RAILROADS AND GROWTH IN PRUSSIA

Erik Hornung
Max Planck Institute for Tax Law and Public Finance

Abstract
We study the effect of railroad access on urban population growth. Using GIS techniques, we match triennial population data for roughly 1,000 cities in 19th-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 propensity score matching, instrumental variables, and fixed-effects estimation techniques. Our instrument identifies exogenous variation in railroad access by constructing straight-line corridors between nodes. Counterfactual models using pre-railroad growth yield no evidence to support the hypothesis that railroads appeared as a consequence of a previous growth spurt. (JEL: O18, O33, N73)

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 19th century, have been repeatedly discussed as being *the* technology that shaped growth during the Industrial Revolution (with seminal work by Fishlow 1965, Fogel 1962, and Rostow 1962). Using the concept of social savings, the effect of railroads on aggregate growth has been comparatively calculated for many countries that were early adopters of railroad technology and ranges from 4% to 25% of the GNP, depending on the country and the period under consideration (see O'Brien 1983).[1]

The editor in charge of this paper was Fabrizio Zilibotti.

Acknowledgments: An earlier version of this paper is part of my PhD thesis which was prepared at the Ifo Institute, Munich. I thank Sascha O. Becker, Francesco Cinnirella, Oliver Falck, Guido Schwerdt, Ludger Woessmann, three anonymous referees, and the editor, Fabrizio Zilibotti, for helpful suggestions. Comments received from seminar participants at Ifo Munich, University of Munich, and University of Mannheim are gratefully acknowledged. I also gratefully acknowledge 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.

E-mail: erik.hornung@tax.mpg.de

1. 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

Journal of the European Economic Association August 2015 13(4):699–736

DOI: 10.1111/jeea.12123

In addition to its macroeconomic effect, technology adoption can be crucial in generating localized comparative advantages and regional economic growth. However, a major challenge in the literature is to establish causality in this relationship. Recent work by Atack et al. (2010) and Banerjee, Duflo, and Qian (2012) attempts 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? The overall results still seem to be ambiguous.[2]

This paper sheds light on the direction of causality between railroad adoption and economic growth in a range of approaches using highly detailed city-level data from the historical German state of Prussia. Using an extensive data set for all 978 Prussian cities, we provide evidence that access to this new technology massively influenced city growth rates—a widely used proxy for economic growth.[3] Following the notion that "city sizes grow with improvements in technology" (Henderson 2005, p. 1577), we estimate that railroad technology induced an additional annual growth ranging between 1 and 2 percentage points for adopting cities compared to nonadopting cities. The size of this effect remains very stable across a range of different periods and specifications. Estimating counterfactual models of railroad access on growth prior to access yields no evidence of a reversed causality.

A recently growing literature analyzes the effects of transport infrastructure on a range of outcomes. Authors address the consequences of establishing railroad systems for a number of economies from very different points of view.[4] Research analyzing aspects of market integration, increasing trade flows, and price convergence usually finds evidence of large gains from increasing trade due to railroad network expansion (see Donaldson 2014; Donaldson and Hornbeck 2013; Keller and Shiue 2013, 2014). Focusing on aspects of development and growth such as urbanization, per capita income growth, and income inequality yields mixed results in terms of size or significance of the effects (see Atack et al. 2010; Atack, Haines, and Margo 2011; Banerjee, Duflo, and Qian 2012). This recent literature predominantly focuses on analyzing the consequences of establishing railroads for the agricultural sector or for agricultural societies with limited factor mobility. As railroads are often strongly connected to the

(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.

2. There is some consensus in the literature on German railroads that the latter is most likely (Hahn 2005, p. 26; Fremdling 1983, p. 122). However, the question of whether regions grew comparatively faster after they gained access to the railroad has not been answered conclusively (Matzerath 1996, p. XI).

3. An expanding body of literature examines the effects of the diffusion of historical innovations on growth—proxied by urban population growth. Such studies use the geographic distribution of an important cultural or technological innovation and analyze its effects on local economic growth. These studies analyze the diffusion of banking in the United States (Bodenhorn and Cuberes 2010), the diffusion of Protestantism in Germany (Cantoni 2014), the diffusion of the printing press in Europe (Dittmar 2011), and the diffusion of potato cultivation in Europe (Nunn and Qian 2011).

4. This literature is also closely related to research in urban economics which analyzes the effects of interstate highways on outcomes such as suburbanization, the composition of industrial activity, or the demand for skill (see Alder 2014; Baum-Snow 2007; Duranton and Turner 2012; Faber 2014; Michaels 2008).

industrial sector, the Prussian environment seems a natural laboratory to assess the consequence of railroads on industrial development.

Different from the recent literature, this paper analyzes a period of strong industrial development during which railroad infrastructure was provided by the private sector. Such circumstances further complicate the identification of causal effects, as the assignment of railroad access to a location is likely to be endogenous. This paper contributes to the literature on transport infrastructure by further refining identification strategies in order to establish causality. Our highly detailed data allow us to combine matching techniques with instrumental variable methods as well as introducing a time-varying instrument to estimate the causal effect of railroad access on growth. We will further discuss our findings in comparison to the literature toward the end of this paper.

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 mid-1860s.[5]

The paper is structured to gradually build up specifications from a cross-sectional approach using ordinary least squares (OLS), instrumental variables (IV), and propensity score matching (PSM) to a fixed-effects panel approach using OLS and IV, thus reflecting the hierarchy between the different specifications. These approaches successively address issues of endogeneity and unobserved heterogeneity to estimate the causal effect of railroad access on growth.

The IV approach rests on the assumption that until the mid-1860s, Prussian railroads were built to connect important cities.[6] Since construction costs were high, lines were mostly built linearly. 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 straight 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 (nodes), we can construct a variable indicating the potential for railroad adoption—being located within a narrow straight-line corridor—that we use to instrument actual railroad access.[7]

We further apply matching techniques to account for city-level heterogeneity in pre-railroad development. This allows for estimations in samples of cities that are highly comparable and ideally differ only in their access to a railroad line. The previous findings are confirmed when applying our instrumental-variable estimation strategy to these matched samples.

5. This corresponds to the prevailing periodization of the German railroad system based on Sombart (1921, p. 239).

6. See Sections 2 and 4.2 for more information.

7. The use of straight-line instrumental variables is well established in the literature on transportation infrastructure and was most prominently started in Michaels (2008).

The instrumental variable proves to be powerful for cross-sectional as well as for fixed-effects panel estimations. Whenever new railroad lines are built, new straight lines can be drawn between nodes, effectively creating exogenous variation across cities as well as over time. Both approaches return significant positive effects of railroad access over a range of different periods. As such, this paper seems to be the first successful attempt to develop a time-varying straight-line instrument to allow for causal inferences regarding the effects of transportation infrastructure using panel data with fixed effects.

Finally, we use the fixed-effects panel setting to estimate event study specifications. Results from such specifications credibly show the absence of differences in pre-railroad growth trends between railroad and nonrailroad cities as well as a sharp upward trend after railroad access has been established.

There may be many, nonexclusive channels through which railroads might affect the economy and our data allow us to shed light on some of them. Atack, Haines, and Margo (2011) argue that railroads increase competition among firms due to their role in increasing market size. Consequently, firms attempt to increase productivity through the division of labor which in turn leads to an increase in establishment size. A part of this hypothesis can actually be tested using Prussian city-level factory data. We find that the average firm size is larger in cities that are connected to the railroad network than in unconnected cities. Furthermore, we do not find evidence that railroads increase the number of factories located in a city. Thus railroads seem to have a causal effect on industrial development at the intensive margin in the form of increasing returns to scale rather than at the extensive margin. Additional results suggest that railroads induce population growth by increasing migration to urban centers while fertility remains unchanged.

The remainder of the paper is structured as follows. Section 2 provides the historical background of the railroad network expansion and urbanization patterns in Prussia. Section 3 introduces and describes the data used for the empirical analysis. Section 4 addresses endogeneity issues and presents results from cross-sectional and fixed-effects panel data using OLS, PSM, and IV estimation techniques, building up to our preferred specifications. Section 5 fits our findings into the recent literature and discusses possible mechanisms and remaining issues. Section 6 concludes.

2. Patterns of Railroad Network Expansion and Urbanization

At the beginning of the 19th century, Germany had an inadequate transportation network when 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 as early as in 1833 (List 1833). List's blueprint for the railroad system connects all major cities throughout Germany. The simultaneous founding of the Zollverein (German Customs Union) led to increasing trade between the many states and fiefdoms and the pan-German transport network expansion became desirable (Keller and Shiue 2013, 2014).[8]

Due to 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 prior to 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).[9]

The government's decision, due to a lack of funds, not to directly construct a railroad network, but to approve and license private railroad enterprises, meant that railroad construction in Prussia lacked a central plan (Fremdling and Knieps 1997, p. 137), but was built according to the expected profitability of the 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).

Such a pattern can also be observed in a periodization of the German railroad network expansion following Sombart (1921, p. 239):[10] (1) preliminary stage until 1845—connecting the 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—completion of a coarse network; (4) ramification until 1913—railroad supply for smaller towns through branch lines.

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.

8. Prussia abolished internal customs barriers and tariffs in 1818 and initiated the Zollverein that covered most parts of Germany by 1834. Consequently, Prussia could trade freely with most of the German states during the period under analysis in the paper. However, the most direct connection between the eastern and western parts of Prussia runs through the Duchy of Brunswick and the Kingdom of Hanover. Brunswick entered into the Zollverein in 1841 while Hanover entered only in 1854. After establishing a railroad line from Minden to Mageburg, Hanover agreed to impose a low transit rate for goods passing through from one part of Prussia to the other. Passengers, mail, and money were able to transit free of duties.

