler | KOB | 0 | 0 | 0 | 51 |
| 2 | Allenstein | KON | 5 | 0 | 1 | 31 |
| 219 | Altena | ARN | 3 | 13 | 0 | 41 |
| 257 | Altenkirchen | KOB | 1 | 0 | 0 | 41 |
| 10 | Angerburg | GUM | 4 | 26 | 0 | 5 |
| 53 | Angermünde | POT | 13 | 2 | 0 | 28 |
| 32 | Anklam | STE | 3 | 0 | 0 | 2 |
| 209 | Arnsberg | ARN | 12 | 4 | 0 | 26 |
| 67 | Arnswalde | FRA | 7 | 3 | 3 | 29 |
| 160 | Aschersleben | MAG | 8 | 5 | 0 | 57 |
| 55 | (Nieder-)Barnim | POT | 8 | 0 | 1 | 30 |
| 54 | (Ober-)Barnim | POT | 18 | 0 | 0 | 36 |
| 190 | Beckum | MUN | 8 | 3 | 0 | 22 |

Note: Extract from iPEHD data file "ipehd_1819_indu_fac.csv".

### *3.2 Areas of Content covered by iPEHD*

The iPEHD data are categorized into the following eight content areas:

**Education**: This area contains, among others, such data as the number of students, teachers, and schools by school type, literacy, and school finance.

**Occupation**: This area contains, among others, data on the labor force in agriculture, in factories, in manufacturing, in crafts, and in services.

**Wages and Income Tax**: This area contains data on daily wages of day laborers, on teacher income, and on income taxes.

**Industry**: This area contains data on a huge number of different factories, technologies, and transportation.

**Agriculture**: This area contains, among others, such data as livestock, crop yields, soil composition, and the distribution of land.

**Population**: This area contains data on population by age, by gender, and by marital status, on births and deaths, and on population with disabilities.

**Religion**: This area contains denomination-specific data on population, literacy, education, occupation, and number of churches.

**Miscellaneous**: This area contains such data as surface area, buildings, municipalities, and residential areas for each county.

Apart from the data gathered in these eight content areas, the **merger file** provides information on merger variables necessary to combine data from different census years (see Section 4.3 below).

### *3.3 Codebooks*

A large number of codebooks provide additional information for each variable contained in iPEHD. There is one codebook for each year, so that explanations for each variable can be found in the codebook for the corresponding year. A summary codebook that combines all years is also provided; this summary codebook allows a content search of the whole iPEHD.

The codebooks list the variable name (“variable name”), the name of the data file where it can be found (“ipehd datasets”), an English label (“label”), and the original label in German language (“original label”). The German language label is similar to the table headings found in the original sources. The English label leads with the year and is a shortened (direct) translation

of the German label; in cases where a translation is not feasible, the original German term was adopted. In addition, the codebooks indicate the source of each set of variables ("source").

## 4. Merging Data from Different Censuses

One of the biggest challenges when analyzing historical data is to ensure comparability over time. A key service of iPEHD is to facilitate the analysis of data from different census years at the county level, holding the administrative boundaries fixed. This section presents the structure in which the data are presented in iPEHD and the suggested procedure of combining different census years.

### *4.1 County-level Structure of the Data*

Starting after the Congress of Vienna in 1815, Prussia reformed its administrative structure and introduced the county level. At the time, the dimension of a county was meant to follow borders of previously existing administrative units. The maximum distance to the administrative center was meant to be two to three Prussian Miles (roughly 15 to 23 km or 9 to 14 miles), such that every inhabitant could travel forth and back within a day. The population size was meant to range between 20,000 in sparsely populated areas and 36,000 in densely populated areas.

Throughout the 19$^{th}$ century, various administrative reforms reshaped the county structure of Prussia. As the population grew over time, it became necessary to divide existing administrative units in order to reduce administrative efforts. Most of these changes were partitions of one county into two or more counties.

Thus, it is usually possible to reconstruct earlier administrative units by aggregating data from later years to the former structure. A drawback of this procedure is that the researcher loses part of the variation provided by having more observations. Still, the procedure appears necessary in order to have intertemporal comparability of the units of observation. The alternative would be to assign the same early data to two or more subsequently parted units, introducing measurement error if observed data were not uniformly distributed in the original area.

