THE UNIVERSITY OF WARWICK

March 2012 No.80

Railroads and Micro-regional Growth in Prussia

Erik Hornung
Ifo Institute

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Railroads and Micro-regional Growth in Prussia *

Erik Hornung
Ifo Institute †

March 1, 2012

Abstract

We study the effect of railroad access on urban population growth. Using GIS techniques, we match triennial population data for roughly 1000 cities in nineteenth-century Prussia to georeferenced maps of the German railroad network. We find positive short- and long-term effects of having a station on urban growth for different periods during 1840-1871. Causal effects of (potentially endogenous) railroad access on city growth are identified using instrumental-variable and fixed-effects estimation techniques. Our instrument identifies exogenous variation in railroad access by constructing straight-line corridors between terminal stations. Counterfactual models using pre-railroad growth yield no evidence in support of the hypothesis that railroads appeared as a consequence of a previous growth spurt.

Keywords: Railroads, Technological Diffusion, City Growth, Prussian Economic History
JEL classification: O18, O33, N73

*Comments from Oliver Falck, Guido Schwerdt, Ludger Woessmann and seminar participants at Ifo Munich, are gratefully acknowledged. Erik Hornung gratefully acknowledges the hospitality provided by CAGE during a research visit at the University of Warwick. This work was financially supported by the Pact for Research and Innovation of the Leibniz Association. Laurenz Detsch provided capable research assistance.

†Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany; hornung@ifo.de

1 Introduction

The statement that technological change is one of the driving forces of economic growth is beyond dispute. Railroads, as one of the most important innovations of the nineteenth century, have been repeatedly discussed to be *the* technology that shaped growth during the Industrial Revolution (Fishlow, 1965; Fogel, 1962; Rostow, 1962). The effect of the railroad on aggregate growth has been comparatively calculated using the concept of social savings for many countries that were early adopters of railroad technology and ranges from 4 to 25 percent of GNP, depending on the country and the period under consideration (see O'Brien, 1983). For Germany, the relationship between railroads and economic growth has been primarily analyzed by calculating the investment induced by railroad construction and the backward linkages to other industries (Fremdling, 1977, 1985). Pierenkemper and Tilly (2004, p. 63) for example note that the demand for iron and coal induced by railroad construction was the engine of the Industrial Revolution in Germany.

In addition to its macroeconomic effect, technology adoption can be crucial in generating localized comparative advantages and regional economic growth. Approaches analyzing variations of railroad diffusion within one country, have the advantage to exclude differences in the institutional and cultural frameworks that might be driving the system but that are often difficult to observe in cross-country studies. Another reoccurring theme in the literature is the causal direction of the relationship. Recently, Atack et al. (2010), Banerjee et al. (2009), and Donaldson (2010) attempted to answer the problem raised by Fishlow (1965): Did railroads have a substantial impact on economic growth or did they appear as a consequence of growth? There is some consensus in the literature that the latter is most likely (Hahn, 2005, p. 26; Fremdling, 1983, p. 122). However, the question whether regions grew comparatively faster after they gained access to the railroad has not been answered conclusively (Matzerath, 1996, p. XI).

This paper analyzes the micro-regional effect of railroad access on economic growth at the city level within the German state of Prussia. Using an extensive dataset of all 978 Prussian cities, we provide evidence that access to this new technology massively influenced city growth rates – a widely used proxy for regional economic growth. Following the notion that "city sizes grow with improvements in technology" (Henderson, 2005, p. 1577), we estimate that cities that adopted railroad technology subsequently increased their growth by an annual rate of roughly 1 to 2 percentage points, compared to the non-adopting cities. The size of this effect remains relatively stable across a range of different periods and specifications.

Using a geographic information system (GIS), we geo-reference historical maps of the German railroad system as well as the location of all Prussian cities to obtain information on railroad diffusion over time. This allows us to test the relationship cross-sectionally, as well as in a panel setting. The period under consideration covers the beginning of railroad construction in Prussia in 1838 until the main railroad framework was laid out during the 1860s.

To address issues of endogeneity, we employ an instrumental-variable (IV) approach and estimate the causal effect of railroad access on growth. This IV approach rests on the fact that until the 1860s, railroads were built exclusively to connect important cities. Since construction costs were high, lines were mostly built in a linear way. Consequently, cities located on a direct line between these important cities were able to gain access to the railroad by chance, whereas cities whose location deviated from the line could gain access only for reasons potentially endogenous to the city's growth. By using a straight line to connect terminal and junction stations, we can construct a variable indicating the potential for railroad adoption – being located within a straight-line corridor – that we use to instrument actual railroad access.

This instrument proves to be powerful in cross-sectional as well as in panel estimations with fixed effects. The instrument exhibits a time dimension because new straight-line corridors were established whenever new railroads were built. Both approaches return significant positive effects of railroad access over a range of different periods. In additional robustness tests, we apply instrumental-variable estimations to different matched samples consisting only of cities similar in terms of geography or in terms of a large set of other matching characteristics.

Above all, the coming of the railroad was a transport revolution. Landes (1969, p. 154) notes that faster transport meant that labor became more mobile and natural obstacles to the movement of the factors of production were eliminated. As cities became connected to the growing rail network, new and existing businesses were able to produce at lower costs, realizing scale economies. Consequently, new jobs were created or existing ones relocated. Therefore, cities connected to the railroad attracted an increasing inflow of migrants seeking employment opportunities. This led to two kinds of city growth: natural population growth caused by increases in fertility and rural-urban migration.

We can actually test the hypothesis that railroad access created additional jobs in the industry in our setting. In an alternative specification, we find that the average firm-size in cities that are connected to the railroad network is larger than in unconnected cities. However, we do not find that railroad access affects the number of factories located in a city. Thus railroads seem to affect industrial growth at the intensive margin rather than at the extensive margin.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 provides the historical background of railroad network expansion and urbanization patterns in Prussia. Section 4 provides insight into the mechanism that drives this relationship. Section 5 introduces and describes the data. Section 6 addresses endogeneity issues and presents results of the OLS, IV, and fixed-effects estimations. Section 7 concludes.

2 Related Literature

An expanding body of literature examines the effects of the diffusion of historical innovations on growth. Studies of the diffusion of banking in the United States (Bodenhorn and Cuberes, 2010), the diffusion of Protestantism in Germany (Cantoni, 2010), the diffusion of the printing press in Europe (Dittmar, 2011), and the diffusion of potato cultivation in Europe (Nunn and Qian, 2011) use the geographic distribution of an important cultural or technological innovation and analyze its effects on micro-regional economic growth, proxied by urban population growth.

Similar to our analysis, some papers study the effect of the diffusion of network innovations on economic growth at the regional level. Duranton and Turner (2010) analyze the effect of interstate highways on employment and population growth of US cities using historical plans of the interstate highway system and historical railroad maps in an instrumental-variable approach. A reduced-form relationship of historical railroad diffusion returns positive effects on long-term city growth.

Atack et al. (2010) find causal effects of gaining railroad access on population density and urbanization in the midwestern United States in the 1850s at the county level. In a differences-in-differences approach, their binary variable identifying if a county had gained railroad access by 1860 explains 58.3 percent of the increase in urbanization during the period. However, their IV setting, which employs a so-called 'congressional survey' instrument determining if a county is located on a straight line between cities that were originally intended to become connected according to congressional surveys, does not yield significant results. Furthermore Atack et al. (2011) find causal effects of gaining railroad access on the establishment size in manufacturing between 1850 and 1870.

Gregory and Henneberg (2010) analyze whether gaining a station had an effect on population growth in England and Wales between 1841 and 1911 at the parish level. Although causality cannot be established, they find that gaining access to the railroad network early led to faster population growth as compared to gaining access later. Banerjee et al. (2009) analyze the effect of access to railroads on per capita GDP growth across Chinese counties in a panel from 1986 to 2003. To establish causality, they instrument distance from a railroad by distance from the straight line between important historical cities. They find that GDP growth increases by 0.12 percent if a county is 1 percent closer to the railroad. Donaldson (2010) focusses on a trade model, however his empirical approach finds positive effects of railroad access on real income levels in India, using district level data and fixed-effects panel estimations. Keller and Shiue (2008) study the effect of institutions and the railroad on market integration using city-level data for five countries in Europe, including Germany. They find that pair-wise price gaps diminish significantly when two cities establish a railroad connection.

We are not aware of any work that analyzes micro-regional effects of railroads on economic growth for Germany or Prussia. However, in a side note, Matzerath (1985,

p. 139) calculates that the average annual population growth (1849-1858) of nine Prussian cites, that were important locations for railroad engines (but not particularly industrial cities) was 1.6 percent and thus slightly higher than the average city growth of 1.5 percent.

3 Patterns of Railroad Expansion and Urbanization

This section provides relevant historical information and periodization for Prussian railroad network expansion as well as urbanization.

3.1 The Expansion of the Railroad Network

At the beginning of the 19th century, Germany had an inadequate transportation network as compared to other European countries (Pierenkemper and Tilly, 2004). This was noted by German economist Friedrich List, who published his thoughts about the benefits of a national German railroad network in 1833 (List, 1833). List's blueprint for the railroad system connected all major cities throughout Germany. The simultaneous founding of the Zollverein (German Customs Union) led to increasing trade between the many states and fiefdoms of Germany and thus transport network expansion became desirable (Keller and Shiue, 2008).

Because of constitutional restrictions, the Prussian government was not able to raise the capital necessary to finance a public railroad network. However, Prussia was intrigued by the British example and in 1838, a law was enacted to allow private parties to build railroads. That same year, the first railroad, linking the capital of Berlin with the residency of Potsdam, was opened. The connection was, like most railroad projects until the 1870s, privately owned, financed, and operated. Since the railroad joint-stock companies easily raised capital, the network grew rapidly and by 1845 had overtaken the French system in length (Pierenkemper and Tilly, 2004). Table 1 shows the expansion of the Prussian railroad until 1880.

The government's decision, due to the lack of funds, not to directly construct a railroad network, but to approve and license private railroad enterprises, meant that railroad construction in Prussia had no central plan (Fremdling and Knieps, 1997, p. 137), but was built according to the expected profitability of lines. Consequently, the sparsely populated eastern provinces of Prussia remained unconnected until the government started building the so-called 'Ostbahn' in 1848. The state then built and operated railroads similar to those privately owned (Fremdling and Knieps, 1997, p. 138).