9. Online Appendix B provides additional information on the expansion of the Prussian railroad network until 1880.

10. This periodization is still used today—for example, in Henning (1995, p. 162).

3. The Data

The variable of interest in our analysis indicates whether a city was connected to the railroad in a given year (the treatment). We use GIS software to collect information on railroad diffusion using maps provided by IEG (2010).[11] The resulting binary variable takes the value 1 if one or more railroad lines intersect the city in a given year. We correct our data using information on actual railroad access from the German handbook of cities (Keyser 1939–1974), which specifies the year in which access was established and indicates the corresponding connection. This information is then checked and verified with information from official Prussian sources (Königlich Preussisches Statistisches Bureau 1883).

In Table 1 we present information on all railroad lines established by 1848, the relevant year for our cross-sectional analysis. Twenty-one railroad lines were built in Prussia during the period 1838–1848. We provide information on the year of construction, as well as their length, passenger, and freight transport statistics for each of the lines during the year 1848. The list of railroad lines also yields one important information—the terminal stations of the lines. These nodes perform a crucial role in a network because they are locations that were chosen to be connected in the first place. Whenever a new railroad line is built, it had originally been planned to connect two or more locations. These locations are obviously not chosen arbitrarily and have higher-level functions than other locations along the line.

This paper assigns the role of a node in the railroad network to two types of cities: terminals and junctions. A city is identified as a node if it is mentioned in the name of the line—see column (1) of Table 1. For example, Berlin and Frankfurt (Oder) are identified as nodes after the Berlin–Frankfurter Eisenbahn (Berlin–Frankfurt railroad) was opened in 1842. In 1845, the line Berlin–Breslau was established and included a section that was actually the Berlin–Frankfurt line. Breslau is thus identified as a node after 1845, while Frankfurt keeps its original node status. Furthermore, the line Berlin–Breslau was built to take a detour to provide a connection to the Görlitz–Dresden railroad line connecting Prussia to the Kingdom of Saxony. The junction to this line was chosen to be at the small village of Kohlfurt, close to the border. Kohlfurt is thus identified as a junction node.[12]

Generalizing from urban population growth to economic growth has shown to be an acceptable approximation in cases where data on income are unavailable (Acemoglu, Johnson, and Robinson 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 light of the fact that urban centers were the places

11. 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 GIS 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.

12. For further information regarding city size and node status see Online Appendix C.

TABLE 1. Railroad lines built by 1848.

Connection (1)	Year built (2)	Length in km (3)	No. of passengers (4)	Freight in cwt (5)	Share of straight lines (6)
Berlin–Stettin	1843	134	279,768	1,302,519	81.7%
Stettin–Posen	1847	205	172,234	727,245	80.0%
Berlin–Frankfurt–Breslau	1843/45	389	632,899	1,730,987	79.3%
Hansdorf–Glogau	1847	72	108,697	204,899	67.2%
Breslau–Schweidnitz–Freiburg	1844	67	193,996	1,314,144	80.9%
Breslau–Myslowitz	1843	198	376,910	2,109,013	79.0%
Brieg–Neisse	1847	44	85,533	211,993	73.7%
Kosel–Oderberg	1846	54	76,098	338,726	82.8%
Berlin–Hamburg	1846	286	523,145	1,831,190	83.0%
Magdeburg–Leipzig	1840	119	725,495	2,294,189	77.5%
Berlin–Potsdam–Magdeburg	1838/46	147	739,608	869,727	81.3%
Magdeburg–Halberstadt–Thale	1843	58	320,215	1,627,154	69.9%
Berlin–Jüterbog–Halle	1841/48	232	330,024	1,098,306	78.6%
Halle–Gerstungen	1846	165	632,943	1,052,009	62.3%
Köln–Minden	1846	267	1,451,703	3,292,257	83.0%
Münster–Hamm	1848	35	134,990	120,095	88.4%
Steele–Vohwinkel	1831/47	33	116,834	1,190,570	40.1%
Elberfeld–Dortmund	1848	58	553,027	2,023,728	53.5%
Düsseldorf–Elberfeld	1842	26	331,112	1,960,077	60.1%
Köln–Bonn	1844	29	608,937	71,509	71.3%
Köln–Aachen	1841	86	514,430	6,033,504	72.4%

Notes: Presented data cover the year 1848. Freight is measured in Prussian hundredweights. The "share of straight lines" measure is adopted from official Prussian records.
Source: Technisches Eisenbahn-Büreau (1855).

where most of the innovation, as well as human and physical capital, was located and accumulated.

For our dependent variable, we use triannual city-level population data provided by Matzerath (1985), originally published by the Prussian Statistical Office. We corrected some errors and included additional years using the original published sources. 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 census years. To achieve a balanced sample, we restrict our data to the 978 cities that held city rights in 1849.[13]

In Tables 2–4 we provide descriptive summary statistics by treatment status for the 1849 cross-section. The treatment group (column (2)) consists of cities that gained railroad access during the period 1838 to 1848; the control group (column (3)) consists of cities that had no access by 1848.[14] In column (4), we compare variable means for railroad cities with nonrailroad cities. Table 2 reports annual growth rates for a range

13. For further information see Online Appendix D.

14. The control group however, includes cities that subsequently gained access in the period 1849–1871.

TABLE 2. Annual population growth rates by railroad access status in 1848.

	Nodes		Railroad cities excluding nodes		Nonrailroad cities		Difference in means between
	Obs.	Mean	Obs.	Mean	Obs.	Mean	(2) and (3)
	(1)		(2)		(3)		(4)
1816–21	30	0.017	67	0.020	824	0.019	0.001
		(0.014)		(0.018)		(0.021)	(0.003)
1821–31	29	0.015	68	0.014	840	0.011	0.002
		(0.010)		(0.010)		(0.013)	(0.002)
1831–37	29	0.017	76	0.015	854	0.013	0.002
		(0.012)		(0.013)		(0.013)	(0.002)
1837–40	29	0.019	76	0.018	860	0.016	0.002
		(0.019)		(0.019)		(0.018)	(0.002)
1840–43	29	0.018	75	0.017	861	0.014	0.003
		(0.010)		(0.014)		(0.016)	(0.002)
1843–46	28	0.023	75	0.023	867	0.013	0.010***
		(0.014)		(0.013)		(0.015)	(0.002)
1846–49	30	0.011	75	0.007	870	0.003	0.004**
		(0.015)		(0.015)		(0.017)	(0.002)
1849–52	29	0.026	76	0.019	867	0.013	0.006***
		(0.015)		(0.016)		(0.016)	(0.002)
1852–55	30	0.013	75	0.012	862	0.003	0.009***
		(0.020)		(0.021)		(0.014)	(0.002)
1855–58	27	0.019	72	0.015	856	0.009	0.006***
		(0.024)		(0.015)		(0.017)	(0.002)
1858–61	30	0.019	74	0.015	867	0.012	0.003*
		(0.013)		(0.016)		(0.014)	(0.002)
1861–64	29	0.024	74	0.021	863	0.011	0.010***
		(0.017)		(0.021)		(0.015)	(0.002)
1864–67	29	0.020	75	0.014	857	0.003	0.012***
		(0.017)		(0.020)		(0.017)	(0.002)
1867–71	30	0.022	75	0.014	858	0.005	0.010***
		(0.016)		(0.018)		(0.016)	(0.002)

Notes: Summary statistics for the 1849 cross-section by treatment status. The number of observations changes due to missing information in the period before 1849 or due to the exclusion of outliers in the annual population growth rates. The "difference between means" is calculated using a two-sided test. Standard deviations in parentheses (columns (1)–(3)). Standard errors in parentheses (column (4)).
***Significant at 1%; **significant at 5%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

of periods from 1816 to 1871, Table 3 reports descriptives for controls variables that will be included in the cross-sectional analysis, and Table 4 reports descriptives for the pre-railroad adoption period that will be used for a matching approach.

The descriptives presented in Table 2 allow a comparison of population growth rates between the treatment and control group as well as pre- and post-treatment trends. Column (4) shows that pre-treatment differences in growth rates between treatment and control group were generally quite small and insignificant for most

TABLE 3. Control variables (post-railroad) by access status in 1848.

	Nodes		Railroad cities excluding nodes		Nonrailroad cities		Difference in means between
	Obs.	Mean (1)	Obs.	Mean (2)	Obs.	Mean (3)	(2) and (3) (4)
Straight-line corridor = 1	30	1.000	76	0.250	871	0.011	0.239***
		(0.000)		(0.436)		(0.107)	(0.019)
Street access = 1	30	0.933	76	0.592	871	0.393	0.199***
		(0.254)		(0.495)		(0.489)	(0.059)
Waterway access = 1	30	0.567	76	0.276	871	0.191	0.086*
		(0.504)		(0.450)		(0.393)	(0.048)
Civilian population (log)	30	9.787	76	8.357	871	7.793	0.564***
		(1.144)		(0.777)		(0.692)	(0.084)
Military population (log)	30	6.379	76	3.159	871	2.265	0.894***
		(2.700)		(2.400)		(1.944)	(0.237)
Factory workers (share)	30	0.069	76	0.060	871	0.038	0.022
		(0.097)		(0.088)		(0.120)	(0.014)
Mining (county level)	30	0.300	76	0.197	871	0.095	0.102***
		(0.466)		(0.401)		(0.294)	(0.036)
Large farming (county level)	30	0.015	76	0.018	871	0.027	−0.009***
		(0.013)		(0.013)		(0.026)	(0.003)
Age composition	30	0.290	76	0.334	871	0.343	−0.009**
		(0.041)		(0.033)		(0.037)	(0.004)
School enrollment rate	30	0.790	76	0.928	871	0.916	0.013
		(0.151)		(0.177)		(0.294)	(0.034)
Distance to node	30	0.000	76	0.281	871	0.803	−0.523***
		(0.000)		(0.198)		(0.887)	(0.102)

Notes: Summary statistics for the 1849 cross-section by treatment status. The "difference between means" is calculated using a two-sided test. Standard deviations in parentheses (columns (1)–(3)). Standard errors in parentheses (column (4)).
***Significant at 1%; **significant at 5%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

periods. Furthermore, we find a strong divergence in growth rates between the groups for post-treatment periods starting from the period 1843–1846.