A peculiarity of the Prussian county system is the city county. Starting with the introduction of the county level in 1815, the so-called *Immediatstädte* (immediate towns) became a county themselves. As urbanization advanced, an increasing number of cities were detached from their

original county and became a county of their own. Thus, the database often contains a *Landkreis* (rural county) and a *Stadtkreis* (city county) with similar names. For example, there are six pairs of *Landkreis/Stadtkreis* information among the 335 county observations in the 1849 classification and 20 pairs among the 458 county observations in the 1874 classification.

### *4.2 County Identifiers*

All data in iPEHD reflect the administrative conditions in place at the date of publication of the census. Since censuses often ordered the counties in different ways, identifiers were assigned reflecting the order of each census. Thus, each county in each census has been assigned a continuous number which is *unique within a census but not across censuses*. The identifiers are named kreiskeyYYYY, where YYYY represents the four-digit year (see Section 4.5 below for additional peculiarities of the 1816-21 data).

The year in the identifier denotes the administrative structure of Prussia, which is not necessarily the same as the census year. In some cases, different identifiers (e.g., kreiskey1871 and kreiskey1874) even had to be assigned to data from the same census year (1871) because the Royal Prussian Statistical Office used different aggregations in different publications of data from the same census.

### *4.3 Intertemporal Comparisons*

Researchers may be interested in intertemporal comparisons and in the construction of panel datasets using iPEHD. The iPEHD county identifiers, together with the merge-county file also provided on the iPEHD website, provide a service that facilitates such linkage of comparable units of observation over time. Our suggestion is that, *in order to obtain a comparable set of observations, researchers should collapse the data to the earliest set of counties in the data*. However, it is important to point out that at the end of the day, the best way to structure and use the data will be specific to every single research project.

To conduct intertemporal comparisons, our suggestion is to take the following nine-step procedure. To construct cross-sections, the procedure should be followed only until step 3.

1. Choose datasets from the same census year.
2. Merge all datasets from the same census year using the identifier (e.g., kreiskey1882).
3. Save the cross-section.

4. Use the merge-county file provided on the iPEHD website.
5. Drop all duplicate and missing observations from the merge-county file according to the identifier in the cross-section (e.g., kreiskey1882): see section 4.4 for an example.
6. Merge the merge-county file with the cross-section using the identifier (e.g., kreiskey1882).
7. Aggregate (sum/mean) all variables in the cross-section to the aggregation level of the earliest census in the analysis using the identifier of the earliest census in the analysis (crucial step!).
8. Repeat steps 1 to 7 for datasets from other census years.
9. Merge the resulting cross-sections using the identifier of the earliest census in the analysis.

### *4.4 Example from the Merger File*

In the example shown in Table 3, the eight illustrative counties observed in 1901 were established from six counties in 1874 and five counties in 1849. Between 1849 and 1874, the 'Elbing Landkreis' had been divided into 'Elbing Stadtkreis' and 'Elbing Landkreis'. Between 1874 and 1901, the 'Danzig Landkreis' had been divided into 'Danzig Niederung', 'Danzig Höhe', and 'Dirschau'.

**Table 3: Example from the Merge File**

| Kreiskey 1901 | County1901 | Kreiskey 1874 | County1874 | Kreiskey 1849 | County1849 |
|---|---|---|---|---|---|
| 38 | ELBING STADTKREIS | 38 | ELBING STADTKREIS | 37 | ELBING LANDKREIS |
| 39 | ELBING LANDKREIS | 39 | ELBING LANDKREIS | 37 | ELBING LANDKREIS |
| 40 | MARIENBURG IN PREUSSEN | 40 | MARIENBURG IN PREUSSEN | 38 | MARIENBURG IN PREUSSEN |
| 41 | DANZIG STADTKREIS | 41 | DANZIG STADTKREIS | 39 | DANZIG STADTKREIS |
| 42 | DANZIG NIEDERUNG | 42 | DANZIG LANDKREIS | 40 | DANZIG LANDKREIS |
| 43 | DANZIG HOHE | 42 | DANZIG LANDKREIS | 40 | DANZIG LANDKREIS |
| 44 | DIRSCHAU | 42 | DANZIG LANDKREIS | 40 | DANZIG LANDKREIS |
| 45 | PREUSSISCH STARGARD | 43 | PREUSSISCH STARGARD | 41 | PREUSSISCH STARGARD |

Note: Extract from the iPEHD merge file "ipehd_merge_county.csv".