Access costs to railroad-knowledge were quite low and the technology could be considered to be a 'free-lunch'. The greater obstacle to railroad adoption was that it is a network technology. Benefits from adoption thus increase with the expected size of the network (Hall, 2005). In the presence of such network externalities, benefits from joining the network increase with the number of adopters. Technological diffusion usually follows an S-shaped curve, and this is also true for network technologies. In the beginning, only those agents whose expected benefits from adoption are larger than the costs of adoption will adopt. With increasing network size, benefits increase, making it feasible for a larger number of agents to adopt the technology. The diffusion often follows the trickle-down pattern observed by Comin and Hobijn (2004), starting from the from economic leader and ending with the laggards.

Such a pattern can also be observed in a periodization of German railroad network expansion following Sombart:[1] 1. Preliminary stage until 1845 – connection of major cities; 2. Construction of a framework until 1860 – uninterrupted connection of most important cities through trunk lines; 3. Full system of standard-gauge railroads until 1880 – finishing of a coarse network; 4. Ramification until 1913 – railroad supply for smaller towns through branch lines.

3.2 Urban Population Growth

The process of Prussian urbanization can be similarly subdivided into four phases following Matzerath (1985): 1. Transitional phase from 1815 to 1840; 2. Start-up phase until 1871; 3. Actual urbanization phase until World War I; 4. Stabilization phase until the end of World War II. Since industrialization and urbanization are closely related, their phases are similar, too. The period we are most interested in is the second phase, which coincides with the start of the railroad diffusion process.

[1] Cited in Henning (1995, p. 162).

Interestingly the first period (1815-1840) is characterized by population growth in cities and rural areas alike. While the urban population grew from 2.8 to 4 million (43 percent), total population grew from 10.3 to 14.9 million (45 percent). Differences can be found, however, particularly between the east and the west. West Prussian city populations grew faster than the rural population, but the reverse was true for the East. Furthermore, while rural population growth was mainly due to a birth surplus, urban population growth resulted from a net migration gain as well as a birth surplus (Matzerath, 1985, pp. 76-80).

The second phase of urbanization (1840-1871) was characterized by an increasing urban population growth. While the urban population grew from 4 to 6.7 million (68 percent), total population grew from 14.9 to 20.2 million (35 percent).[2] West Prussian cities grew much more than those in the middle provinces, which grew much more than cities in the East. Also, population growth is positively correlated with city size (Matzerath, 1985, pp. 117-123).

4 How Railroads Affect City Growth

There are various channels through which railroads might affect growth. None of these can be truly tested in our setting. Nevertheless, our proposed mechanism works through productivity gains.

The Industrial Revolution was a period characterized by a series of innovations that increased the level of productivity. Railroads in particular decreased overland transportation costs and therefore the price of inputs. Furthermore, Atack et al. (2011) argue that railroads increased competition among firms because of an increased market size. Consequently, firms attempted to increase productivity through division of labor. As industrial productivity increased, wages increased as well, attracting an inflow of workers from rural areas to urban centers (Malanima, 2010). We thus assume that the effect of railroads on urban population growth can be interpreted as an indirect effect: railroad access proxies for subsequent productivity advances, which lead to urban population growth.

Job opportunities created by factories in cities with railroad access attracted a massive inflow of rural workers (Boelcke, 1996). In fact, since railroads were usually built so that

[2] Calculated within the borders of 1840.

they passed a city tangentially, the development of cities itself changed such that they grew toward the station. The road leading toward the station usually developed into an important commercial street, attracting industry, and working-class quarters were built to surround the factories (Matzerath, 1985, p. 156).

Railroads also may foster technological diffusion, leading to further productivity gains. According to Mokyr (2002, p. 30), the 'technology of knowledge transmission' is important to the diffusion of knowledge and technology itself. The railroad can be seen as *the* knowledge diffusion technology during the Industrial Revolution. Since, at that time, most knowledge was still transferred through face-to-face contact, reducing transportation costs allowed technology to travel more easy. Being connected to the railroad network therefore meant access to a network of knowledge exchange, which reduced search and information costs, another location advantage, in addition to lower transportation costs and agglomeration gains, for firms.

During the early days of the Prussian railroad, railroads took over the function formerly performed by stage coaches – the transport of passengers and mail. Borchardt (1972) thus describes the coming of the railroad as a communication revolution. The increasing possibility for knowledge exchange through direct personal contact and the acceleration of the mail traffic led to all sorts of new possibilities for technological diffusion and knowledge spillovers.

Furthermore, railroad construction was often accompanied by the development of telegraphs lines, which were built along the railroad line and in many cases even incorporated into the railroad embankment. Thus, in many cases, railroad adoption also meant adoption of telegraphy which even further advanced the speed of communication.

5 The Data

Generalizing from urban population growth to economic growth has shown to be an acceptable approximation in cases where data on income are unavailable (Acemoglu et al., 2002). In similar vein, the outcome of interest in our empirical setup is urban population growth, which serves as a proxy for economic growth. This seems an appropriate choice in the light of the fact that urban centers were the places where most of the innovation, as well as human and physical capital, was located and accumulated.

Comprehensive and systematic population accounts were published by the Prussian Statistical Office starting in 1816, soon after Prussia's new borders were established. Urban population was counted on a triennial basis. Only places that held city rights in the year of the census were included.[3] The censuses usually provided separate accounts for civilian and military residents.[4] These data have been digitized and made available by Matzerath (1985). We corrected some digitization mistakes in the data and added missing variables using the original sources.

In contrast to much of the literature discussed in Section 2, this city-level dataset does not make use of an ad-hoc population threshold, that was recently criticized by Ploeckl (2011). Making use of the legal definition of township, the dataset consists of all Prussian cities during the nineteenth century. Consequently, the data also include a set of very small cities.[5] The average city size in the dataset increases from 3,804 civilian inhabitants in 1837, to 4,525 in 1849, to 6,703 in 1871. From these data we calculate the dependent variable for the cross-sectional analysis, the annual growth rate of the civilian population for the periods between the censuses.

Our variable of interest indicates whether a city was connected to the railroad in a given year. We use GIS-sofware to measure railroad diffusion. Using point coordinates of the city centers, we create a map of all cities in Prussia. We then overlay the city map with annual maps of the German railroad system (see IEG, 2010) to discover which cities had access to the railroad in a given year. The resulting binary variable takes the value 1 if one or more railroad lines intersect the city in a given year.

This approach sometimes returns inaccurate results because cities are represented only by point coordinates, which do not reflect their historical dimensions. Thus, it often appears as if a city had no railroad access. We correct our data using information on actual railroad access from the German handbook of cities (Keyser, 1939-1941), which specifies

[3] After the establishment of the German Reich in 1871, censuses were conducted only in years ending with 0 and 5.

[4] Unfortunately, some of the censuses between 1819 and 1837 provide only the civilian population or were not published at all.

[5] The dataset encompasses 434 cities below the usual ad-hoc threshold of 2,000 inhabitants in 1837, 364 in 1849, and 266 in 1871.

the year in which access was established and indicates the corresponding connection. This information is checked and verified with information from Königlich Preussisches Statistisches Bureau (1883).

To achieve consistency in the data, we restrict our sample to the 978 cities that held city rights in 1849. Cities that lost or gained city rights before or after 1849 are excluded from analysis with the dataset.[6] This restriction is further motivated by the census of 1849 (Statistisches Bureau zu Berlin, 1851-1855), which is unique in providing a wealth of information at the city level. We thus analyze the 1849 cross section in depth. Our various control variables[7] include access to rivaling infrastructure such as (i) main roads and (ii) navigable rivers and ports. Indicators of urbanization include (iii) pre-railroad city growth 1831-1837 and the size of the (iv) civilian and (v) military population in 1849. Indicators of industrial development include (vi) the share of citizens employed in factories and (vii) the occurrence of mining activity at the county level. As geographical endowments usually are among the major determinants of city growth, we control for (viii) the county-level concentration of large landholdings. As shown As shown by Cinnirella and Hornung (2011), the concentration of large landholdings is correlated with soil-quality and thus can be viewed as a proxy for geographical endowments and therefore the supply of food for urban markets. Further controls include (ix) the age composition and (x) the education of the urban population. These controls are aimed at capturing differences in future population growth as well as the city's progressiveness. We also calculate and control for (xi) the distance to the closest terminal or junction station of railroad lines since nearby cities are more likely to become connected to the network.

The unobserved incorporation of suburbs and smaller municipalities, as well as mergers between cities, sometimes introduces measurement error in the data and population appears to grow erratically in some cases. Our estimates could be biased in cases where cities that had access to railroads systematically had higher growth rates because of incorporations. We can control for such (xii) incorporations using the dataset provided

[6] The legal definition of township results in the omission of a couple of locations. According to Ploeckl (2011), in Saxony, these locations were actually some of the fastest growing during the Industrial Revolution. When looking at the Prussian census data, we find 39 towns entering the census during the period 1849-1885 and thus gaining legal city rights. These cities had an average size of 4,915 inhabitants in 1871 and an average annual growth of 3 percent during the period 1871-1885. The Prussian-wide average was 1 percent in this period.

[7] See Appendix Appendix A for more specific definitions and sources.

by Matzerath (1985), which also indicates whether a city changed dimensions in a given period.[8]

We provide descriptive summary statistics for the 1848 cross section in Table 2. Unless specified otherwise, all data refer to the base year 1849. Roughly 8 percent of the cities in the sample were connected to the railroad, 41 percent had access to a main street, 20 percent had access to navigable waterways, and 10 percent were located in a county with mining activity.[9]

Table 3 compares the annual urban population growth rate of the cities that were connected to railroads to that of those that were not. We report growth rates for a range of periods from 1821 to 1871, thus allowing a comparison of growth rates before and after railroad access was gained. In this case, the control group consists of cities that were not connected by 1848 (column 1); the treatment group consists of those cities gained access during the period 1838 to 1848 (column 2).[10] In column 3, we find that there was a difference in growth rates, albeit a small one, before the railroads were established. However, we find a strong divergence in growth rates between the groups from the period 1843-1846.

We also present the growth rates of the cities with a terminal or junction station, which will be excluded from our subsequent analysis, so as to present a complete picture. These growth rates are separately presented in column 4 of Table 3 and behave very similar to those of the treatment group at first, but start to be slightly higher, on average, after the period 1846-1849.