We also present the growth rates of node cities in column (1) of Table 2. These cities will, in most cases, be excluded from our subsequent analysis. Growth rates of node cities behave similarly to those of the treatment group at first, but tend to be slightly higher, on average, after the period 1846–1849.

Table 3 presents the various control variables[15] including access to rivaling infrastructure such as main roads and navigable rivers and ports. Indicators of urbanization include pre-railroad city growth 1831–1837 and the size of the civilian and military population in 1849. Indicators of industrial development include the share of citizens employed in factories and the county-level occurrence of mining activity. As

15. Unless otherwise specified, all data refer to the base year 1849. See Appendix A for more specific definitions and sources.

TABLE 4. Matching variables (pre-railroad) by access status in 1848.

	Nodes		Railroad cities excluding nodes		Nonrailroad cities		Difference in means between
	Obs.	Mean (1)	Obs.	Mean (2)	Obs.	Mean (3)	(2) and (3) (4)
City size 1837 (log)	29	9.631	76	8.168	861	7.665	0.502***
		(1.093)		(0.769)		(0.683)	(0.083)
Annual city growth 1821-37	29	0.016	68	0.015	836	0.012	0.003**
		(0.009)		(0.006)		(0.010)	(0.001)
Merchants 1819 (p.c.)	28	0.016	62	0.016	793	0.015	0.002
		(0.006)		(0.011)		(0.010)	(0.001)
Looms 1819 (p.c.)	28	0.013	62	0.017	797	0.019	−0.002
		(0.020)		(0.019)		(0.024)	(0.003)
Protestants 1816 (share)	28	0.608	65	0.786	828	0.626	0.160***
		(0.375)		(0.326)		(0.400)	(0.051)
Private dwellings 1821 (p.c.)	29	0.105	67	0.130	845	0.138	−0.009**
		(0.042)		(0.027)		(0.033)	(0.004)
Commercial buildings 1821 (p.c.)	29	0.005	67	0.005	844	0.005	0.000
		(0.007)		(0.005)		(0.006)	(0.001)
Insurance value of buildings against fire 1821 (log)	29	6.338	67	5.806	819	5.385	0.421***
		(0.844)		(0.640)		(0.710)	(0.090)

Notes: Summary statistics for the 1849 cross-section by treatment status. The "difference between means" is calculated using a two-sided test. Standard deviations in parentheses (columns (1)–(3)). Standard errors in parentheses (column (4)).
***Significant at 1%; **significant at 5%.
Source: See main text and Appendix A for data sources and details.

geographical endowments are usually among the major determinants of city growth, we control for the county-level concentration of large landholdings. As shown by Cinnirella and Hornung (2013), the concentration of large landholdings is correlated with soil quality and can thus be viewed as a proxy for geographical endowments and therefore the supply of food for urban markets.[16] Further controls include the age composition and the primary education of the urban population.[17] These controls are aimed at capturing differences in future population growth as well as the city's progressiveness. We also calculate and control for the distance to the closest node of railroad lines since nearby cities are more likely to become connected to the network.

The incorporation of suburbs and smaller municipalities, as well as mergers between cities, sometimes introduce jumps in the data and in some cases population appears to grow erratically. We exclude cities for those periods in which incorporations

16. Additionally controlling for soil texture, a proxy for geographical endowments does not change the results (see Online Appendix G, Table G.1, panel E).

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

took place using the data set provided by Matzerath (1985), which also indicates whether a city changed dimensions in a given period.[18]

Comparing cities by treatment status, we find substantial heterogeneity in many of the control variables. However, it is not clear if these differences arise due to the existing railroad access or if they were predetermined. Our subsequent analysis will gradually build toward eliminating issues arising due to these differences.

Table 4 presents pre-treatment variables which will subsequently be used for a matching exercise.[19] Comparing the treatment and control group, we find that cities systematically differ in some aspects such as size but were highly comparable in many other aspects such as commercial development previous to any railroad construction. We will discuss these aspects in more detail in Section 4.3.

4. The Effect of Railroad Access on City Growth

This section will analyze the relationship between railroads and growth. To address concerns regarding endogeneity, we take several different econometric approaches lending evidence for causal effects of railroad access on urban population growth. The section is structured to gradually build up specifications—from a cross-sectional approach using OLS and IV estimations and combining IV estimation with PSM techniques, to a fixed-effects panel approach using OLS and IV—reflecting a hierarchy between the different specifications.

4.1. Cross-sectional OLS Estimates

In a first step, we estimate the effect of railroad access on urban growth in a standard cross-city growth regression. 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 the short- and long-term effects. This results in a model in which the urban population growth rate *PGR*[20] in a variety of periods t is a function of railroad access *RA* in 1848 and other explanatory factors X,

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

18. The exclusion of incorporations explains the varying number of observations over the subperiods. Unfortunately, after excluding these cases we sometimes still observe implausible jumps in the population accounts that might be due to unobserved incorporations or similar artificial changes in the census population. After careful inspection of the data, our sanity check finds that growth rates that exceed minus or plus 10% are hardly due to natural changes in the population. Thus we decided to exclude such observations from future estimations in subperiods that surpass this threshold. For results using the full sample of cities including nodes as well as outliers please refer to Online Appendix G, Table G.1, panels A and B. This table also shows results excluding outliers according to a standardized residuals threshold (see panel C). Point estimates in these models are similar in magnitude to our baseline specification.

19. See Appendix A for more specific definitions and sources.

20. The urban population growth rate is defined as $(\ln(POP_{t2}) - \ln(POP_{t1}))/(t_2 - t_1)$.

We emphasize here that the explanatory factors X include a lagged dependent variable to account for the dynamic aspects of urban growth.[21]

As previously mentioned, 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 growth and railroad access. Reverse causality, unobserved heterogeneity, and omitted variable bias could be serious issues in this setting.

We start addressing endogeneity concerns by excluding 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 nodes of the railroad network. Since, up until the 1860s, railroads were built to connect important cities, nodes 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 (see Section 2).

Each column of panel A in Table 5 reports OLS estimates of urban population growth on railroad access for different periods between 1831 and 1871—excluding the nodes.[22] We find that being connected to the railroad in 1848 significantly increased the annual population growth by 0.9 percentage points during the period 1849–1871 (column (2)). Comparing all periods across columns (3)–(9), we find that the annual population growth generated by railroad access varies between 0.4 and 1.1 percentage points. The coefficient seems to stabilize in the later periods under consideration, which hints at long-term effects from railroad access.

Note that the counterfactual specification in column (1) yields no significant effect when we regress railroad access until 1848 on pre-rail population growth 1831–1837. Prior to the advent of the rail, cities that were connected by 1848 thus had very similar growth patterns compared to those that were not. We find no pre-trend in rail access that favored cities with high growth rates.[23] Similarly, we find no effect on previous growth for railroad lines established in the period 1872–85 (for more information on this placebo test see Online Appendix F).

4.2. Cross-sectional IV Estimates

OLS estimates of the relationship might be biased in cases of omitted variables. Thus, we use an IV approach to resolve the omitted variable concern. Similar to the approaches taken by Atack et al. (2010) and Banerjee, Duflo, and Qian

21. However, specifications not including the lagged dependent variable will not yield substantially different results (see Online Appendix G, Table G.1, panel D).

22. Online Appendix E shows results when adding control variables one after the other and coefficients for the full set of control variables for all periods.

23. We find similar results when extending the period to 1821–1837 in all our specifications (available from the author upon request). For better comparability, we show the period 1831–1837 since more observations are missing for 1821.

TABLE 5. The impact of railroad access, cross-sectional estimates.

	Main periods		Subperiods						
	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: OLS; population growth rate and actual railroad access (DepVar: population growth rate)									
Rail access 1838–48	0.002	0.009***	0.006***	0.008***	0.006***	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)
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.04	0.22	0.09	0.09	0.10	0.04	0.13	0.14	0.14
Panel B: IV first stage; actual railroad access and location within straight-line corridor (DepVar: rail access 1838–48)									
Straight-line corridor = 1	0.531***	0.553***	0.547***	0.567***	0.578***	0.525***	0.508***	0.507***	0.491***
	(0.096)	(0.090)	(0.097)	(0.093)	(0.094)	(0.096)	(0.099)	(0.097)	(0.097)
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.20	0.27	0.20	0.21	0.22	0.22	0.21	0.23	0.25
Panel C: IV second stage; population growth rate and actual railroad access (DepVar: population growth rate)									
Rail access 1838–48	0.000	0.021***	0.015**	0.017***	0.020**	0.011**	0.021***	0.021**	0.022***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.005)	(0.007)	(0.010)	(0.007)
Observations	898	906	929	924	914	926	924	919	919
Kleibergen–Paap *F*-statistic	30.39	38.18	31.57	37.41	37.78	29.74	26.46	27.46	25.71
Panel D: reduced form; population growth rate and location within straight-line corridor (DepVar: population growth rate)									
Straight-line corridor = 1	0.000	0.012***	0.008***	0.010***	0.012**	0.006**	0.011***	0.011**	0.011***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)	(0.004)	(0.005)	(0.004)
Observations	898	906	929	924	914	926	924	919	919
R-squared	0.04	0.20	0.09	0.08	0.10	0.04	0.11	0.13	0.12

Notes: The table shows city-level OLS and IV estimates regressing annual population growth for different periods on railroad access in 1848. In panels B and C Railroad access is instrumented by *SLC* location. Standard errors, clustered at the county level, in parentheses. Controls include: subsequent 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 node, and a constant.