In order to have a comparable set of observations when performing intertemporal comparisons between 1901 and 1849, one has to aggregate the observations of 'Danzig Niederung', 'Danzig Höhe', and 'Dirschau' to match 'Danzig Landkreis'. Thus, one should always aggregate the data to the aggregation level of the earliest census year in the specific analysis (step 7).

However, to perform intertemporal comparisons between, e.g., 1874 and 1849, one needs to drop the duplicate entries of 'Danzig Landkreis' from the merger file first (step 5). In addition, one needs to drop entries from the merger file that have missing observations on the county identifier in the respective year. Such missing observations exist because some territories were annexed by Prussia only after the respective census year.

As one example of how to merge datasets from 1874 and 1849, the following Stata code exemplifies the nine steps of the suggested procedure:

```
insheet using "ipehd_1871_edu_literacy_part2.csv", clear   /* Step 1 */
save "ipehd_1871_edu_literacy_part2.dta"

insheet using "ipehd_1871_pop_demo_part2.csv", clear
save "ipehd_1871_pop_demo_part2.dta"

merge 1:1 kreiskey1874 using "ipehd_1871_edu_literacy_part2.dta"  /* Step 2 */
drop _merge

save "ipehd_1871_part2.dta"   /* Step 3 */

insheet using "ipehd_merge_county.csv", clear   /* Step 4 */

duplicates drop kreiskey1874, force   /* Step 5 */
drop if kreiskey1874==.

merge 1:1 kreiskey1874 using "ipehd_1871_part2.dta"   /* Step 6 */

collapse (sum) pop* lit* edu*, by (kreiskey1849)   /* Step 7 */
drop if kreiskey1849==.
save "ipehd_1871_part2_2.dta"

insheet using "ipehd_1849_rel_deno.csv", clear   /* Step 8 */
save "ipehd_1849_rel_deno.dta"

insheet using "ipehd_merge_county.csv", clear
duplicates drop kreiskey1849, force
drop if kreiskey1849==.
merge 1:1 kreiskey1849 using "ipehd_1849_rel_deno.dta"
collapse (sum) rel*, by (kreiskey1849)
save "ipehd_1849.dta"

merge 1:1 kreiskey1849 using "ipehd_1871_part2_2.dta"   /* Step 9 */
drop _merge
save "ipehd_1849_1871.dta"
```

### *4.5 Peculiarity of the Data from 1816 to 1821*

By 1816, Prussia had just started her administrative reform that established the county level. In some parts of the country, the reforms had not been finalized even in 1821. Thus, the data from the censuses in 1816 until after 1821 sometimes reflect old administrative units.

Unfortunately, due to the reform, these old units were subsequently aggregated and then newly divided in order to establish the new counties. This makes it impossible to accurately match the data of (some of) the administrative units from the early censuses to (some) counties in subsequent censuses. The *kreiskey1800* is thus coded so as to aggregate the data to a higher level. The *kreiskey1800* can be used to link the 1816-1821 data to later periods.

However, iPEHD also provides a unique identifier that allows merging data from the same census for these cross-sections. These identifiers are named 'id1816' and 'id1819'. In order to merge data from 1816 to other data from 1816, one should use id1816. In order to merge data from 1819 or 1821 to other data from 1819 or 1821, one should use id1819.

In order to merge data from 1816, 1819, or 1821 to data from subsequent censuses, one should take the following steps:

1. Choose datasets from 1816, 1819, or 1821.
2. Merge all datasets from the same census using the identifier (idYYYY).
3. Aggregate (sum/mean) all cross-sections using the identifier 'kreiskey1800'.
4. Merge the cross-section with aggregated data from subsequent censuses using the identifier 'kreiskey1800'.