6 The Causal Effect of Railroad Access on City Growth

As mentioned earlier, the direction of causality between railroad access and urban growth is not straightforward. Railroads might induce population growth in connected cities, but having access itself might not be independent of a city's importance, wealth, and growth prospects. Thus, there might be an omitted variable that is correlated with both city

[8]Unfortunately we sometimes still observe abnormal jumps in the population accounts which is why we exclude observations whose annual population growth exceeds minus or plus 10 percent.

[9]The share of factory workers, as well as the school enrollment rate, exceeds 100 percent in some cases, presumably due to workers and schoolchildren commuting from outside of the city.

[10]The control group however includes cities that gained access in the period 1849-1871.

growth and railroad access. Reverse causality and omitted variable bias could be serious issues in this setting.

To address issues of reverse causality and omitted variables, we take two different econometric approaches to establish causality from railroad access to urban population growth: instrumental-variable estimations and fixed-effects panel regression. At this point, we exclude from our sample all cities that are most likely to have gained access to the railroad for reasons endogenous to our dependent variable, namely, the terminal and junction stations of the railroad network. Since, up until the 1860s, railroads were built to connect major cities, terminal stations are located in those cities that were the reason for the construction of the line and thus do not qualify for the assumption of random assignment.[11]

Twenty-one railroad lines were built in Prussia during the period 1838-1848. In Table 4, we provide information on year of construction and length, as well as passenger and freight transport statistics, for each of the lines during the year 1848. For now, we assume that all other cities had access to railroad technology simply because they were located en route between two major cities.

6.1 Cross-Sectional Estimates (OLS)

In a first step, we estimate the effect of railroad access on urban growth in a standard cross-regional growth regression by ordinary least squares (OLS). By doing so, we can draw on a variety of unique city-level control variables provided by the Prussian census of 1849. In addition, we can calculate population growth rates between different censuses in order to analyze short- and long term effects. This results in a model where the urban population growth rate PGR[12] in a variety of periods (t) is a function of railroad access RA in 1848 and other explanatory factors X:

$$PGR_t = \alpha_1 + \beta_1 RA_{1848} + X'_{1849}\gamma_1 + \varepsilon_1. \quad (1)$$

We emphasize here that the explanatory factors X include a lagged dependent variable

[11] Thirteen (25) of the 20 (60) cities with more than 20,000 (10,000) inhabitants were terminal or junction stations of a railroad line by 1848.

[12] The urban population growth rate is defined as $\frac{ln(POP_{t2})-ln(POP_{t1})}{(t_2-t_1)}$.

to account for dynamic aspects of urban growth. Each column of Table 5 reports OLS estimates of urban population growth on railroad access for different periods between 1831 and 1871. We find that being connected to the railroad in 1848 significantly increased annual population growth by 0.9 percentage points during the period 1849-1871 (column 2). Comparing all periods across columns 3 to 9, we find that the annual population growth generated by railroad access varies between 0.4 and 1.1 percentage points.[13] The coefficient increases in the later periods under consideration, which hints at increasing long-term effects from railroad access.

Note that the counterfactual specification in column 1, where we regress pre-rail population growth 1831-1837 on railroad access until 1848, is insignificant. This means that cities that were connected by 1848 had similar or worse growth patterns before the coming of the rail compared to those who were not. Thus we find no pre-trend in rail access that favored cities with high growth rates.[14]

Throughout the specifications, we control for the increasing number of cities that became connected to railroads during the period under consideration with a dummy variable. The point estimates of this dummy also hint at significant positive effects for the later adopters, albeit the coefficients are comparatively smaller in most cases.

Interestingly, neither connection to a main street nor, in most cases, to navigable waterways has positive effects on urban population growth after 1849. Controlling for these other modes of transportation also rules out the possibility that railroad lines were just built alongside the main trade routes and reinforced their status. As expected, we find that the lagged dependent variable pre-railroad city growth and city size in 1849 significantly determine subsequent growth in many specifications.

To separate railroad access from other indicators of industrialization, we control for the share of factory workers in the city population and the occurrence of mining activity at the county level. Both indicators are significant and positively correlated with urban

[13] We exclude all cities that had a terminal or junction station on a railroad line in 1848. Furthermore, some cities drop from the sample as they lose city rights. Additionally, we exclude all cities with incorporations and cities with unrealistically low (-10 percent) or high (+10 percent) growth rates, since these increases might be caused by unobserved incorporations. These exclusions explain the decreasing number of observations over the periods. Point estimates are similar or higher if we do not exclude observations with unrealistic jumps.

[14] We find similar results when extending the period to 1821-1837 in all our specifications (not shown). For better comparability, we show the period 1831-1837 since some observations are missing for 1821.

population growth. Also, cities in counties with a high share of large farms, which proxies for the supply of agricultural products to urban markets grow faster than others. The dominance of small family farms thus might have retarded urban population growth. Interestingly, distance to the next terminal or junction station is often significant with a positive sign. This means that increasing the distance to those cities seems to foster population growth. From this could be concluded that major railroad cities attracted most of the regional migration, leaving less migration to nearby cities. We also test for interaction effects of streets and rivers with railroad access – both turn out to be insignificant (not shown).

6.2 Instrumental-Variable Estimates

To this point, we have assumed that railroad lines were built to connect major cities and that cities located along the way were able to gain access to the railroad network by chance. Nevertheless, OLS estimates of the relationship might be biased in cases of omitted variables. Thus, we use an instrumental-variable approach to resolve the omitted variable concern. Similar to the approaches taken by Atack et al. (2010) and Banerjee et al. (2009), we predict actual railroad access RA in 1848 with the potential for railroad adoption in 1848, being located within a straight-line corridor SLC:

$$RA_{1848} = \alpha_2 + \beta_2 SLC_{1848} + X'_{1849}\gamma_2 + \varepsilon_2. \quad (2)$$

Until the 1860s, Prussian railroads were built to connect major cities. Under the assumption that lines were exclusively built to establish a fast connection between major cities A and B, cities en route had the chance to become connected to the railroad simply because they were located on this straight line. Thus, all cities on a straight line between A and B were randomly assigned to being able to adopt railroad technology. If only these cities gained access, our OLS estimates would be unbiased. In reality, we observe that connections sometimes deviate from the straight line. Cities located on such a deviation might have gained access for endogenous reasons.

Our instrument SLC is a dummy that takes the value 1 for all cities located on a straight line between terminal or junction stations and the value 0 for all others. We thus use variation in the potential for railroad adoption to instrument actual access. The idea behind this instrument is that deviation from the straight line bears additional costs.[15] If the railroad actually deviates from the straight line in order to connect a city, the additional costs of land acquisition, building tracks and stations, and additional operational costs, as well as the extension of travel time between the major cities, would be immense.[16] On the other hand, deviation from the straight line might reduce costs in the event of natural geographical obstacles such as lakes and hills. Column 3 of Table 4 shows that large shares of the lines were built linearly, implying the high costs of deviation from the straight line.

Using GIS techniques, we connect the terminal and junction cities between which railroads were constructed with straight lines. Furthermore, we create a buffer around these railroad lines.[17] All cities within this corridor could potentially connect to the railroad due to the fact that they were accidentally located on a linear line between major cities. The instrument takes the value 1 for all observations within the corridor while all other observation take the value 0. This means that all cities that had access to railroads despite not being located on a straight line are taken as endogenous.

Obviously, deviation from the straight line did not happen exclusively in order to connect a certain city and geography introduces random measurement error into our instrument. Rivers are one of the main reasons to deviate from the straight line since bridge building was expensive and orthogonality was required. Thus we allow the buffer to expand the linear line by 1.5 kilometers in each direction.[18] In contrast to Banerjee et al. (2009), we do not use the distance to the straight line as an instrument as it might be correlated with the distance to a major city. In such a setting, the exclusion restriction would be violated if proximity to major cities affects growth directly.

[15] For example, the connection Cologne-Duisburg-Minden was originally intended to pass through the city of Lünen, which is located close to the straight line. This routing would have bypassed the city of Dortmund, which was to become a major industrial center. It was only the city's willingness to build the station at its own expense and an additional contribution of 3,000 Thaler that convinced the railroad company to build the costly detour, with extra mileage of roughly 10 km, to connect Dortmund (Ziegler, 1996, p. 310).

[16] The average construction stock for a Prussian mile (7.53 km) of railroad was roughly 350,000 Thalers for lines built until 1848.

[17] See Figures 1 and 2 for examples.

[18] Extending the buffer size to 2, 3, or 5 kilometers leads to similar results and will increase the power of the instrument. Note that the average distance is 10.8 km to the next nearby city and 17.4 km to the next nearby city with more than 3,000 inhabitants.

Figure 2 provides some helpful examples. The map displays a section of the railroad system centering around Berlin. We observe five railroad lines radiating from the Prussian capital. The hollow circles mark cities that have a railroad station, while black circled cities do not. The hash lines show the actual routing of the railroads, whereas the dark tubes show the straight-line corridor.[19]

For example, in 1842/43, the connection between Berlin and Stettin (today Szczecin) was built to provide Berlin with fast access to the Baltic Sea. The city of Biesenthal did not become connected even though it was located only 3 km from the actual line. However, it is not located within the corridor either, and is assigned the value 0 for the instrument. Interestingly, Biesenthal did not open a station until 1865 after a street was built toward the railroad line. On the other hand, it looks like the Berlin-Stettin line takes a marked deviation from the straight line in order to connect the city of Greiffenberg. Actually, Greiffenberg became connected to a different line only in 1863 and therefore the noted deviation seems to be for geographical reasons or because land could not be acquired.

Another, very different, example is the connection Berlin-Potsdam, which was extended to Magdeburg and completed in 1850. This line deviates markedly from the straight line in order to connect the cities of Brandenburg, Genthin and Burg. Since these cities are not located within the straight-line corridor, the instrument will take the value 0, assuming these cities gained access to the railroad for endogenous reasons.

Another observation that can be made from Figure 2 is that railroads do not exclusively follow the established main roads. In many cases, it seems as if railroads took a more direct line, linking previously unconnected cities with the network.