***Significant at 1%; **significant at 5%.

Source: See main text and Appendix A for data sources and details.

(2004),[24] we predict actual railroad access *RA* in 1848 with the potential for railroad adoption in 1848—being located within a straight-line corridor *SLC* between nodes,

$$RA_{1848} = \alpha_2 + \beta_2 SLC_{1848} + X'_{1849}\gamma_2 + \eta_t. \tag{2}$$

Until the 1860s, Prussian railroads were built to connect important cities (see also Section 2). Under the assumption that lines were exclusively built to establish a fast connection between important cities A and B, cities en route were able to connect 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 adopt railroad technology. If it were only these cities that had 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 binary variable determined by location on a straight line between nodes. 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.[25] 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.[26] 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 (6) of Table 1 shows that large shares of the lines were built linearly, indicating the high costs of deviation from the straight line.[27]

Using GIS techniques, we connect the nodes between which railroads were constructed with straight lines. The straight lines are chosen to follow the routing of an existing railroad line from node to node. Thus, the instrument proxies potential railroad access along the straight line of existing routes. Furthermore, we create a buffer around these railroad lines (see Figure 1 for examples). 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 for a deviation from the straight line since bridge building

24. The revised version of this paper (Banerjee, Duflo, and Qian 2012), uses distance to the straight line as the main explanatory variable.

25. 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). Online Appendix H provides further helpful examples.

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

27. The "share of straight lines" measure is adopted from official Prussian records.

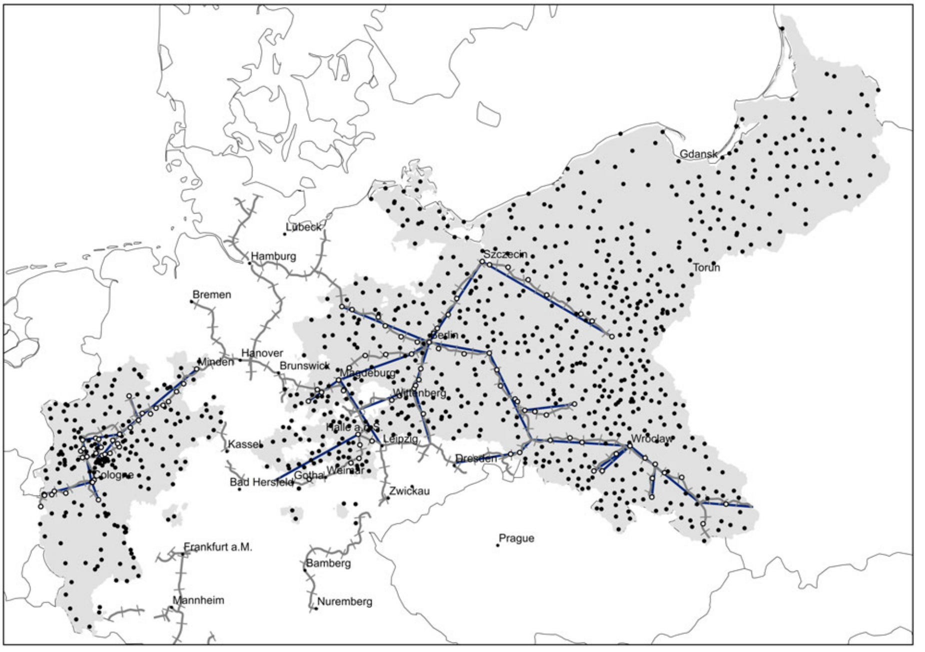

FIGURE 1. German–Prussian railroad network, 1848. Gray area depicts Prussian territory in 1848. Hash lines indicate railroad routings in the German Reich. Tubes indicate the straight-line corridor using a 1.5 km buffer. Hollow circles indicate cities that had a railroad station by 1848. Filled circles indicate cities that did not have access by 1848. Source: Own illustration; see main text for details.

was expensive and orthogonality was required. Thus we allow the buffer to expand the linear line by 1.5 kilometers in each direction.[28]

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.[29]

Table 5 also reports estimates using the straight-line corridor location as an instrument. Panel B shows first-stage results of the IV approach. The instrument *SLC*

28. We find that coefficients remain significant using corridors with a width of 2, 4, 6, 20, 30, or 40 kilometers. Although not significantly different from each other, point estimates decrease when increasing the corridor width. Furthermore, increased corridor width will increase the power of the instrument. See Online Appendix I for a graph that plots beta coefficients against corridor width. Note that a corridor width of 40 km might already pick up cities from other corridors. This could explain the increased beta coefficient as compared to the 30 km corridor width. Also, note that the average distance to the next nearby city is 10.8 km, and 17.4 km to the next nearby city with more than 3,000 inhabitants.

29. We also consider two alternative instruments in Online Appendix J. The first approach draws straight lines between nodes selected in List's railroad plan of 1833. The second approach connects nodes using a least-cost path based on terrain slope and rivers. Results from IV estimations using these approaches are presented in Online Appendix J, Table J.1, and are qualitatively similar to the *SLC* approach.

is significantly correlated with actual railroad access. First-stage F-statistics are high and confirm the power of the instrument. Given that the exclusion restriction holds, second-stage results reported in panel C show the causal effect 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 1849–1871. Across all subperiods under consideration, the effect varies between 1.1 and 2.2 percentage points for a city that gained access by 1848.[30] Again, it is reassuring that the counterfactual model for the period 1831–1837 does not yield significant results (column (1) of Table 5). Coefficients estimated by IV are approximately twice as large as coefficients derived from OLS estimations. The OLS coefficients might be biased downward in the case of an omitted variable—for example, cities with lower growth prospects might have influenced routing in order to become connected.

The exclusion restriction would be violated if the instrument was correlated with the error term. This would be the case if location in the *SLC* was associated with urban population growth through a channel other than the railroad; for example, if the corridor coincided with historical trade routes that still fostered growth. The coefficients would be biased if cities in the *SLC* were larger or better integrated in trade than cities outside of the corridor. When estimating the reduced-form relationship of urban growth on location in the *SLC*, we find no correlation with the pre-railroad growth during 1831–1837 (column (1) in panel D of Table 5).

Although violations of the exclusion restriction cannot be tested formally, we further address such concerns by showing bivariate regressions of the instrument *SLC* on the full set of control variables in Table 6 (columns (1)–(2)). Indeed, we find that *SLC* location is positively associated with street access as well as mining activity and negatively associated with agricultural endowments in the full sample of cities. Thus, controlling for the full set of control variables proves to be important for eliminating such channels.

However, we also find that *SLC* location is correlated with a range of city characteristics measured at a time prior to railroad building. Anticipating our subsequent PSM approach (see Section 4.3), we also show bivariate regression of the instrument *SLC* on the full set of control and matching variables in Table 6 (columns (3) and (4)). In the matched sample of cities, based on weights resulting from a kernel PSM, we find that cities inside and outside the corridor do not differ significantly regarding the control and matching variables.[31] However, we find that cities in the *SLC* are also located closer to a node by 9 km—this indicates that it is harder to find good matches in close proximity to nodes.

30. Robustness tests that introduce 25 district dummies, that exclude the sparsely populated eastern provinces (these are the predominantly Polish-speaking provinces of Prussia, Poznan, Pomerania, and Silesia) or that include continuous variables for roads and waterways instead of dummies, do not yield qualitatively different results (available from the author upon request).

31. A radius-matching approach yields similar results (available from the author upon request).

TABLE 6. Testing for plausibility of the exclusion restriction.