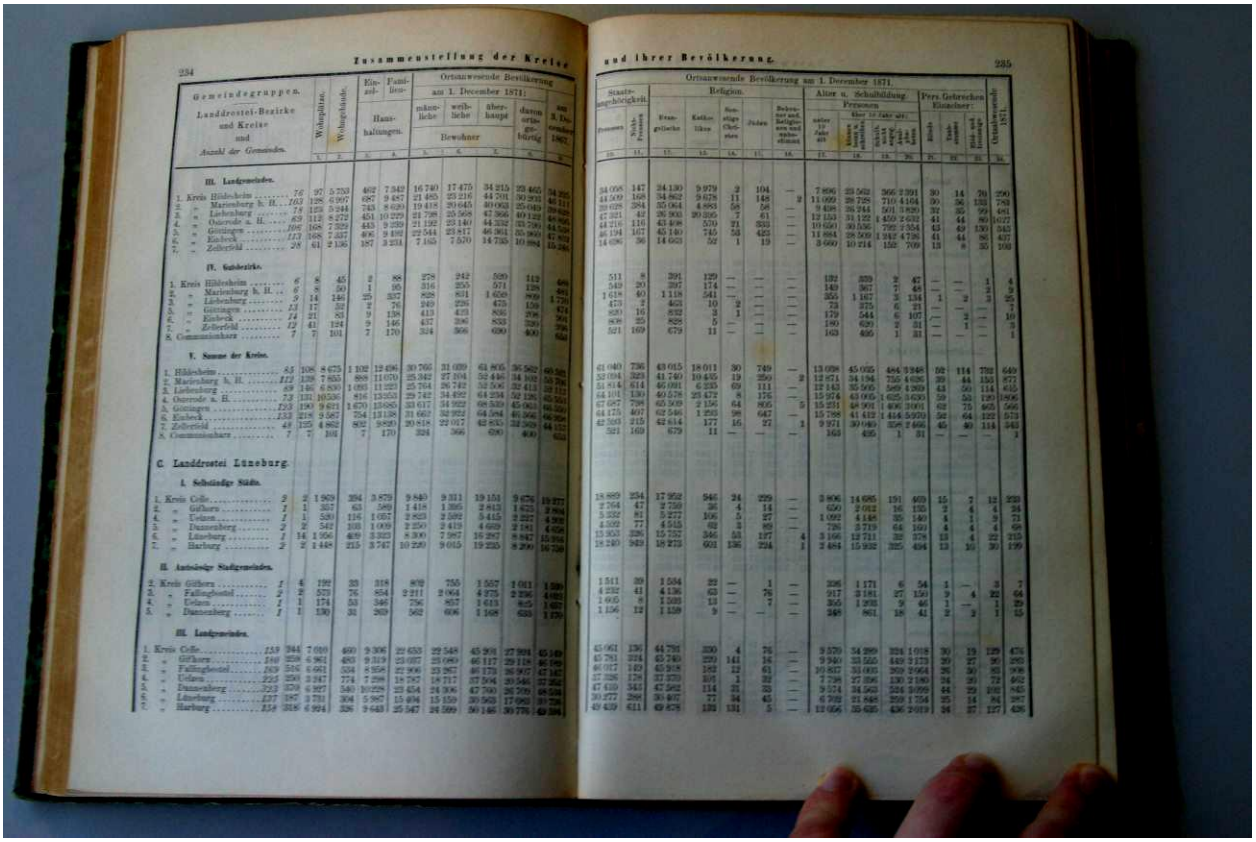
## 5. Original Sources underlying the iPEHD Data