Table 6 reports estimates using the straight-line corridor as an instrument. Panel A shows first-stage results of the IV approach. The instrument *SLC* is significantly correlated with actual railroad access. First-stage F-statistics are high and confirm the power of the instrument. Second-stage results, reported in Panel B, show causal effects of railroad access on urban population growth. We find a significant increase in urban population growth due to railroad access of 2.1 percentage points during the period 1849-

[19] Black lines mark the routing of main streets.

1871. Across all shorter periods under consideration, the effect varies between 1.1 and 2.2 percentage points for a city that had gained access until 1848. Coefficients estimated by IV are approximately twice as large as coefficients derived from OLS estimations. The OLS coefficients might be biased downward in case of an omitted variable – for example, cities with lower growth prospects might have influenced routing in order to become connected.

Again, it is reassuring that the counterfactual model for the period 1831-1837 does not yield significant results (column 1 of Table 6).

The exclusion restriction would be violated if the instrument had a direct or indirect effect on the outcome. This would be the case if location in the straight-line corridor was associated with urban population growth through a channel other than the railroad; for example, if the corridor coincided with historical trade routes that foster growth. Since we control for street and waterway access in every specification, we eliminate such channels. Nevertheless, by estimating the reduced-form relationship of urban growth on being located within the straight-line corridor, we find no correlation with pre-railroad growth 1831-1837 (column 1 in Panel C of Table 6). This suggests that the instrument works only through the channel of railroad access and the exclusion restriction does not appear to be violated.

Thus far, we controlled for those cities that became connected to the railroad during the periods under consideration by using a single dummy variable. Alternatively, we can use different dummy variables for additional adopters for each period. Results of such an approach are presented in Table 7. Column 1 shows that cities that gained access to railroads during *any* of the periods did not grow significantly faster than other cities in the period 1831-1837. On the other hand, in column 2 we find that adopters of *all* periods grew comparatively faster during the period 1849-1871. As columns 3-9 show, this increase in growth rates often occurred shortly prior to the year of becoming connected and increased thereafter. Presumably, this is a preceding effect of railroad construction work. This result also helps to answer the question if railroads followed or induced growth.

6.3 Robustness Tests in Different Subsamples

This section presents four different approaches for ensuring that the treated and untreated cities in the sample are as comparable as possible.

In addition to controlling for the size of a city, we also test whether restricting our sample to cities smaller than 3,000 inhabitants in 1837 (before the first railroad was built in Prussia) changes the results.[20] Cities with fewer than 5,000 inhabitants were considered as small by the Prussian administration. Restricting the sample to even smaller cities with a population of only 3,000, guarantees that none of these were important enough to become connected because of sheer size. Although the instrument loses some of its power, Panel A of Table 8 shows that this sample yields results similar to those for the full sample.[21] Nevertheless, it seems as if these smaller cities needed some time to attract immigration and the result for the period 1849-1852 is insignificant.

Furthermore, we find that the introduction of 25 district dummies or the exclusion of the sparsely populated eastern provinces[22] does not change the results qualitatively (not shown).

To confirm our previous results, we also employ matching techniques. By using geographical matching, we compare cities that had access to railroads in 1848 directly with their two next unconnected neighbors.[23]

We repeat the instrumental-variable estimations in this matched sample.[24] The advantage of such an approach is that nearby cities might be very similar in many aspects, reducing the omitted variable bias. Since all cities in the matched sample are located close to the railroad, we also exclude remote areas from our regressions. This methodology thus directly estimates the effect of having a railroad station compared to merely being located in proximity to one. Panel B in Table 8 reports results using the geographically matched sample. They confirm previous results and show that cities with a railroad station grow significantly faster than their nearby neighbors.

[20] Forty of the cities below 3,000 inhabitants had access to railroads by 1848; 627 did not.

[21] Restricting the sample to cities with a population between 1,500 and 5,000 inhabitants, thus excluding a large number of very small and very big cities, yields similar results (not shown).

[22] These are the predominantly Polish-speaking provinces of Prussia, Poznan, Pomerania and Silesia.

[23] The mean distance is 11.8 km to the first nearby city and 15.2 km to the second nearby city.

[24] Unconnected cities that were matched to two or more connected cities are weighted correspondingly.

We further expand the matching approach to propensity score matching. The aim of propensity score matching is to compare the outcome (population growth) for observations (cities) that are as similar as possible and – at best – differ only in their assignment to the treatment (railroad access). Propensity score matching is particularly useful in cases where assignment to the treatment group is not explicitly random. In our case, the worry might be that even though cities are located on a straight line between terminal or junction stations, they did not gain access just because of this fact.

To obtain a highly comparable sample, we match treated and untreated observations using a set of pre-railroad controls. For this purpose, we exclude all cities that gained access to railroads during the period 1849-1871 from the matching. Since the first Prussian railroad was built in 1838, we match cities by (i) their size in 1837, (ii) their population growth during the period 1821-1837, and normalized numbers of (iii) merchants in 1819, (iv) looms in 1819, (v) Protestants in 1816, (vi) private dwellings in 1821, (vii) commercial buildings in 1821, and (viii) the insurance value of buildings against fire in 1821. The matching variables are targeted at matching cities regarding their size and commercial development previous to railroad construction.

Since endogeneity of railroad access could still be an issue, we do not show results of the propensity score matching but instead use the instrumental-variable approach on the matched and weighted sample. Propensity score matching is done using radius and kernel matching techniques. To reduce the inclusion of poor matches, we make use of the common support condition.

Results of the IV estimation in a radius-matched sample are shown in Panel C of Table 8. Radius matching finds all untreated observations that are within distance of a specified caliper to a treated observation according to the propensity score. We find a significant positive increase of 1.1 percentage points in annual population growth over the entire period. Although point estimates are similar compared to previous samples, standard errors are higher and coefficients become insignificant in two subperiods.

Results of the IV estimation in a sample matched using a nonparametric kernel approach are shown in Panel D of Table 8. Kernel matching compares the outcome of treated observations to a weighted average of outcomes of untreated observations. Observations that are more similar receive more weight than others. We find a significant positive increase of 1.7 percent in annual population growth over the whole period. Again, standard errors are higher and coefficients become insignificant in three subperiods.

It is reassuring that the counterfactual model for the period 1831-1837 does not yield significant results in any specification (column 1 in each panel); rather, the relationship seems to be negative in this period.

6.4 Panel Data Estimates

In a second approach to estimate the effect of railroad access on urban growth, we use panel techniques. The advantage of such an approach is the possibility of overcoming time-invariant unobserved heterogeneity by including fixed effects. For our purpose, we can include city fixed effects, allowing us to exploit exclusively within-city variation.

To eliminate concerns of reverse causality, we regress city size, measured as the natural logarithm of the total civilian population *ln(POP)* in city i in year t, on a dummy variable indicating railroad access RA in the previous year.

$$lnPOP_{it} = \alpha_i + \tau_t + \beta_3 RA_{t-1} + X'_{it}\gamma_3 + \nu_{it}. \tag{3}$$

We further include city fixed effects α_i as well as time fixed effects τ_t, capturing national trends in population growth in our regressions. In such a panel setting, the estimated coefficient of interest β_3 returns the additional growth in population levels for cities that had access to railroads, compared to those that did not, after gaining access. The covered period ranges from 1840, just after the first railroad was built in Prussia, to 1864, just after the end of Phase 2 of railroad network expansion – the connection of major cities. Since the censuses provide triennial data, we derive a panel consisting of nine repeated cross sections. The only information published at this frequency at the city level is population counts. Thus, the only available control variables *X'* are the military population and a dummy that controls for incorporation of municipalities provided by Matzerath (1985). By excluding all cities that had a terminal or junction station, we again try to minimize issues of endogeneity.

We present panel estimates in Table 9. The first specification reports estimates in a pooled sample, including year fixed effects (column 1). Column 2 introduces city fixed effects and thus shows the within-city effect of railroad access in one year on subsequent additional population growth. The dummy variable indicating railroad access switches from 0 to 1 when a city becomes connected to the railroad network. The results indicate that railroad access additionally increases urban population levels by 4.7 percent over a period of three years. At an annual rate of 1.6 percent, this coefficient is very close in magnitude to the ones derived in the cross-sectional settings. Interestingly, the coefficient estimated in the fixed-effects model is also very close to the pooled sample, indicating low levels of unobserved heterogeneity at the city level.

A third specification in column 3 adds 324 county fixed effects interacted with year fixed effects. Such a specification captures county-wide shocks during a period that affected all cities within the same county equally. An obvious example could be the discovery of mineral resources that introduces a shock to a county's economy. Other examples include shocks to the food supply or epidemics. Interestingly, the point estimate increases in magnitude to the previous specification using only city fixed effects.

Using a dummy that switches from 0 to 1 after gaining access only allows identifying changes in growth rates in the subsequent period. To be able to show more long-term effects of gaining access, we code separate dummies for all periods after adoption. The first dummy takes the value 1 in the first period after a city became connected to the railroad and switches back to 0 for all periods after adoption. The second dummy takes the value 1 only in the second period after a city gained access to the railroad, and so forth. This way, we compare cities in the same stage of railroad access to all others. Such a model allows us to identify nonlinear trends from railroad access. Column 4 in Table 9 shows significant positive effects of railroad access over all periods. Furthermore, the coefficients increase throughout the periods, indicating a positive nonlinear growth trend.

In column 5 of Table 9 we further restrict our sample to cities that had less than 3,000 inhabitants in 1837, before the first railroad was built. By doing so, we aim to exclude all major cities that might have been connected to the railroad because of their size and importance. We find very similar results in this subsample, again indicating that the estimated effect is not driven by the growth of larger cities.

One drawback of our panel estimation is the lack of time-variant control variables. Thus, we are not able to account for trends in, for example, industrialization occurring during the period, which might have influenced both railroad access and city growth.

Nevertheless, we can address endogeneity issues in the panel setting using the instrumental-variable approach discussed in Section 6.2. Whenever new lines were built, new straight-line corridors were established, providing over-time variation in the instrument. Thus we construct straight-line corridors on a triennial basis and use them to generate an instrument that carries time variation.

Results of the first-stage relationship between potential railroad diffusion and actual access are presented in column 6. We find a similar correlation as in the cross-sectional setting. The second-stage results are presented in column 7. The estimated coefficient is 7.7, meaning that a city that was connected to the railroad subsequently experienced an increase in growth of roughly 2.6 percent per year.