DepVar: Straight-line corridor = 1	Full sample β (1)	Full sample SE (2)	Matched sample β (3)	Matched sample SE (4)
Control variables				
Street access = 1	0.022*	(0.012)	0.058	(0.065)
Waterway access = 1	0.015	(0.016)	0.086	(0.078)
Annual city growth 1831–1837	−0.209	(0.415)	−1.908	(2.596)
Civilian population (log)	0.011	(0.008)	−0.008	(0.053)
Military population (log)	0.001	(0.003)	−0.009	(0.016)
Factory workers (share)	0.028	(0.038)	0.412	(0.424)
Mining (county level)	0.068**	(0.035)	0.116	(0.152)
Large farming (county level)	−0.485***	(0.148)	−1.707	(1.041)
Age composition	−0.100	(0.160)	−0.044	(1.073)
School enrolment rate	0.049	(0.031)	0.104	(0.128)
Distance to node	−0.023***	(0.005)	−0.090***	(0.026)
Matching variables				
Annual city growth 1821–1837	0.010	(0.008)	−0.027	(0.050)
City size 1837 (log)	1.021*	(0.524)	1.345	(4.486)
Merchants 1819 (p.c.)	0.465	(0.679)	0.685	(3.612)
Looms 1819 (p.c.)	−0.175	(0.206)	0.057	(1.790)
Protestants 1816 (share)	0.034***	(0.011)	0.095	(0.068)
Private dwellings 1821 (p.c.)	−0.022	(0.162)	0.228	(0.976)
Commercial buildings 1821 (p.c.)	−1.292***	(0.417)	−3.961	(4.037)
Insurance value of buildings 1821 (log)	0.010*	(0.006)	−0.010	(0.040)
Number of observations	947 controls, 859 matching		623 total, 106 weighted	

Notes: The table shows bivariate regression of straight-line corridor location with all control and matching variables introduced in Tables 3 and 4. Columns (1) and (2) present results in the full sample of cities excluding nodes. Columns (3) and (4) present results in a matched sample of cities (excluding nodes), based on weights resulting from the kernel PSM approach applied in Table 8, panel B, column (2). Standard errors, clustered at the county level, in parentheses. Constant omitted.
***Significant at 1%; **significant at 5%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

4.3. Propensity Score Matching on Pre-Railroad Development

As observed in Tables 3 and 4 in column (5), differences in means between railroad and nonrailroad cities are significant for some of the post-treatment as well as the pre-treatment variables. This casts doubts on the suitability of using the entire sample of nonrailroad cities as a control group. After having established our baseline specifications using the full population of cities in Prussia, this section presents specifications combining PSM and IV estimation.

We employ PSM techniques including indicators of pre-railroad development, size, and geography. The aim of PSM is to compare the outcome for cities that are as similar as possible and—ideally—differ only in their assignment to the treatment. 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 the set of pre-railroad variables presented in Table 4.[32] Since the first Prussian railroad was built in 1838, we match cities by their size in 1837, their population growth during the period 1821–1837, and normalized numbers of merchants in 1819, looms in 1819, Protestants in 1816, private dwellings in 1821, commercial buildings in 1821, and the insurance value of buildings against fire in 1821. These variables are targeted at matching cities regarding their size and commercial development prior to railroad construction. 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.[33]

Table 7 presents descriptive statistics by treatment status for a radius and a kernel matching approach.[34] Column (3) shows differences in means between the matched samples. Both approaches yield highly comparable sets of cities. In this sample we find that otherwise similar cities that gained access to a railroad line before 1848 achieve significantly higher population growth in the period 1849–1871.

As matching can only resolve observed heterogeneity, the endogeneity of railroad access due to unobserved heterogeneity might still be an issue. Thus we combine the outcomes of the PSM with our IV approach. The weights obtained from matching are included to estimate the effect of railroad access on growth using the IV *SLC* in the matched sample.

Results of the IV estimation in a radius-matched sample are shown in panel A of Table 8. We find a significant positive increase of 1.1 percentage points in annual population growth over the period 1849–1871. Since the number of matched observations is small, standard errors are higher and coefficients become insignificant in four subperiods. However, in terms of magnitude, the point estimates range between earlier OLS and IV estimates using the full sample. Since one of the matching variables is the population growth rate for the period 1821–1837, it does not come as a surprise that the counterfactual model for the period 1831–1837 does not yield significant results in the matched specifications (column (1) in each panel). Results of the IV estimation in a sample matched using a nonparametric kernel approach are presented

32. For this purpose, we exclude all cities that gained access to railroads during the period under consideration from the matching. As the number of cities with railroad access increases, finding suitable matches becomes more difficult. This is the reason why the sample size decreases over the subperiods when using a radius matching.

33. Online Appendix K shows the frequency distribution of the propensity score by treatment status.

34. Radius matching finds all untreated observations that are within distance of a specified caliper (0.001) to a treated observation according to the propensity score. 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.

TABLE 7. Descriptive statistics after propensity score matching.

	Railroad cities excl. nodes Mean (1)	Nonrailroad cities Mean (2)	Difference in means between (1) and (2) (3)	SE of the difference in means (4)
Panel A: Radius-matched sample				
Annual city growth 1849–1871	0.012	0.006	0.006***	(0.001)
City size 1837 (log)	7.823	7.843	−0.020	(0.115)
Annual city growth 1821–1837	0.015	0.014	0.001	(0.001)
Merchants 1819 (p.c.)	0.014	0.016	−0.002	(0.002)
Looms 1819 (p.c.)	0.021	0.021	0.000	(0.004)
Protestants 1816 (share)	0.788	0.755	0.033	(0.065)
Private dwellings 1821 (p.c.)	0.138	0.139	−0.001	(0.005)
Commercial buildings 1821 (p.c.)	0.005	0.004	0.001	(0.001)
Insurance value of buildings 1821 (log)	5.557	5.572	−0.015	(0.113)
Observations	36	208	244	
Weighted observations	36	36	72	
Panel B: Kernel-matched sample				
Annual city growth 1849–1871	0.014	0.007	0.007***	(0.002)
City size 1837 (log)	8.086	7.945	0.141	(0.105)
Annual city growth 1821–1837	0.015	0.014	0.001	(0.001)
Merchants 1819 (p.c.)	0.016	0.015	0.000	(0.001)
Looms 1819 (p.c.)	0.017	0.018	−0.001	(0.003)
Protestants 1816 (share)	0.780	0.744	0.036	(0.050)
Private dwellings 1821 (p.c.)	0.132	0.134	−0.002	(0.004)
Commercial buildings 1821 (p.c.)	0.004	0.004	0.000	(0.001)
Insurance value of buildings 1821 (log)	5.659	5.585	0.074	(0.099)
Observations	53	570	623	
Weighted observations	53	53	106	

Notes: The table presents descriptive statistics for matching variables by treatment status. Panel A presents results in a matched sample based on weights resulting from the radius PSM approach applied in Table 8, panel A, column (2). Panel B presents results in a matched sample based on weights resulting from the kernel propensity score matching approach applied in Table 8, panel B, column (2). Standard errors in parentheses.
***Significant at 1%.
Source: See main text and Appendix A for data sources and details.

in panel B of Table 8. Results obtained in this sample are qualitatively similar to the radius-matched sample. We find that railroad access significantly increases annual population growth by 1.7% over the period 1849–1871 using a kernel-matched sample.

In panel C, we expand the matching variables to include the geographic location of a city, namely longitude and latitude. In doing so, we aim at finding pairs of cities that are similar in terms of location as well as in size and commercial development previous to railroad construction. Estimates using weights from this matching approach behave

TABLE 8. The impact of railroad access, cross-sectional IV estimations in matched samples.

	Main periods		Subperiods						
DepVar: Population growth rate	1831–1837 (1)	1849–1871 (2)	1849–1852 (3)	1852–1855 (4)	1855–1858 (5)	1858–1861 (6)	1861–1864 (7)	1864–1867 (8)	1867–1871 (9)
Panel A: IV estimates; sample and weights obtained from PSM on pre-railroad variables (radius: caliper = 0.001)									
Rail access 1838–1848	–0.003	0.011**	0.011	0.005	0.018***	0.007	0.031*	0.028	0.039**
	(0.010)	(0.004)	(0.009)	(0.011)	(0.007)	(0.008)	(0.018)	(0.020)	(0.015)
Observations	323	244	328	285	342	278	262	232	191
Weighted obs.	92	72	88	80	84	82	68	74	62
Kleibergen–Paap F-statistic	41.75	14.91	36.26	40.14	38.27	44.04	35.15	23.67	41.00
Panel B: IV estimates; sample and weights obtained from PSM on pre-railroad variables (kernel)									
Rail access 1838–1848	–0.004	0.017**	0.014*	0.010	0.022**	0.007	0.020*	0.020	0.024***
	(0.009)	(0.008)	(0.008)	(0.010)	(0.011)	(0.008)	(0.011)	(0.017)	(0.009)
Observations	810	623	803	786	759	724	705	676	619
Weighted obs.	118	106	118	112	112	110	108	104	104
Kleibergen–Paap F-statistic	62.76	78.38	74.99	83.58	80.90	71.35	75.67	72.03	61.44
Panel C: IV estimates; sample and weights obtained from PSM on pre-railroad variables and geography (kernel)									
Rail access 1838–1848	–0.004	0.017**	0.012	0.010	0.022**	0.008	0.020*	0.021	0.025***
	(0.010)	(0.008)	(0.008)	(0.010)	(0.011)	(0.008)	(0.011)	(0.017)	(0.009)
Observations	808	620	800	786	759	724	705	678	616
Weighted obs.	114	100	112	112	112	110	108	108	98
Kleibergen–Paap F-statistic	61.77	41.71	80.35	80.30	76.08	57.67	56.09	55.26	32.32

Notes: The table shows city-level IV estimates regressing annual population growth for different periods on railroad access in 1848 in matched and weighted samples. Railroad access is instrumented by straight-line corridor location. Standard errors, clustered at the county level, in parentheses. All regressions include the full set of baseline controls: subsequent 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 node, and a constant.

***Significant at 1%; **significant at 5%; *significant at 10%.

Source: See main text and Appendix A for data sources and details.

very similarly to prior matching approaches—railroad access significantly increases annual population growth by 1.7% over the period 1849–1871.[35]

Comparing the overall results from our matching approach to the baseline estimates, we find qualitatively similar results.[36] However, point estimates are usually somewhat smaller, indicating that unobserved heterogeneity between cities might account for part of the railroad effect.