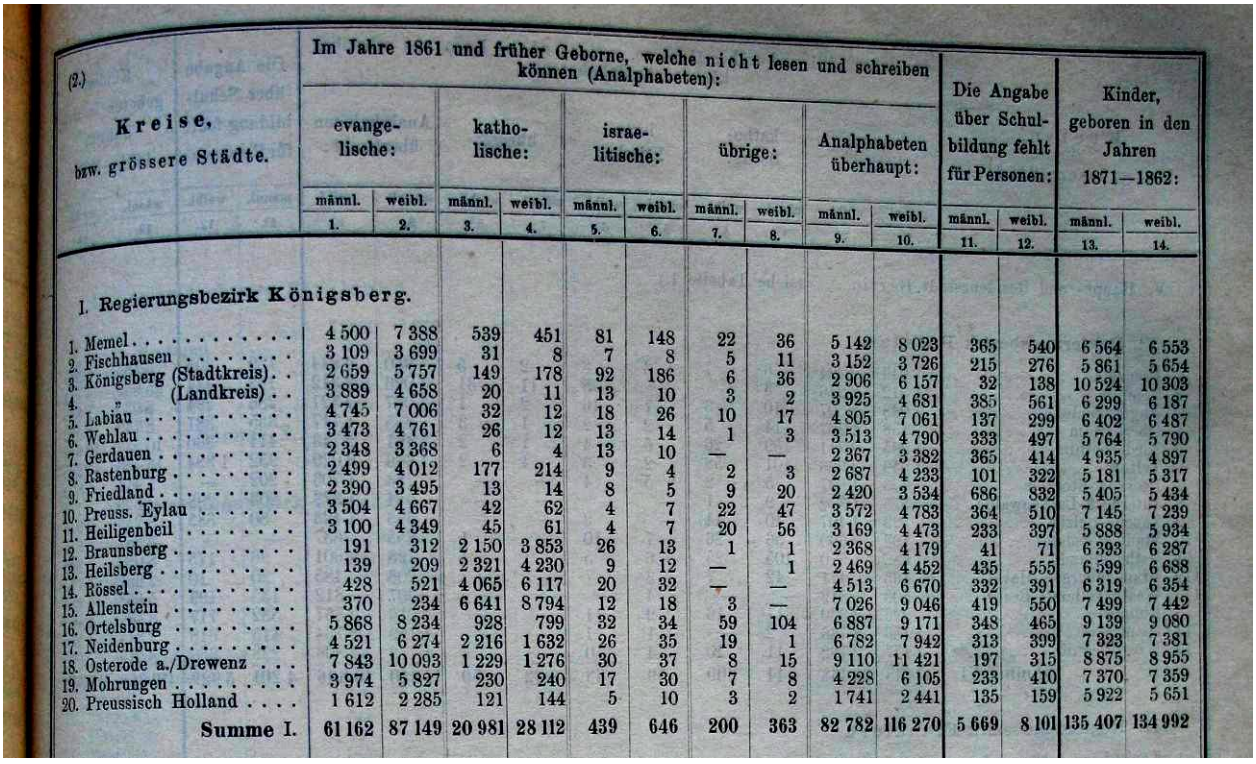
The iPEHD data have been digitized from different sources originally published by the Royal Prussian Statistical Bureau or its employees. These original historical volumes should be consulted for detailed information regarding the exact attributes of the data. Figure 2 shows two example pages from such source volumes.

The following list documents all the volumes used as sources for iPEHD. There are a total of 15 original sources, many of which consist of several volumes.

**1816-21**: Mützell, Alexander A. (1821-25). *Neues Topographisch-statistisch-geographisches Wörterbuch des Preussischen Staats*, Vol. 1-6. Halle: Karl August Kümmel.

**Figure 2: Example Pages from Source Volumes**





(2.)

Im Jahre 1861 und früher Geborne, welche nicht lesen und schreiben können (Analphabeten):