6.5 Causal Effects on Firm Size and the Number of Firms

Thus far, we have estimated a reduced-form relationship of railroad access on population growth, without testing the proposed mechanism that access to a railroad increases the productivity of firms and creates more jobs. Atack et al. (2011) argue that, in the nineteenth-century, access to railroads decreased transportation costs which led to the extension of markets and to an increase of competition among firms in the US. This, in turn, led to an increase in the division of labor and consequently an increase in establishment size. In an additional approach, we test if railroad access increased firm size and thus might have caused urban population growth in nineteenth-century Prussia.

The 1849 census (Statistisches Bureau zu Berlin, 1851-1855) includes a factory census which allows us to test the effect of railroad access on firm size at the city level. The census reports the number of factories and of workers in 119 different product categories. We calculate the average size of all factories at the city level and use it as an alternative dependent variable in the same empirical set-up employed in the previous sections 6.1

and 6.2.[25] Since such data is exclusively available for the 1849 cross section, we can only estimate level effects of railroad access on firm size.

The results of this approach are presented in Table 10, starting with an OLS estimation in column 1. We then proceed to estimate the relationship using the straight-line corridor instrumental-variable approach of Section 6.2. Column 2 shows that firms located in a city with a railroad station were 73 percent larger than in cities without a station.

This raises the question if firms in cities with railroad access grew bigger or if these cities were also able to attract more firms? Columns 3 and 4 in Table 10 show that this is not the case. We find a negative effect of railroad access on the number of firms. This effect, though significant in the OLS estimation, is insignificant in the instrumental-variable approach. This means that railroad access affects industry location at the intensive margin rather than the extensive margin.

7 Conclusion

This paper tests the hypothesis that railroads induced economic growth at the micro-regional level. Since data constraints do not allow us to directly analyze whether railroads affect productivity and thereby growth, our identification estimates the reduced-form relationship of railroad access and growth. By using city-level population growth as a proxy for economic growth, we add sub-national evidence to the literature, which to date was mostly focussed on macroeconomic effects. Basically, we test whether cities that gained access to the new transport technology – railroads – grew faster than others analyzing the case of nineteenth-century Prussia.

We find that railroads had a significant effect on urban population growth in the short as well as in the long run. Cities that were connected to the railroads as early as 1848 grew roughly 1 to 2 percent faster than their unconnected counterparts. Added to an average growth rate of 0.9 percent during the period 1849-1871, this effect is quite substantial.

To address basic issues of endogeneity, all our estimates exclude cities that were locations of terminal or junction stations of the railroads. This means that the estimated

[25] Please note that these regressions do not control for the share of factory workers in the city population. However, if they do, the coefficient on railroad access 1838-48 is hardly affected.

effects cannot to be generalized to all cities. Nevertheless, as we have shown, the excluded cities developed even higher growth rates, meaning that our analysis might underestimate the effect of railroad access on growth.

We also contribute further evidence in the debate over whether railroads induced economic growth or just appeared as a consequence of it. By estimating counterfactual models, in a series of different specifications, where we regress pre-railroad growth on subsequent railroad adoption, we find no evidence that railroads appeared as a consequence of a previous growth spurt.

Moreover, results from our panel approach show that it was particularly advantageous to become connected to the railroad network at an early stage. The longer a city was connected to the network, the higher were its long-term growth rates.

Finally we are also able to provide evidence that railroads affected city growth through creating industrial jobs thus attracting an inflow of workers and their families. Our results show that cities with railroad access hosted much bigger factories than unconnected cities (73 percent), triggering population growth through the demand for workers from the industrial sector.

In summary, we conclude that railroads reshaped the economic geography of Prussia during the nineteenth century. During that period, previously unimportant cities were able to achieve economies of agglomeration that prevail even today, while cities of earlier historical importance fell into oblivion.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, 117, 1231 – 1294.

Atack, J., F. Bateman, M. Haines, and R. A. Margo (2010): "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-1860." *Social Science History*, 34, 171 – 197.

Atack, J., M. R. Haines, and R. A. Margo (2011): "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70." in *Economic Evolution and Revolution in Historical Time*, ed. by P. Rhode, J. Rosenbloom, and D. Weiman, Stanford: Stanford University Press, 162 – 179.

Banerjee, A., E. Duflo, and N. Qian (2009): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," mimeo.

Bodenhorn, H. and D. Cuberes (2010): "Financial Development and City Growth: Evidence from Northeastern American Cities, 1790-1870," NBER Working Papers 15997, National Bureau of Economic Research.

Boelcke, W. A. (1996): "Städtewachstum und Eisenbahnentwicklung in deutschen Bundesstaaten," in *Stadt und Verkehr im Industriezeitalter*, ed. by H. Matzerath, Köln: Böhlau Verlag, 23–40.

Borchardt, K. (1972): *Die Industrielle Revolution in Deutschland*, München: Piper.

Cantoni, D. (2010): "The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands," Economics Working Papers 1260, Department of Economics and Business, Universitat Pompeu Fabra.

Cinnirella, F. and E. Hornung (2011): "Landownership Concentration and the Expansion of Education," CESifo Working Paper Series 3603, CESifo Group Munich.

Comin, D. and B. Hobijn (2004): "Cross-country Technology Adoption: Making the Theories Face the Facts," *Journal of Monetary Economics*, 51, 39–83.

Dittmar, J. E. (2011): "Information Technology and Economic Change: The Impact of The Printing Press," *Quarterly Journal of Economics*, 126, 1133–1172.

Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Working Paper 16487, National Bureau of Economic Research.

Duranton, G. and M. A. Turner (2010): "Urban Growth and Transportation," mimeo, University of Toronto.

Eisenbahn-Büreau, T. (1855): *Statistische Nachrichten von den Preussischen Eisenbahnen*, vol. 1, Berlin: Verlag von Ernst & Korn.

Fishlow, A. (1965): *American Railroads and the Transformation of the Antebellum Economy*, Cambridge: Harvard University Press.

Fogel, R. W. (1962): "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings," *Journal of Economic History*, 22, 163–197.

Fremdling, R. (1977): "Railroads and German Economic Growth: A Leading Sector Analysis with a Comparison to the United States and Great Britain." *Journal of Economic History*, 37, 583 – 604.

——— (1983): "Germany," in *Railways and the Economic Development of Western Europe, 1830 - 1914*, ed. by P. O'Brien, New York: St. Martin's Press, 121–147.

——— (1985): *Eisenbahnen und deutsches Wirtschaftswachstum 1840-1879: Ein Beitrag zur Entwicklungstheorie und zur Theorie der Infrastruktur*, Dortmund: Gesellschaft für Westfälische Wirtschaftsgeschichte, 2nd ed.

Fremdling, R. and G. Knieps (1997): "Competition, Regulation and Nationalization: The Prussian Railway System in the Nineteenth Century," in *Selected Cliometric Studies on German Economic History*, ed. by J. Komlos and S. Eddie, Stuttgart: Steiner, 136 – 150.

Gregory, I. N. and J. M. Henneberg (2010): "The Railways, Urbanization, and Local Demography in England and Wales, 1825-1911." *Social Science History*, 34, 199 – 228.

Hahn, H.-W. (2005): *Die Industrielle Revolution in Deutschland*, München: Oldenbourg, 2 ed.

Hall, B. H. (2005): "Innovation and Diffusion," in *The Oxford Handbook on Innovation*, ed. by J. Fagerberg, D. C. Mowery, and R. R. Nelson, Oxford: Oxford University Press, 459–484.

Henderson, J. V. (2005): "Urbanization and Growth," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Amsterdam: Elsevier, vol. 1B of *Handbook of Economic Growth*, chap. 24, 1543–1591.

Henning, F.-W. (1995): *Die Industrialisierung in Deutschland 1800 bis 1914*, Paderborn: Verlag Ferdinand Schöningh, 9th ed.

IEG (2010): "IEG-Maps - Kartenserver am Institut für Europäische Geschichte Mainz," http://www.ieg-maps.uni-mainz.de/, accessed November 2010.

Keller, W. and C. H. Shiue (2008): "Institutions, Technology, and Trade," NBER Working Papers 13913, National Bureau of Economic Research.

Keyser, E. (1939-1941): *Deutsches Städtebuch - Handbuch städtischer Geschichte*, Stuttgart: Kohlhammer.

Königlich Preussisches Statistisches Bureau (1883): *Die historische Entwicklung des Deutschen und Deutsch-Oesterreichischen Eisenbahn-Netzes vom Jahre 1838 bis 1881*, Berlin: Verlag des Königlichen Statistischen Bureaus.

Landes, D. S. (1969): *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*, New York: Cambridge University Press.

List, F. (1833): *Über ein sächsisches Eisenbahn-System als Grundlage eines allgemeinen deutschen Eisenbahn-Systemes, und insbesondere über die Anlegung einer Eisenbahn von Leipzig nach Dresden*, Leipzig: A. G. Liebeskind.

Malanima, P. (2010): "Urbanization," in *The Cambridge Economic History of Modern Europe Volume 1: 1700-1870*, ed. by S. Broadberry and K. H. O'Rourke, Cambridge University Press, 235–263.

Matzerath, H. (1985): *Urbanisierung in Preußen 1815-1914*, Stuttgart.: Kohlhammer/Deutscher Gemeindeverlag.

——— (1996): "Verkehr und Stadtentwicklung. Stand und Möglichkeiten historischer Forschung," in *Stadt und Verkehr im Industriezeitalter*, ed. by H. Matzerath, Köln: Böhlau Verlag, VII–XX.

Mokyr, J. (2002): *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.

Mützell, A. A. (1823-1825): *Neues Topographisch-Statistisch-Geographisches Wörterbuch des Preussischen Staats*, vol. 5 & 6, Halle: Karl August Kümmel.

Nunn, N. and N. Qian (2011): "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment," *Quarterly Journal of Economics*, 126, 593–650.

O'Brien, P. (1983): "Transport and Economic Development in Europe 1789-1914," in *Railways and the Economic Development of Western Europe, 1830 - 1914*, ed. by P. O'Brien, New York: St. Martin's Press, 1–27.

Pierenkemper, T. and R. Tilly (2004): *The German Economy during the Nineteenth Century*, New York and Oxford: Berghahn Books.