4.4. Panel Data Estimates

In our preferred approach to estimate the effect of railroad access on urban growth, we use panel techniques. The advantage of the panel approach is the possibility to overcome time-invariant unobserved heterogeneity by including fixed effects. Here, the city-level heterogeneity in pre-railroad development observed in our matching variables in Table 4 can be excluded by including city fixed effects. This allows us to exclusively exploit 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:

$$\ln POP_{it} = \alpha_i + \tau_t + \beta_3 RA_{it-1} + X'_{it}\gamma_3 + \nu_{it}. \tag{3}$$

We can 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 1861, just at the end of stage 2 of the railroad network expansion—connecting major cities. Since the censuses provide triennial data, we derive a panel consisting of eight repeated cross-sections. The only information published in this frequency at the city level is population counts. Thus, only a few control variables X' from the original model in equation (1) are available in the panel setting. Available city-level controls include the military population, distance to the next node, and a dummy that controls for the incorporation of municipalities as provided by Matzerath (1985).

We present panel estimates in Table 9. The first specification reports estimates in a pooled sample, including time fixed effects (column (1)). Column (2) introduces city fixed effects and thus shows the within-city effect of gaining railroad access in

35. Again the PSM successfully reduces the differences in means between railroad and nonrailroad cities. Results are qualitatively similar but coefficients seem less stable when using radius-matching techniques. Furthermore, results show similar patterns when using the matching variables as control variables instead of using them for the matching approach (results available upon request from the author).

36. Similar results are found when estimating in samples consisting only of cities smaller than 3,000 inhabitants, cities matched to their two next geographical neighbors, or only of cities within 15 km distance to the *SLC* (see Online Appendix G, Table G.1, panels G–I).

TABLE 9. The impact of railroad access, panel estimates.

	OLS					
DepVar: (ln) Population	Pooled regression (1)	Fixed effects (2)	Nodes control (3)	County × period FE (4)	East sample (5)	West sample (6)
Railroad access = 1	0.088***	0.077***	0.055***	0.068***	0.050***	0.059**
	(0.013)	(0.013)	(0.011)	(0.017)	(0.010)	(0.029)
Period fixed effects	Y	Y	Y	Y	Y	Y
City fixed effects	N	Y	Y	Y	Y	Y
Nodes control	N	N	Y	Y	Y	Y
County×year fixed effects	N	N	N	Y	N	N
Observations	7,737	7,737	7,737	7,737	6,007	1,730
Number of cities	978	978	978	978	756	222
R-squared		0.40	0.40	0.72	0.55	0.32

	OLS				IV	
DepVar: (ln) Population	Geographic sample (7)	Small city sample (8)	Large city sample (9)	State controls (10)	First stage (11)	Second stage (12)
Railroad access = 1	0.035***	0.060***	0.029*	0.042***		0.072*
	(0.013)	(0.017)	(0.017)	(0.011)		(0.040)
State owned = 1				0.003		
				(0.019)		
State administered = 1				0.067**		
				(0.032)		
Straight-line corridor = 1					0.475***	
					(0.060)	
Period fixed effects	Y	Y	Y	Y	Y	Y
City fixed effects	Y	Y	Y	Y	Y	Y
Nodes control	Y	Y	Y	Y	Y	Y
County×year fixed effects	N	N	N	N	N	N
Observations	3,443	5,296	1,185	7,737	7,737	7,737
Number of cities	438	667	153	978	978	978
Kleibergen–Paap *F*-statistic.						63.60
R-squared	0.41	0.35	0.59	0.41	0.34	0.40

Notes: Panel estimates at the city–year level using triennial data for the period 1840–1861. Railroad access indicates if a city had access to the railroad network in a previous year. Column *Fixed effects* introduces county fixed effects, column *Nodes control* includes a dummy for the nodes of the network, column *County×period FE* introduces a full set of interactions of county fixed effects with time period fixed effects, columns *East sample* and *West sample* restrict to regional samples, column *Geographic sample* restricts to cities close to the *SLC*, columns *Small city sample* and *Large city sample* distinguish by city size prior to 1837, column *State controls* introduces dummies for state involvement. Columns *First stage* and *Second stage* indicate first-stage and second-stage estimates, instrumenting actual railroad access with straight-line corridor location. Further controls: military population (log), distance to next node, and a dummy for incorporations. Standard errors, clustered at the county level, in parentheses.

***Significant at 1%; **significant at 5%; *significant at 10%.

Source: See main text and Appendix A for data sources and details.

one period on subsequent additional population growth, similar to a difference-in-differences approach. The dummy variable indicating railroad access switches from 0 to 1 after a city is connected to the railroad network. Interestingly, the coefficient estimated in the fixed-effects model is also close to the pooled sample, indicating low levels of unobserved heterogeneity at the city level.

For the cross-sectional analysis, we excluded nodes since their access to railroads is likely to be driven by characteristics not fully observed. Such characteristics are time invariant and should be absorbed by the city fixed effects included in our panel estimates. However, in column (3), we include a dummy that accounts for time-invariant characteristics of the node status. The dummy switches to 1 after a city becomes a node. After controlling for nodes, the results indicate that railroad access additionally increases urban population levels by 5.5% over a period of three years. This translates into an annual rate of 1.8%.[37]

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. To address this issue, the specification in column (4) adds 324 county fixed effects interacted with time fixed effects. Such a specification captures county-wide shocks during one period that affected all cities within the same county. 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. The point estimate increases in magnitude to the previous specification implying that there might indeed have been negative shocks at the county level. The difference between coefficients is, however, not significant.

It is also interesting to test for heterogeneous treatment effects since railroad access might be something that does not affect every city in the same vein. Columns (5) and (6) report estimates in separate samples for the six eastern provinces and the two western provinces. We find that the effects are larger in the west, confirming expectations due to strong industrialization in this region. The difference is however not significant. In column (7) we restrict our sample to cities within close proximity of 15 km to the *SLC* in 1861. Compared to the full-sample estimates the coefficient is lower. Further estimates suggest that the coefficient increases with distance to the straight line (available from the author upon request). This finding indicates the existence of spatial spillovers. Nonrailroad cities located in proximity to a railroad line might benefit from positive spillovers and have higher growth rates; cities far away from any railroad line might suffer from remoteness and have lower growth rates.

Furthermore, we report estimates in separate samples for small cities and large cities in columns (8) and (9).[38] Interestingly, we find that effects are larger for smaller cities, indicating that the estimated effect is not driven by cities that were already

37. Alternatively, the exclusion of node cities from the sample leads to very similar results.

38. A city is identified as small if its size was below 3,000 inhabitants in 1837, before the first railroad was built; a large city had more than 5,000 inhabitants in 1837. The coefficient found in a sample of medium cities is very similar to the small-cities sample.

TABLE 10. Year of establishment and the profitability of railroad lines.

DepVar: ROI in	Year 1 (1)	Year 2 (2)	Year 3 (3)	Year 4 (4)	Year 5 (5)	Year 6 (6)	Year 7 (7)	Year 8 (8)	Year 9 (9)
Year built	−0.061	−0.219***	−0.225***	−0.140*	−0.081	−0.083	−0.074	−0.055	−0.092
	(0.067)	(0.075)	(0.067)	(0.069)	(0.074)	(0.068)	(0.066)	(0.071)	(0.066)
Observations	34	34	33	34	34	34	34	34	34
R-squared	0.04	0.23	0.28	0.13	0.04	0.05	0.03	0.02	0.04

Notes: The table shows bivariate regressions of return on investment (*ROI*), normalized to years after establishment, on year of establishment of railroad lines for all lines built until 1860. Standard errors, clustered at the county level, in parentheses. Constant omitted.
***Significant at 1%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

large. Furthermore, these results indicate that much of the action took place in initially smaller towns. The literature's general focus on large towns might thus actually neglect seizable effects.

We also test for heterogeneity in the provision of railroad lines. From 1850 the Prussian state became more involved in railroads and started building and running railroad lines on its own account.[39] In column (10), we include two dummies each taking the value of 1 just after a railroad became state owned or state administered. The coefficient on state-administered railroads is significant and within the range of our previous findings. However, the point estimate on state-owned railroads is close to zero and insignificant suggesting that these lines did not induce similar growth.

We can further address endogeneity issues in the panel setting using the straight-line corridor approach introduced in Section 4.2. A strong advantage of our setting is that the instrument actually exhibits time variation. Whenever new lines were built, new straight-line corridors are established, providing over-time variation in the instrument. We construct straight-line corridors on a triennial basis and use them to generate an instrument that varies over time.

Second-stage results presented in column (12) show a causal effect of 7.2%. This means that a city that was connected to the railroad subsequently experienced an additional growth of roughly 2.4% per year. This result is actually close to the 2.1% annual growth found in the baseline IV specification presented in Table 5 for the period 1849–1871.

The exclusion restriction might be violated if the establishment of railroad lines over time was correlated with, for example, their expected profitability. We test for this possibility by running bivariate regressions of profitability measured by return on investment (*ROI*) on the year of establishment. For each railroad line, the *ROI* is normalized to years after establishment. The results presented in Table 10 show

39. Railroad building occurred to increase profits, adding to the budget, as well as for military reasons.

that after an initial stage of roughly four years there is no systematic difference in profitability for lines that had been established earlier.[40]

4.5. *Event Study Analysis*

To determine the timing of a long-term shift in growth rates we estimate event-study specifications. Since the event of gaining railroad access occurs at different times for different locations, the panel analysis with city and time fixed effects will prove extremely helpful. Such an identification strategy can test whether results are driven by underlying pre-event trends since it traces out the trend in growth rates for the periods leading up to and following railroad access. The presence of a pre-trend would raise concerns about our identification strategy.