| Kreise, bzw. grössere Städte. | evangelische: | | katholische: | | israelitische: | | übrige: | | Analphabeten überhaupt: | | Die Angabe über Schulbildung fehlt für Personen: | | Kinder, geboren in den Jahren 1871—1862: | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | männl. | weibl. | männl. | weibl. | männl. | weibl. | männl. | weibl. | männl. | weibl. | männl. | weibl. | männl. | weibl. |
| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. |
| **1. Regierungsbezirk Königsberg.** | | | | | | | | | | | | | | |
| 1. Memel | 4 500 | 7 388 | 539 | 451 | 81 | 148 | 22 | 36 | 5 142 | 8 023 | 365 | 540 | 6 564 | 6 553 |
| 2. Fischhausen | 3 109 | 3 699 | 31 | 8 | 7 | 8 | 5 | 11 | 3 152 | 3 726 | 215 | 276 | 5 861 | 5 654 |
| 3. Königsberg (Stadtkreis) | 2 659 | 5 757 | 149 | 178 | 92 | 186 | 6 | 36 | 2 906 | 6 157 | 32 | 138 | 10 524 | 10 303 |
| 4. " (Landkreis) | 3 889 | 4 658 | 20 | 11 | 13 | 10 | 3 | 2 | 3 925 | 4 681 | 385 | 561 | 6 299 | 6 187 |
| 5. Labiau | 4 745 | 7 006 | 32 | 12 | 18 | 26 | 10 | 17 | 4 805 | 7 061 | 137 | 299 | 6 402 | 6 487 |
| 6. Wehlau | 3 473 | 4 761 | 26 | 12 | 13 | 14 | 1 | 3 | 3 513 | 4 790 | 333 | 497 | 5 764 | 5 790 |
| 7. Gerdauen | 2 348 | 3 368 | 6 | 4 | 13 | 10 | — | — | 2 367 | 3 382 | 365 | 414 | 4 935 | 4 897 |
| 8. Rastenburg | 2 499 | 4 012 | 177 | 214 | 9 | 4 | 2 | 3 | 2 687 | 4 233 | 101 | 322 | 5 181 | 5 317 |
| 9. Friedland | 2 390 | 3 495 | 13 | 14 | 8 | 5 | 9 | 20 | 2 420 | 3 534 | 686 | 832 | 5 405 | 5 434 |
| 10. Preuss. Eylau | 3 504 | 4 667 | 42 | 62 | 4 | 7 | 22 | 47 | 3 572 | 4 783 | 364 | 510 | 7 145 | 7 239 |
| 11. Heiligenbeil | 3 100 | 4 349 | 45 | 61 | 4 | 7 | 20 | 56 | 3 169 | 4 473 | 233 | 397 | 5 888 | 5 934 |
| 12. Braunsberg | 191 | 312 | 2 150 | 3 853 | 26 | 13 | 1 | 1 | 2 368 | 4 179 | 41 | 71 | 6 393 | 6 287 |
| 13. Heilsberg | 139 | 209 | 2 321 | 4 230 | 9 | 12 | — | 1 | 2 469 | 4 452 | 435 | 555 | 6 599 | 6 688 |
| 14. Rössel | 428 | 521 | 4 065 | 6 117 | 20 | 32 | — | — | 4 513 | 6 670 | 332 | 391 | 6 319 | 6 354 |
| 15. Allenstein | 370 | 234 | 6 641 | 8 794 | 12 | 18 | 3 | — | 7 026 | 9 046 | 419 | 550 | 7 499 | 7 442 |
| 16. Ortelsburg | 5 868 | 8 234 | 928 | 799 | 32 | 34 | 59 | 104 | 6 887 | 9 171 | 348 | 465 | 9 139 | 9 080 |
| 17. Neidenburg | 4 521 | 6 274 | 2 216 | 1 632 | 26 | 35 | 19 | 1 | 6 782 | 7 942 | 313 | 399 | 7 323 | 7 381 |
| 18. Osterode a./Drewenz | 7 843 | 10 093 | 1 229 | 1 276 | 30 | 37 | 8 | 15 | 9 110 | 11 421 | 197 | 315 | 8 875 | 8 955 |
| 19. Mohrungen | 3 974 | 5 827 | 230 | 240 | 17 | 30 | 7 | 8 | 4 228 | 6 105 | 233 | 410 | 7 370 | 7 359 |
| 20. Preussisch Holland | 1 612 | 2 285 | 121 | 144 | 5 | 10 | 3 | 2 | 1 741 | 2 441 | 135 | 159 | 5 922 | 5 651 |
| **Summe I.** | **61 162** | **87 149** | **20 981** | **28 112** | **439** | **646** | **200** | **363** | **82 782** | **116 270** | **5 669** | **8 101** | **135 407** | **134 992** |

Note: The top picture is from Königliches Statistisches Bureau (1873), Vol. VIII, pp. 234-235. The bottom picture is from Königliches Statistisches Bureau in Berlin (1875), p. 117.

**1829**: Preussisches Statistisches Landesamt (1829). *Beiträge zur Statistik der Königlichen Preussischen Rheinlande, aus amtlichen Nachrichten zusammengestellt*. Aachen: J.A. Mayer.

**1849**: Statistisches Bureau zu Berlin (1851-55). *Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849*, Vol. 1-6b. Berlin: Statistisches Bureau zu Berlin.

**1858**: Meitzen, August (1868). *Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates*, Vol. 1-4. Berlin: Verlag von Paul Parey.

**1862**: Königlich Preussisches Statistisches Bureau (1863). Die Eisen-, Stein- und Wasserstrassen des preussischen Staates im Jahre 1862, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 3, 206-214. Berlin: Verlag der Königlichen Geheimen Ober-Hofbuchdruckerei.

**1864**: Königliches Statistisches Bureau in Berlin (1867). *Die Ergebnisse der Volkszählung und Volksbeschreibung, der Gebäude und Viehzählung, nach den Aufnahmen vom 3. December 1864, resp. Anfang 1865 und die Statistik der Bewegung der Bevölkerung in den Jahren 1862, 1863 und 1864*. Preussische Statistik Vol. 10. Berlin: Verlag von Ernst Kuehn.