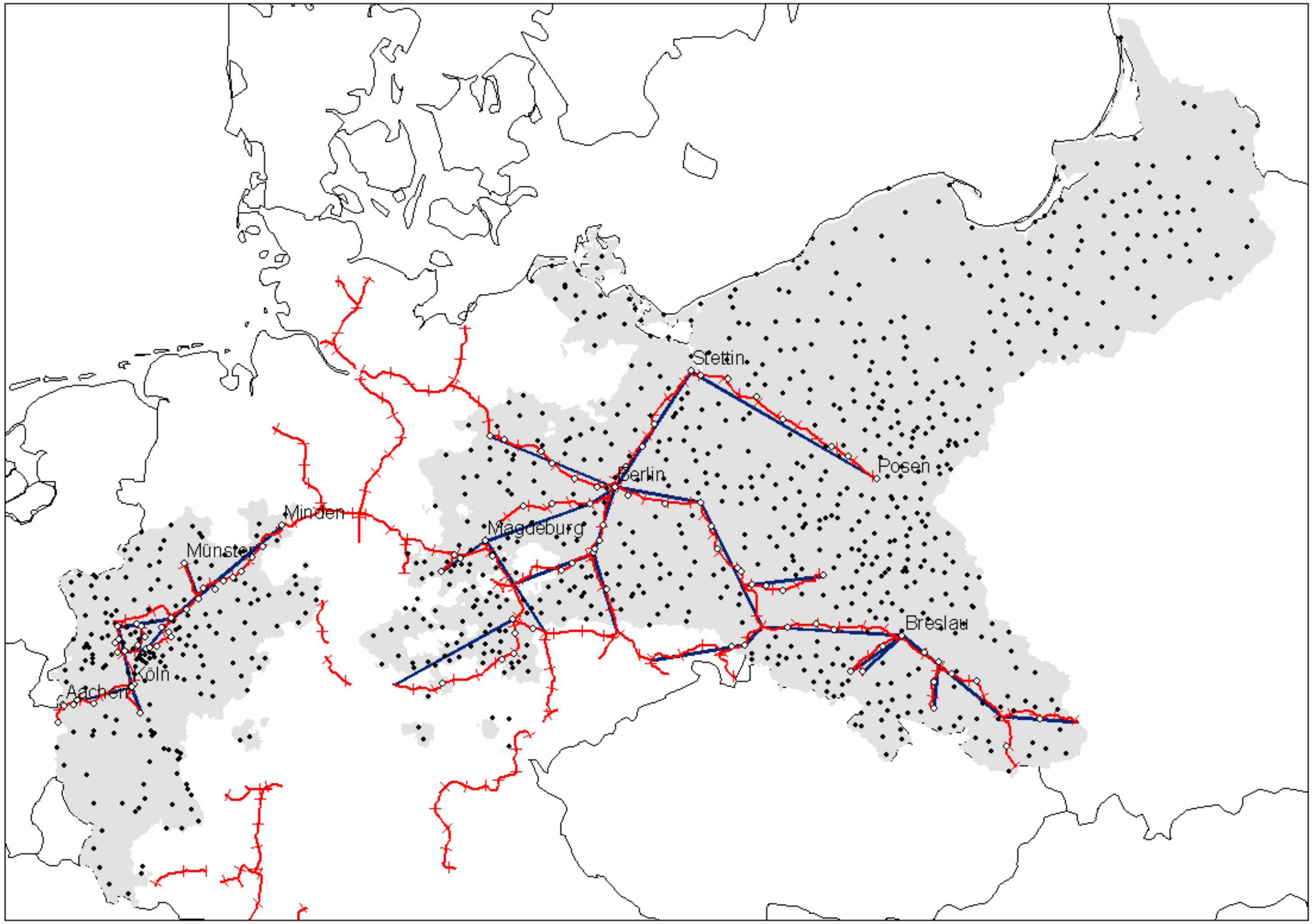
Ploeckl, F. (2011): "Towns (and Villages); Definitions and Implications in a Historical Setting," Economics Series Working Papers 536, University of Oxford, Department of Economics.

Rostow, W. W. (1962): *The Process of Economic Growth*, W. W. Norton & Co, 2nd ed.

Statistisches Bureau zu Berlin (1851-1855): *Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849*, vol. 1-6b, Berlin: Statistisches Bureau zu Berlin.

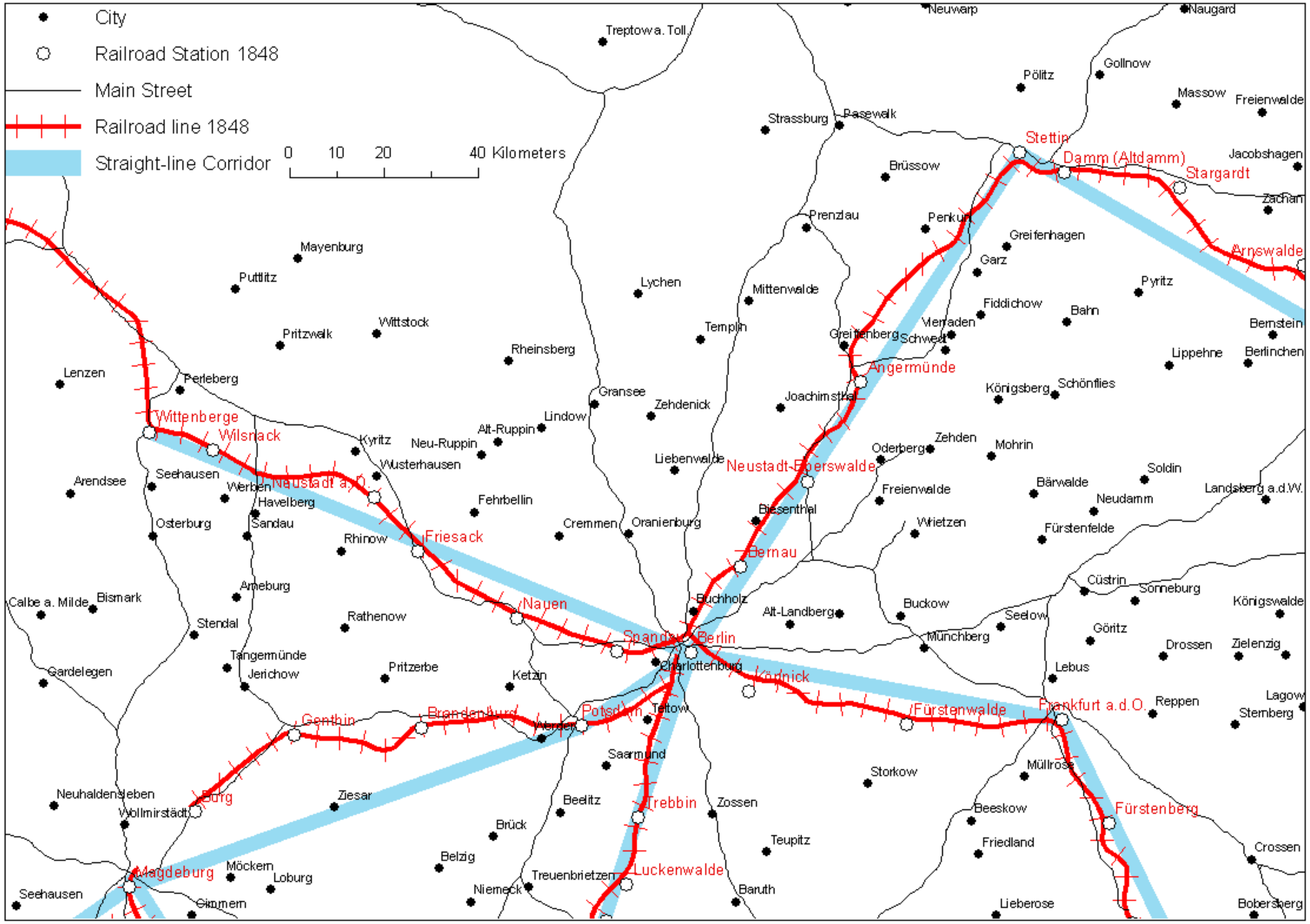
Ziegler, D. (1996): *Eisenbahnen und Staat im Zeitalter der Industrialisierung: Die Eisenbahnpolitik der deutschen Staaten im Vergleich*, Stuttgart: Franz Steiner Verlag.

Figure 1: German-Prussian Railroad Network, 1848.



Note: Gray area indicates Prussian territory in 1848. Hash lines indicate railroad routings in the German Reich. Tubes indicate the straight-line corridor using a 1.5 km buffer. Hollow circles indicate cities that had a railroad station by 1848. Black circles indicate cities that did not have access by 1848. Source: Own illustration; see main text for details.

Figure 2: Straight-Line Corridors and Actual Railroad Acess Around Berlin.



Note: Hash lines indicate railroad routings from Berlin to Stettin, Frankfurt (Oder), Magdeburg, and Wittenberge. Black lines indicate main streets. Tubes indicate the straight-line corridor using a 1.5 km buffer. Hollow circles indicate cities that had a railroad technology by 1848. Black circles indicate cities that did not have access by 1848. Source: Own illustration; see main text for details.

Table 1: Prussian Railroad Network Expansion

Year	State owned State administration	Private owned State administration	Private owned Private administration	Total
1838	-	-	34.7	34.7
1840	-	-	232.2	232.2
1845	70.0	-	1308.6	1378.6
1850	657.8	480.9	2729.9	3868.6
1855	1859.3	510.7	2719.4	5089.4
1860	2550.4	1278.0	3340.9	7169.3
1865	2986.5	1430.2	4237.5	8654.2
1870	3505.7	1820.5	6196.8	11523.0
1875	4390.9	2735.5	9750.7	16877.1
1880	11455.3	3649.5	5243.6	20348.4

Note: Length of the railroad network in kilometers at the end of the specified year. Source: Königlich Preussisches Statistisches Bureau (1883, p. 161)

Table 2: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Annual city growth 1816-31	0.014	0.010	-0.030	0.052	910
Annual city growth 1831-37	0.014	0.013	-0.043	0.089	898
Annual city growth 1849-71	0.009	0.010	-0.021	0.081	906
Annual city growth 1849-52	0.014	0.016	-0.075	0.095	929
Annual city growth 1852-55	0.004	0.015	-0.063	0.100	924
Annual city growth 1855-58	0.009	0.017	-0.074	0.095	914
Annual city growth 1858-61	0.012	0.014	-0.080	0.097	926
Annual city growth 1861-64	0.012	0.016	-0.039	0.096	924
Annual city growth 1864-67	0.004	0.017	-0.071	0.087	919
Annual city growth 1867-71	0.006	0.016	-0.049	0.090	919
Rail access 1838-48	0.081	0.274	0	1	934
Straight-line corridor = 1	0.032	0.176	0	1	934
Street access = 1	0.411	0.492	0	1	934
Waterway access = 1	0.199	0.400	0	1	934
Civilian population (log)	7.847	0.713	5.568	11.159	934
Military population (log)	2.351	1.999	0	8.684	934
Factory workers (share)	0.040	0.118	0	1.605	934
Mining (county level)	0.103	0.304	0	1	934
Large farming (county level)	0.027	0.025	0	0.213	934
Age composition	0.342	0.037	0.173	0.439	934
School enrolment rate	0.915	0.287	0	3.860	934
Distance to next railroad start	0.771	0.868	0.013	4.714	934
Firm size (log)	1.053	0.911	-2.028	5.220	922
Number of firms (log)	2.624	1.000	0	7.002	924

Note: Summary statistics for the 1849 cross section. The number of observations decreases when cities drop from the sample because they lose city rights, due to incorporations, or due to cities having implausible growth rates. Source: See main text and Appendix Appendix A for data sources and details.