To be able to show the dynamics of the effects from gaining access, we code separate dummies for all periods before and after adoption.[41] Each dummy only takes the value 1 in a single period prior to or after a city becomes connected to the railroad and is 0 for all other periods. Since the occurrence of railroad access varies over time across cities, this approach allows us to pool the statistical information for each separate stage of railroad access. Effectively, we can now compare cities in the same stages of railroad access across periods.

Table 11 shows significant positive effects of railroad access for all periods after gaining access for different specifications. All coefficients are measured relative to the omitted coefficient which is the period prior to access. It is important to emphasize that we find no differences in growth rates prior to the event, which excludes the possibility of pre-trends. This finding becomes visible in Figure 2 which plots the beta coefficients over time allowing an inspection of the pre- and post-event growth for cities that gained railroad access during the period 1838–1861. The figure shows an absence of pre-railroad trends in growth rates and a sharp upward trend after railroad access is established. In combination with city fixed effects, the absence of pre-trends confirms the perception that the results do not suffer from unobserved heterogeneity issues.

Our results are further confirmed by IV estimates instrumenting periods of adoption with periods of *SLC* location. Results are presented in column (5) of Table 11 and Figure 2.

5. Discussion of the Results

The previous analysis focuses on establishing causality between railroad access and population growth. Such reduced-form estimates provide a very general lesson on the relative impact of gaining railroad access which can be attributed to a range of different mechanisms. This section provides evidence for one of the possible mechanism through

40. The initial stage is characterized by a high number of lines effectively realizing losses with a *ROI* of zero.

41. Each period spans a triennium and is thus of the same length.

TABLE 11. The impact of railroad access, nonlinear panel estimates.

DepVar: (ln) Population	Fixed effects (1)	Nodes control (2)	Small city sample (3)	Geographic sample (4)	IV second stage (5)
4 periods prior to access	0.007	0.007	0.011	0.011	−0.047**
	(0.006)	(0.006)	(0.008)	(0.012)	(0.020)
3 periods prior to access	0.007	0.007	0.014	0.017	−0.048**
	(0.008)	(0.008)	(0.014)	(0.016)	(0.022)
2 periods prior to access	0.002	0.002	0.013	0.003	−0.046**
	(0.009)	(0.009)	(0.017)	(0.014)	(0.020)
1 period prior to access	0.003	0.004	0.010	0.000	−0.041***
	(0.008)	(0.008)	(0.015)	(0.009)	(0.016)
Access for 1 period	0.051***	0.035***	0.045**	0.023*	−0.019
	(0.012)	(0.011)	(0.020)	(0.013)	(0.022)
Access for 2 periods	0.069***	0.051***	0.062***	0.035**	0.016
	(0.015)	(0.013)	(0.023)	(0.015)	(0.024)
Access for 3 periods	0.098***	0.078***	0.087***	0.059***	0.074*
	(0.018)	(0.015)	(0.027)	(0.016)	(0.040)
Access for 4 periods	0.113***	0.093***	0.107***	0.069***	0.067
	(0.026)	(0.024)	(0.034)	(0.025)	(0.053)
Access for 5 periods	0.146***	0.123***	0.125***	0.090***	0.172**
	(0.022)	(0.020)	(0.034)	(0.020)	(0.076)
Access for more than 5 periods	0.162***	0.134***	0.100***	0.090***	0.128**
	(0.026)	(0.026)	(0.030)	(0.031)	(0.053)
Year fixed effects	Y	Y	Y	Y	Y
City fixed effects	Y	Y	Y	Y	Y
Nodes control	N	Y	Y	Y	Y
Observations	7,737	7,737	5,296	3,443	7,737
Number of cities	978	978	667	438	978
Kleibergen–Paap F-statistic.					7.52
R-squared	0.41	0.41	0.36	0.42	0.39

Notes: Panel estimates at the city–year level using triennial data for the period 1840–1861. Each period has a length of three years. Column *Fixed effects* introduces county fixed effects, column *Nodes control* introduces a dummy for the nodes of the network, column *Small city sample* restricts sample to cities smaller than 3,000 inhabitants in 1837, column *Geographic sample* restricts to cities close to the *SLC*, column *IV second stage* indicates second-stage estimates from instrumenting actual period of railroad access with periods of straight-line corridor location. Further controls: military population (log), distance to next node, and a dummy for incorporations. Standard errors, clustered at the county level, in parentheses.
***Significant at 1%; **significant at 5%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

which railroads affect growth, integrates our findings into the recent literature on railroads, and further discusses issues regarding economic spillovers, heterogeneous treatment effects, and the generalizability of the results.

5.1. Possible Mechanisms

The literature identifies the immediate effects of transport infrastructure expansion in terms of reductions in the cost of trading, increases in trade volumes, increases

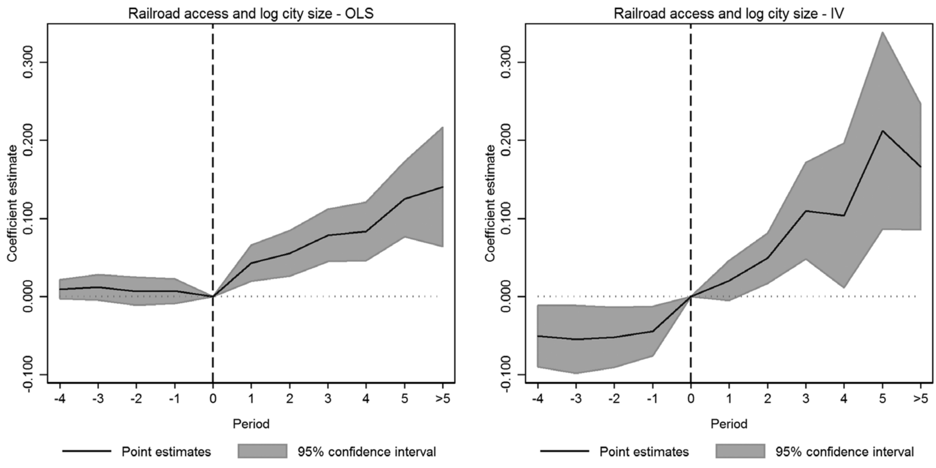

FIGURE 2. Event study—railroad access and log city size. The figure plots coefficients from the event study analysis presented in Table 11, columns (2) (OLS) and (5) (IV). Each period has a length of three years. The solid black lines plot estimated coefficients over periods of railroad access; the gray areas plot the corresponding 95% confidence intervals. Dashed lines indicate period zero. Coefficients are estimated relative to this period zero, before railroad access was established in a city. Specifications include controls for military population (log), distance to next node, and a dummy for incorporations, as well as city and period fixed effects.

in market access, and reductions in price gaps (see Donaldson 2014; Donaldson and Hornbeck 2013; Keller and Shiue 2013, 2014; Michaels 2008). Other works show that such advantages might further translate into increases in firm size, firm profit, the number of firms, the female labor force participation rate, and employment (see Atack et al. 2010; Atack, Haines, and Margo 2011; Banerjee, Duflo, and Qian 2012; Duranton and Turner 2012). Regarding urban expansion this implies that cities can support a larger number of individuals due to decreases in the cost of living—that is, in the form of a reduction in food prices or due to higher incomes—for example, higher wages due to productivity gains.

Atack, Haines, and Margo (2011) argue that railroads increase competition among firms due to an increase in the market size. Consequently, firms attempt to increase productivity through the division of labor which in turn leads to an increase in establishment size. As industrial productivity increases, so do wages, attracting an inflow of workers from rural areas to urban centers (Malanima 2010).[42]

Similar to Atack, Haines, and Margo (2011), we test whether railroad access increased firm size, which might have translated into urban population growth in 19th-century Prussia. The 1849 census (Statistisches Bureau zu Berlin 1851) includes a

42. 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 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).

factory census allowing the effects of railroad access on firm size to be tested at the city level.[43] We calculate the average size of factories at the city level and use it as an alternative dependent variable in the cross-sectional set-up using OLS, IV, and matching techniques similar to the approaches introduced previously. Since such data are exclusively available for the 1849 cross-section, we can only estimate the level effects of railroad access on firm size.

The results of this approach are presented in Table 12. Column (2) shows that firms located in a city with a railroad station were 74% larger than in cities without a station. This effect increases to 109% in the kernel-matched sample, implying that establishments in treated cities are more than twice as large. Reassuringly, we find no counterfactual effect of railroads built after 1848 on firm size in 1849.[44]

Do firms in railroad cities just grow bigger or are these cities also able to attract a larger number of firms? Columns (4)–(6) in Table 12 show that the latter was not the case. The estimated effect of railroad access on the number of firms is insignificant in the IV and PSM specifications. Our results might thus indicate that firms indeed increased the division of labor as a response to railroad access. Railroad access seems to have affected industry location only at the intensive margin rather than the extensive margin.[45]

The additional employment opportunities generated in railroad cities can induce population growth either by attracting immigration or by increasing fertility. The data allow us to distinguish between sources of population growth in our cross-sectional setting. Columns (7)–(9) present results when using fertility as an outcome. Throughout specifications we find no significant effect of railroad access on the child–woman ratio in 1849.[46] Columns (10)–(12) present results when using migration as an outcome. IV estimates show that the share of urban dwellers that were born outside of the city was 13.8% higher in railroad cities in 1871 (12.9% in the unmatched sample).