**1866**: Meitzen, August (1868). *Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates*, Vol. 1-4. Berlin: Verlag von Paul Parey.

**1871**: Königliches Statistisches Bureau (1873-74). *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1.December 1871*, Vol. 1-11. Berlin: Verlag des Königlichen Statistischen Bureaus.

Königliches Statistisches Bureau in Berlin (1875). *Die Ergebnisse der Volkszählung und Volksbeschreibung im Preussischen Staate vom 1. December 1871*. Preussische Statistik Vol. 30. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1878**: Herrfurth, Ludwig and Conrad Studt (1880). *Finanzstatistik der Kreise des preussischen Staates für das Jahr 1877/78.* Zeitschrift des Preussischen Statistischen Landesamtes, Ergänzungshefte, Vol. 7. Berlin: Verlag des Königlichen Statistischen Bureaus

**1882**: Königliches Statistisches Bureau in Berlin (1884/85). *Die Ergebnisse der Berufsstatistik vom 5. Juni 1882 im preussischen Staat*. Preussische Statistik Vol. 76 a-c. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1886 (Education)**: Königliches Statistisches Bureau in Berlin (1889). *Das gesammte Volksschulwesen im preußischen Staate im Jahre 1886.* Preussische Statistik Vol. 101. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1886 (Agriculture)**: Königliches Statistisches Bureau in Berlin (1887). *Die Ergebnisse der Ermittelung des Ernteertrags im preussischen Staate für das Jahr 1886*. Preussische Statistik Vol. 92. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1892**: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 44, 310-346. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1896**: Königliches Statistisches Bureau in Berlin (1897). *Die Ergebnisse der Ermittelung des Ernteertrags im preussischen Staate für das Jahr 1896*. Preussische Statistik Vol. 147. Berlin: Verlag des Königlichen Statistischen Bureaus.

**1901**: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 44, 310-346. Berlin: Verlag des Königlichen Statistischen Bureaus.

## 6. Existing Research using the iPEHD Data

A lot of research in economic history has used data from the iPEHD by now. This research is briefly described in this section. For those papers already published in academic journals, the iPEHD website provides ready-made datasets and codes in Stata to replicate the tables published in the papers. In addition, many more projects are currently under way and will be added to the website as publications become available. There is also a non-technical survey that summarizes some of the research conducted using the iPEHD data: Becker and Woessmann (2011a), "The Effects of the Protestant Reformation on Human Capital."

### *6.1 Protestant Economic History and Education*

Becker and Woessmann (2009), "Was Weber Wrong? A Human Capital Theory of Protestant Economic History" (started in 2006, first working-paper version released in 2007): This paper uses data from several censuses (Population 1871, Occupation 1882, Education 1886) and additional sources (including the Income Tax Statistics 1877) to show that the higher economic prosperity of Protestant relative to Catholic counties can be accounted for by

Protestants' higher literacy (presumably spurred by instruction in reading the Bible), suggesting that explanations based purely on differential work ethics may have limited explanatory power.

Becker and Woessmann (2008), "Luther and the Girls: Religious Denomination and the Female Education Gap in 19th Century Prussia": Using data from the first Prussian census in 1816, among others, this paper shows that a larger share of Protestants in a county's population decreased the gender gap in basic education.

Becker and Woessmann (2010), "The Effect of Protestantism on Education before the Industrialization: Evidence from 1816 Prussia" (first working-paper version released in 2009): This paper shows that Protestantism led to more schooling already in 1816, before the Industrial Revolution, ruling out that Protestant education just resulted from industrialization.

Becker and Woessmann (2011b), "Knocking on Heaven's Door? Protestantism and Suicide": Using data from 1816-21 and 1869-71, this paper finds a substantial positive effect of Protestantism on suicide.

### *6.2 Education and the Industrial Revolution*

Becker, Hornung, and Woessmann (2011), "Education and Catch-up in the Industrial Revolution" (first working-paper version released in 2009): This paper combines school-enrollment and factory-employment data from 1816, 1849, and 1882 to show that – in contrast to the state-of-the-art view based on British evidence – basic education was significantly associated with non-textile industrialization in both phases of the Industrial Revolution.

Cinnirella and Hornung (2011), "Landownership Concentration and the Expansion of Education": Combining data from several censuses that effectively span the entire 19th century (1816, 1849, 1864, 1886, and 1896), as well as data from a 1866 classification of soil composition, this paper finds that landownership concentration, a proxy for the institution of serf labor, had a negative effect on school enrollment which diminished in the second half of the century.