Table 3: Differences in Growth for Connected and Unconnected Cities

Period	No railroad in 1848 (1)	Railroad in 1848 (2)	Difference (3)	Terminal & junction cities (4)
1816-31	0.014	0.017	0.003*	0.015
	(0.010)	(0.010)		(0.007)
1831-37	0.013	0.016	0.003*	0.016
	(0.013)	(0.014)		(0.011)
1837-40	0.016	0.017	0.002	0.018
	(0.018)	(0.019)		(0.018)
1840-43	0.014	0.017	0.003*	0.018
	(0.016)	(0.014)		(0.011)
1843-46	0.013	0.024	0.010***	0.024
	(0.015)	(0.014)		(0.014)
1846-49	0.003	0.007	0.004*	0.012
	(0.017)	(0.015)		(0.017)
1849-52	0.013	0.020	0.006***	0.025
	(0.016)	(0.017)		(0.015)
1852-55	0.003	0.012	0.009***	0.012
	(0.014)	(0.021)		(0.020)
1855-58	0.009	0.015	0.006**	0.019
	(0.017)	(0.015)		(0.023)
1858-61	0.012	0.015	0.003*	0.019
	(0.014)	(0.016)		(0.013)
1861-64	0.011	0.021	0.010***	0.025
	(0.015)	(0.021)		(0.020)
1864-67	0.003	0.014	0.012***	0.021
	(0.017)	(0.020)		(0.020)
1867-71	0.005	0.014	0.009***	0.023
	(0.016)	(0.018)		(0.016)

Note: Annual population growth rates for connected and unconnected cities. The difference between growth rates is calculated using a two-sided test. Annual population growth rates of terminal and junction cities are reported for comparison. Standard deviations in parentheses. Source: See main text and Appendix Appendix A for data sources and details.

Table 4: Railroad Lines Built by 1848

Connection	Built in (1)	Length in km (2)	Share of straight lines (3)	Passengers (4)	Freight in cwt (5)
Berlin-Stettin	1843	134	81.7%	279,768	1,302,519
Stettin-Posen	1847	205	80.0%	172,234	727,245
Berlin-Frankfurt-Breslau	1843/45	389	79.3%	632,899	1,730,987
Hansdorf-Glogau	1847	72	67.2%	108,697	204,899
Breslau-Schweidnitz-Freiburg	1844	67	80.9%	193,996	1,314,144
Breslau-Myslowitz	1843	198	79.0%	376,910	2,109,013
Brieg-Neisse	1847	44	73.7%	85,533	211,993
Kosel-Oderberg	1846	54	82.8%	76,098	338,726
Berlin-Hamburg	1846	286	83.0%	523,145	1,831,190
Magdeburg-Leipzig	1840	119	77.5%	725,495	2,294,189
Berlin-Potsdam-Magdeburg	1838/46	147	81.3%	739,608	869,727
Magdeburg-Halberstadt-Thale	1843	58	69.9%	320,215	1,627,154
Berlin-Jüterbog-Halle	1841/48	232	78.6%	330,024	1,098,306
Halle-Gerstungen	1846	165	62.3%	632,943	1,052,009
Köln-Duisburg-Minden	1846	267	83.0%	1,451,703	3,292,257
Münster-Hamm	1848	35	88.4%	134,990	120,095
Steele-Vohwinkel	1831/47	33	40.1%	116,834	1,190,570
Elberfeld-Dortmund	1848	58	53.5%	553,027	2,023,728
Düsseldorf-Elberfeld	1842	26	60.1%	331,112	1,960,077
Köln-Bonn	1844	29	71.3%	608,937	71,509
Köln-Aachen	1841	86	72.4%	514,430	6,033,504

Note: Presented data cover the year 1848. Freight is measured in Prussian hundredweights. Source: Eisenbahn-Büreau (1855)

Table 5: Railroad Access and Growth in a Cross Section of Cities

DepVar: Population growth rate	1831-37	1849-71	1849-52	1852-55	1855-58	1858-61	1861-64	1864-67	1867-71
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rail access 1838-48	0.002	0.009***	0.006***	0.008***	0.005***	0.004**	0.011***	0.010***	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Later access = 1		0.005***	0.009**	0.008**	0.002	0.002	0.003*	0.005***	0.006***
		(0.001)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Street access = 1	0.001	0.001	-0.000	0.000	-0.001	-0.001	-0.001	0.002	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Waterway access = 1	-0.002*	-0.001	0.002*	0.001	0.001	0.001	-0.001	-0.003*	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Annual city growth 1816-31	-0.071								
	(0.045)								
Annual city growth 1831-37		0.084***	0.086**	0.053	0.076*	0.035	0.116***	0.091**	0.042
		(0.022)	(0.036)	(0.040)	(0.043)	(0.029)	(0.033)	(0.035)	(0.030)
Civilian population (log)	0.001	0.000	0.001	0.002*	0.003**	0.001	0.002*	0.001	-0.000
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Military population (log)	0.000	0.000	0.000	-0.000	-0.000	0.000	-0.000	0.001**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Factory workers (share)	0.008**	0.011**	0.010*	0.011***	0.016**	0.009	0.005	0.015***	0.015***
	(0.004)	(0.005)	(0.006)	(0.004)	(0.007)	(0.007)	(0.004)	(0.004)	(0.003)
Mining (county level)	-0.002	0.007***	0.006***	0.009***	0.011***	0.000	0.008***	0.005**	0.005**
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Large farming (county level)	-0.003	0.044***	0.127***	0.065**	0.058**	0.054**	0.048**	-0.010	0.001
	(0.018)	(0.014)	(0.028)	(0.026)	(0.029)	(0.021)	(0.023)	(0.025)	(0.022)
Age composition	0.056***	0.014	-0.052**	0.014	0.037*	0.028*	-0.016	-0.025	0.047***
	(0.015)	(0.010)	(0.021)	(0.021)	(0.019)	(0.016)	(0.016)	(0.017)	(0.017)
School enrolment rate	-0.002	-0.001	0.001	0.002	-0.001	-0.001	-0.002	-0.003	-0.004
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Distance to next railroad start	-0.001	0.001***	0.001	0.000	0.002***	0.001	0.004***	0.001	-0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.012	-0.005	0.018	-0.020*	-0.030**	-0.007	-0.002	-0.003	-0.010
	(0.010)	(0.008)	(0.014)	(0.012)	(0.012)	(0.010)	(0.011)	(0.015)	(0.014)
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.04	0.21	0.09	0.10	0.10	0.04	0.13	0.14	0.13

Note: OLS estimates at the city level for different periods, using the diffusion of railroad technology in 1848. Standard errors, clustered at the county level, in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text and Appendix Appendix A for data sources and details.

Table 6: Instrumenting Railroad Access with Straight-Line Corridors

	1831-37 (1)	1849-71 (2)	1849-52 (3)	1852-55 (4)	1855-58 (5)	1858-61 (6)	1861-64 (7)	1864-67 (8)	1867-71 (9)
Panel A: First stage - Actual railroad access and location within straight-line corridor.									
Straight-line corridor = 1	0.546***	0.566***	0.560***	0.580***	0.591***	0.539***	0.523***	0.521***	0.505***
	(0.094)	(0.087)	(0.095)	(0.090)	(0.092)	(0.094)	(0.097)	(0.095)	(0.095)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.21	0.28	0.21	0.22	0.23	0.22	0.22	0.24	0.25
Panel B: Second stage - Population growth rate and actual railroad access.									
Rail access 1838-48	0.002	0.021***	0.016**	0.017***	0.019**	0.011**	0.022***	0.022**	0.021***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.005)	(0.007)	(0.009)	(0.007)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	898	906	929	924	914	926	924	919	919
Kleibergen-Paap F statistic	33.48	41.85	34.63	41.12	41.71	32.68	29.23	30.22	28.32
Panel C: Reduced form - Population growth rate and location within straight-line corridor.									
Straight-line corridor = 1	0.001	0.012***	0.009***	0.010***	0.011**	0.006**	0.011***	0.011**	0.011***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.002)	(0.004)	(0.005)	(0.004)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.04	0.20	0.09	0.09	0.10	0.04	0.12	0.13	0.12

Note: Estimations at the city level for different periods. Standard errors, clustered at the county level, in parentheses. Controls include: additional access dummy, street access dummy, waterway access dummy, annual city growth 1816-1831 (column 1), annual city growth 1831-1837 (columns 2-9), civilian population (log), military population (log), factory workers (share), mining (county level), large farming (county level), age composition, school enrollment rate, distance to next railroad start, and a constant. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text and Appendix Appendix A for data sources and details.