### *6.3 Fertility and Education*

Becker, Cinnirella, and Woessmann (2010), "The Trade-off between Fertility and Education: Evidence from before the Demographic Transition" (first working-paper version released in 2009): This paper uses data from the 1849 census and other sources to show that a

trade-off between child quantity and quality existed already in the 19th century and that causation between fertility and education runs both ways.

Becker, Cinnirella, and Woessmann (2012b), "The Effect of Investment in Children's Education on Fertility in 1816 Prussia" (first working-paper version released in 2010): Using data from the 1816 census, this paper finds a significant negative causal effect of education on fertility – evidence for a child quantity-quality trade-off – already several decades before the demographic transition and shows that it is robust to accounting for spatial autocorrelation.

Becker, Cinnirella, and Woessmann (2012a), "Does Parental Education Affect Fertility? Evidence from Pre-Demographic Transition Prussia" (first working-paper version released in 2011): Combining data from three censuses – 1816, 1849, and 1867 – this paper finds a negative residual effect of women's education on fertility, despite controlling for several demand and supply factors.

## 7. Additional Features of the iPEHD Website

The iPEHD website contains a number of additional features. For example, it provides a collection of thematic maps, produced using ArcGIS, that visualize the geographical distribution of several interesting variables across the Prussian territory. One such example is shown in Figure 1 above.

Furthermore, iPEHD is certainly not the only project dealing with historical Prussian data at the county level. Other projects provide such services as maps, information on territorial changes, additional data, and other material on Prussian counties. Links to websites of several of these projects, whose work is highly appreciated and can be viewed as complementary to iPEHD, are provided on the iPEHD website.

Finally, the iPEHD website contains a section on frequently asked questions, providing answers to standard problems encountered by iPEHD users.

## 8. Conclusions

The data contained in iPEHD, originally collected by the Royal Prussian Statistical Office, is an impressive collection of information whose quality, already in the 19th century, is generally regarded as excellent. Now digitized from censuses located in archives, these county-level data provide information on education, occupation, income and tax measures, industry, agriculture,

demographics, religion, and more. This database should facilitate future quantitative research on the economic history of 19th-century Prussia.

However, while iPEHD provides the service of supplying the historical data in a digitized and structured way and suggests ways on how to merge the data from different sources, researchers need to think carefully how to use the data in the context of their specific research projects. For instance, building panel datasets from the different census waves with varying administrative boundaries is a complex task that requires particular thought, meticulous care, and acquaintance with the structure of the original data. More generally, anybody planning to use the raw data contained in iPEHD should make sure to be well acquainted with the data structure and specifics as described in this documentation.

We hope that iPEHD provides a major service to researchers interested in Prussian economic history. Anybody who uses data from iPEHD is kindly requested to cite this paper as a source. Please also send one electronic copy of any work that uses data from iPEHD to iPEHD@ifo.de.

## References


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, no. 3: 177-204.

Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2012a. "Does parental education affect fertility? Evidence from pre-demographic transition Prussia." *European Review of Economic History*: forthcoming.

Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2012b. "The effect of investment in children's education on fertility in 1816 Prussia." *Cliometrica* 6, no. 1: 29-44.

Becker, Sascha O., Erik Hornung, and Ludger Woessmann. 2011. "Education and catch-up in the Industrial Revolution." *American Economic Journal: Macroeconomics* 3, no. 3: 92-126.

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, no. 4: 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, no. 2: 531-596.

Becker, Sascha O., and Ludger Woessmann. 2010. "The effect of Protestantism on education before the industrialization: Evidence from 1816 Prussia." *Economics Letters* 107, no. 2: 224-228.

Becker, Sascha O., and Ludger Woessmann. 2011a. "The effects of the Protestant Reformation on human capital." In *The Oxford Handbook of the Economics of Religion*, edited by Rachel M. McCleary. Oxford: Oxford University Press: 93-110.

Becker, Sascha O., and Ludger Woessmann. 2011b. "Knocking on heaven's door? Protestantism and suicide." CESifo Working Paper 3499. Munich: CESifo.

Cinnirella, Francesco, and Erik Hornung. 2011. "Landownership concentration and the expansion of education." CESifo Working Paper 3603. Munich: CESifo.

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, no. 1: 135-158.