Table 7: Estimations for Later Adopters

DepVar: Population growth rate	1831-37	1849-71	1849-52	1852-55	1855-58	1858-61	1861-64	1864-67	1867-71
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rail access 1838-48	0.001	0.021***	0.018**	0.019***	0.021**	0.014**	0.024***	0.025***	0.021***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)	(0.010)	(0.007)
Rail access 1849-51	-0.004	0.010***	0.012**	0.016***	0.013***	0.011**	0.012***	0.013***	0.006*
	(0.004)	(0.003)	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)
Rail access 1852-54	-0.002	0.012***	0.005	0.004	-0.006	0.002	0.004	0.010*	0.013***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.004)
Rail access 1855-57	-0.007***	0.009***	0.005*	0.008**	0.009*	0.008***	0.010**	0.011***	0.007***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)
Rail access 1858-60	-0.002	0.006***	0.002	0.004	0.017***	0.001	0.004	0.007*	0.008**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Rail access 1861-63	-0.007**	0.004**	0.003	0.001	0.008*	0.009***	0.004	0.002	0.002
	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.002)
Rail access 1864-66	-0.001	0.007***	0.001	0.005**	0.003	0.005**	0.009***	0.013***	0.008**
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
Rail access 1867-70	0.001	0.006***	0.005	0.002	0.004*	0.004*	0.006**	0.013***	0.011***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	898	906	929	924	914	926	924	919	919
Kleibergen-Paap F statistic	30.14	42.21	30.94	37.09	40.50	28.72	26.12	27.65	28.29

Note: Estimations at the city level for different periods. Rail access 1838-48 is instrumented by straight-line corridors. Standard errors, clustered at the county level, in parentheses. Controls include: street access dummy, waterway access dummy, annual city growth 1816-1831 (column 1), annual city growth 1831-1837 (columns 2-9), civilian population (log), military population (log), factory workers (share), mining (county level), large farming (county level), age composition, school enrollment rate, distance to next railroad start, and a constant. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text and Appendix Appendix A for data sources and details.

Table 8: Estimations in Constraint Samples

DepVar: Population growth rate	1831-37 (1)	1849-71 (2)	1849-52 (3)	1852-55 (4)	1855-58 (5)	1858-61 (6)	1861-64 (7)	1864-67 (8)	1867-71 (9)
Panel A: IV estimates using a sample of cities smaller than 3000 inhabitants in 1837.									
Rail access 1838-48	-0.006	0.012***	0.011	0.010*	0.010*	0.014**	0.026***	0.017*	0.013*
	(0.007)	(0.004)	(0.007)	(0.006)	(0.006)	(0.006)	(0.009)	(0.010)	(0.007)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	632	641	659	657	645	657	655	649	649
Kleibergen-Paap F statistic	19.12	24.76	18.28	22.30	26.24	16.41	15.85	15.57	15.19
Panel B: IV estimates using a sample and weights obtained from geographical matching.									
Rail access 1838-48	-0.000	0.022***	0.016*	0.017**	0.023**	0.016**	0.029***	0.013	0.020**
	(0.008)	(0.008)	(0.009)	(0.008)	(0.012)	(0.007)	(0.010)	(0.015)	(0.009)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	267	267	281	277	271	278	277	271	270
Kleibergen-Paap F statistic	19.65	21.77	16.76	20.04	19.94	15.88	14.30	18.20	13.07
Panel C: IV estimates using a sample and weights obtained from radius propensity score matching on pre-railroad controls.									
Rail access 1838-48	-0.003	0.011**	0.015*	0.013	0.025**	0.023**	0.022*	0.023	0.031**
	(0.011)	(0.005)	(0.009)	(0.012)	(0.012)	(0.010)	(0.013)	(0.017)	(0.014)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	314	209	331	344	296	273	263	215	193
Kleibergen-Paap F statistic	29.44	16.56	51.87	59.12	37.34	26.95	43.98	23.36	25.67
Panel D: IV estimates using a sample and weights obtained from kernel propensity score matching on pre-railroad controls.									
Rail access 1838-48	-0.001	0.017**	0.016**	0.011	0.021**	0.009	0.022**	0.022	0.023***
	(0.010)	(0.008)	(0.008)	(0.009)	(0.010)	(0.007)	(0.011)	(0.016)	(0.009)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	810	623	803	785	759	724	705	677	619
Kleibergen-Paap F statistic	61.85	76.13	78.04	83.85	83.39	73.87	78.29	72.23	59.92

Note: Estimations at the city level for different periods. Rail access 1838-48 is instrumented by straight-line corridors. Standard errors, clustered at the county level, in parentheses. Controls include: additional access dummy, street access dummy, waterway access dummy, annual city growth 1816-1831 (column 1), annual city growth 1831-1837 (columns 2-9), civilian population (log), military population (log), factory workers (share), mining (county level), large farming (county level), age composition, school enrollment rate, distance to next railroad start, and a constant. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text, Appendices Appendix A and Appendix B for data sources and details.

Table 9: The Impact of Railroad Access: Panel Estimates

DepVar: (ln)Population	Pooled (1)	FE (2)	CFE (3)	Non-linear (4)	3000 (5)	FS (6)	SS (7)
Railroad access = 1	0.053***	0.047***	0.060***				0.077***
	(0.013)	(0.013)	(0.018)				(0.024)
Potential railroad diffusion						0.498***	
						(0.019)	
Access for 1 period				0.028***	0.035**		
				(0.011)	(0.018)		
Access for 2 periods				0.036***	0.040**		
				(0.013)	(0.020)		
Access for 3 periods				0.059***	0.076***		
				(0.019)	(0.029)		
Access for 4 periods				0.055**	0.073**		
				(0.022)	(0.029)		
Access for 5 periods				0.082***	0.093**		
				(0.027)	(0.042)		
Access for 6 periods				0.103***	0.107**		
				(0.030)	(0.045)		
Access for 7 periods				0.121***	0.089***		
				(0.033)	(0.032)		
Access for 8 periods				0.184***	0.129***		
				(0.046)	(0.044)		
Access for 9 periods				0.103***	0.099***		
				(0.022)	(0.029)		
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
City fixed effects	N	Y	Y	Y	Y	Y	Y
County × year fixed effects	N	N	Y	N	N	N	N
Observations	7,680	7,680	7,680	7,680	5,715	7,680	7,680
R-squared		0.38	0.74	0.38	0.34	0.18	
Number of Cities	862	862	862	862	640	862	862

Note: Panel estimates at the city-year level 1840-1864. Railroad access indicates if a city had access to the railroad network in a previous year. Column FE introduces county fixed effects, column CFE introduces a full set of interactions of county fixed effects with time period fixed effects, column Non-linear introduces dummies for periods of access, column 3000 excludes all cities larger than 3,000 inhabitants before 1838. Columns FS and SS indicate first-stage and second-stage estimates, instrumenting actual railroad access with straight-line corridors. Further controls: military population (log) and a dummy for incorporations. Robust standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text and Appendix Appendix A for data sources and details.

Table 10: Railroads, Firm Size and the Number of Firms

DepVar:	Firm size (log)		Number of firms (log)	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Rail access 1838-48	0.299***	0.734***	-0.158**	-0.188
	(0.092)	(0.271)	(0.068)	(0.127)
Controls	yes	yes	yes	yes
Observations	922	922	924	924
R-squared	0.56		0.59	
Kleibergen-Paap F statistic		31.02		31.02

Note: Estimations at the city level for different industries. Column 2 and 4 show second-stage results, instrumenting Rail access 1838-48 by straight-line corridors. Standard errors, clustered at the county level, in parentheses. Controls include: additional access until 1871 dummy, street access dummy, waterway access dummy, annual city growth 1831-1837, civilian population (log), military population (log), mining (county level), large farming (county level), age composition, school enrollment rate, distance to next railroad start, and a constant. Significance: *** p<0.01, ** p<0.05, * p<0.1. Source: See main text and Appendix Appendix A for data sources and details.

Appendix A Definition and Sources of Control Variables

(i) *Street access*: Dummy variable indicating whether a city was connected to a main road. Similar to the maps on railroad access, we geo-reference the corresponding map for paved and unpaved main roads (*Hauptstraßen*) in 1848 and match it with the location of Prussian cities.

(ii) *Waterway access*: Dummy variable indicating whether a city has at least one cargo ship for river navigation or one seagoing vessel in 1849.

(iii) *Annual city growth 1831-37*: Measured as the average annual growth of the civilian population as counted in the censuses of 1831 and 1837.

(iv) *Civilian population (log)*: Measured as the natural logarithm of the resident civilian population in 1849.

(v) *Military population (log)*: Measured as the natural logarithm of the military population in 1849.

(vi) *Factory workers (share)*: Measured as the share of total population employed in factories of all kinds in 1849.

(vii) *Mining (county level)*: Dummy variable indicating whether the city is located in a county that has a least one steam engine in mining.

(viii) *Large farming (county level)*: Measured as the county-level share of land holdings larger than 300 Prussian Morgen (roughly 75 hectare) over the total number of land holdings in 1849.

(ix) *Age composition*: Measured as the share of the population younger than 15 years over the total population in 1849.

(x) *School enrollment rate*: Measured as the share of children at compulsory school age (6-14) that attended school in 1849.

(xi) *Distance to next railroad start*: Measured as the linear distance to the closest terminal or junction station of railroad lines in 100 kilometers in 1848.

(xii) *incorporations*: Dummy variable indicating whether a city changed its dimension through incorporation of surrounding parishes.

Control variable (i) is coded using maps provided by IEG (2010), variables (ii) and (iv) to (viii) are digitized data from the 1849 census (Statistisches Bureau zu Berlin, 1851-1855), and variables (iii) and (xii) are from data provided by Matzerath (1985).

Appendix B Definition and Sources of Matching Variables

(i) *City size 1837*: Measured as the natural logarithm of the total number of civilian inhabitants in 1837.

(ii) *Annual city growth 1821-37*: Measured as the average annual growth of the civilian population as counted in the censuses of 1821 and 1837.

(iii) *Merchants*: Measured as the share of merchants, hawkers and victual mongers in the total population in 1819.

(iv) *Looms*: Measured as the number of looms on different fabrics over the total population in 1819.

(v) *Protestants*: Measured as the share of the Protestant population in 1816.

(vi) *Private dwellings*: Measure as the number of private dwellings over the total population in 1821.

(vii) *Commercial buildings*: Measured as the number of manufactories, mills and warehouses over the total population in 1821.

(viii) *Insurance-value of buildings against fire*: Measured as the natural logarithm of the average value of buildings insured by the local fire insurance company (*Feuersocietät*) in 1821.

Matching variables (i) and (ii) are calculated using the data provided by Matzerath (1985), variables (iii) to (viii) are digitized data from the 1816-1821 censuses (Mützell, 1823-1825).