Furthermore, we can use differences in market prices to test whether market integration is the only channel through which railroads affect growth. Using a cross-section of county-level data on the average market prices for different crops for the period 1837–1860, we include the price of the most important crops in our cross-sectional IV specification. Prices enter the baseline model with significant coefficients.[47] The coefficient on railroad access is hardly affected. Keeping in mind that this analysis can only estimate a level effect and might be flawed due to the unavailability of city-level market prices, we find no evidence that differences in market integration account for the entire effect of railroads on growth.

43. The census reports the number of factories and workers in 119 different product categories. For further information see Becker, Hornung, and Woessmann (2011).

44. Note that these regressions exclude the control for the share of factory workers that is usually included in all other regressions. However, if they do, the coefficient on railroad access 1838–1848 is hardly affected.

45. Similarly, Gutberlet (2014) finds that railroads increased manufacturing employment at the district level in Germany.

46. The child–woman ratio is calculated as the ratio of children under 5 to women aged 15–45.

47. See Online Appendix G, Table G.1, panel F.

TABLE 12. Railroads, industrial development and sources of population growth.

DepVar:	Firm size 1849 (log)			Number of firms 1849 (log)		
Sample:	Full		Matched	Full		Matched
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Rail access 1838–1848	0.312**	0.743**	1.089**	−0.161**	−0.203	−0.117
	(0.121)	(0.338)	(0.523)	(0.075)	(0.144)	(0.163)
Rail access 1849–1871	0.006	0.086		0.072	0.064	
	(0.079)	(0.097)		(0.062)	(0.063)	
Controls	yes	yes	yes	yes	yes	yes
Observations	922	922	618	924	924	620
Weighted obs.			106			106
Kleibergen–Paap *F*-statistic		28.29	76.94		28.29	77.00

DepVar:	Child–woman ratio 1849			Born outside city 1871 (share)		
Sample:	Full		Matched	Full		Matched
	OLS (7)	IV (8)	IV (9)	OLS (10)	IV (11)	IV (12)
Rail access 1838–1848	−0.001	0.005	−0.047	0.067***	0.129***	0.138**
	(0.013)	(0.041)	(0.051)	(0.013)	(0.038)	(0.057)
Rail access 1849–1871	−0.007	−0.006		0.030***	0.042***	
	(0.009)	(0.011)		(0.008)	(0.011)	
Controls	yes	yes	yes	yes	yes	yes
Observations	934	934	630	926	926	623
Weighted obs.			106			106
Kleibergen–Paap *F*-statistic		28.56	84.35		28.58	78.38

Notes: Estimations at the city level for different outcomes. Columns (1), (4), (7), and (10) show OLS estimates in the full sample excluding nodes. Columns (2), (5), (8), and (11) show second-stage results, instrumenting Rail access 1838–1848 with *SLC*. Columns (3), (6), (9), and (12) show IV estimates in a matched sample based on weights resulting from the kernel PSM approach applied in Table 8, panel B, column (2). Standard errors, clustered at the county level, in parentheses. Controls include: street access dummy, waterway access dummy, annual city growth 1831–1837, civilian population (log), military population (log), mining (county level), factory workers (share, only included in columns (7)–(12)), large farming (county level), age composition, school enrollment rate, distance to node, and a constant.
***Significant at 1%; **significant at 5%.
Source: See main text and Appendix A for data sources and details.

We can also think of various other channels through which railroads might affect growth. Railroad lines were accompanied by a series of other improvements such as new forms of information exchange by telegraph lines[48] or by improved

48. Railroad construction was often accompanied by the development of telegraph lines, which were built along the railroad line and in a number of cases even incorporated into the railroad embankment. Thus, in many cases, railroad adoption also meant the adoption of telegraphy which might have advanced the speed of communication.

postal service.[49] Railroads may thus foster technological diffusion through knowledge exchange. According to Mokyr (2002, p. 30), the "technology of knowledge transmission" is important to the diffusion of knowledge and technology itself. Benefits from such improvements due to railroads are captured in our access measure and cannot be separated with this data set.

5.2. *Discussion of the Results in Context of the Literature*

Recently the literature has seen an increasing amount of research on the effects of transport infrastructure expansion, which can be roughly divided into two groups: research focusing on aspects of market integration due to the expansion of the transport network infrastructure and research focusing on macro aspects of development and growth. Here, we will focus on discussing findings by the latter group that is more closely related to our paper.

Besides reduced price gaps and increased trade flows across Indian districts, Donaldson (2014) also finds increases in real income levels due to railroad access. He estimates that railroads increased agricultural income by 16% and that only 14% of the effect cannot be attributed to increases in trade. This study on historical India fundamentally differs from the Prussian setting in that labor was less mobile, railroads did not foster growth in the negligible industrial sector and lines were built mainly for military purposes. Banerjee, Duflo, and Qian (2012) focus on the long-run effects of railroad building in contemporary China. Estimates yield only small effects on a range of development indicators, a result that might also be attributed to a lack of factor mobility in this particular Chinese setting.

The Indian and Chinese examples thus provide benchmarks for the impact of railroads in rural societies that are characterized by low factor mobility. However, mid–19th century Prussia was characterized by an increasing factor mobility and strong industrial development. Thus, our results might be more suitable to be extended to countries in periods of industrial development.

For example, Atack et al. (2010) find that railroads explain 58.3% of urbanization in the midwestern United States in the 1850s at the county level. The somewhat smaller effects found for population density might be caused by the special US case with abundant land. Furthermore, Atack, Haines, and Margo (2011) find that railroads induced industrial development by increasing the likelihood that an establishment was a factory by roughly 16%. These findings are much more in line with our findings on the industrial development in Prussia presented previously.

Donaldson and Hornbeck (2013) estimate the effect of changes in market access due to railroads on changes in agricultural land values in a panel of US counties.

49. From the early days of the Prussian railroad network on, railroads took over the function formerly performed by stage coaches—the transport of passengers and mail. Borchardt (1972) even 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. For an assessment of the effect of postal services on the spatial structure of the population distribution in the German Empire see Ploeckl (2012).

The authors calculate that removing all railroads built by 1890 would translate into a reduction in agricultural land values equal to an annual loss of 3.4% of GNP.[50] However, this setting can only provide evidence for the nonindustrial sector of the economy.

The works presented by Keller and Shiue (2013, 2014) differ from our approach in analyzing bilateral trade flows between major German cities using 19th-century city-level wheat prices, suggesting that trade is one of the most important channels through which railroads affect growth. Such a setting is particularly useful in assessing whether a particular connection between two cities changes their relative economic environment. Here, railroads as determinants of trade function as a proximate factor through which a fundamental factor, the institution of the German Customs Union, determines development.

There are institutions which might be qualified as necessary preconditions for railroads to be effectively fostering growth. Prussian railroads probably would have been less effective without the German Customs Union which created free trade (see Keller and Shiue 2014). Similarly, the free movement of the factors of production might be a necessary precondition as indicated by Banerjee, Duflo, and Qian (2012) for China. These necessary institutions allowing for the free movement of labor were introduced with the agricultural reforms in Prussia at the beginning of the 19th century.[51]

In sum, the recent literature has predominantly focused on analyzing the consequences of establishing railroads for the agricultural sector or for agricultural societies with limited factor mobility. As railroads are often strongly connected to the industrial sector, the Prussian environment seems a natural laboratory to assess the consequence of railroads on industrial development.

5.3. Spillovers and Heterogeneous Treatment Effects

Some issues arise when assessing the effects of place-based network policies. Growth due to railroad access might either induce positive spillovers or happen at the expense of other regions.[52] Furthermore, heterogeneous-treatment effects become an issue in networks where parts of the network are of a higher local value. More specifically, railroad access might be something that is not of fixed importance: different lines might have different effects on otherwise similar cities. Even cities on the same line might be affected in very different ways depending on their local endowments.

---

50. As changes in market access represent a cumulative measure of all changes in the transportation infrastructure network, the results of such estimations show an aggregated effect on the economy.

51. Landes (1969, p. 154) notes that faster transportation meant that labor became more mobile and that natural obstacles to the movement of the factors of production were eliminated.

52. City growth usually results in large parts from migration. We can think of a scenario where the positive growth effect for treated cities is entirely due to urban–urban immigration from untreated cities, leading to an aggregate effect of zero. In mid-19th century Prussia, where factor mobility was already very high, it is however much more plausible to think of growth due to rural–urban migration. Ziegler (1996, p. 304) notes that we know today that railroad adoption did not end in a zero sum game at the expense of other regions. In such a setting railroads might have worked as a pull factor for rural–urban migration.

TABLE 13. Railroads and urbanization in a panel of counties.

DepVar: Urbanization	Railroad–stations OLS (1)	Railroad–stations IV (2)	Railroad–population OLS (3)	Railroad–population IV (4)
Number of cities with stations	0.007*	0.012*		
	(0.004)	(0.007)		
Share of population with access			0.095***	0.087**
			(0.029)	(0.035)
Year fixed effects	Y	Y	Y	Y
County fixed effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Observations	670	670	670	670
Number of counties	335	335	335	335
Kleibergen–Paap *F*-statistic		46.08		23.10
R-squared	0.19	0.19	0.25	0.25

Notes: Panel estimates at the county level with a full set of county and year dummies using two cross-sections for 1849 and 1864. The dependent variable is urbanization—measured as the share of city dwellers in the total county population. Columns (1) and (2) show OLS and IV estimates using the number of cities in a county that have a station as the explanatory variable. In column (2) this endogenous variable is instrumented with the number of cities in a county located within the 1.5 km *SLC*. Columns (3) and (4) show OLS and IV estimates using the share of a county's population that lives in a city with a station as the explanatory variable. In column (2) this endogenous variable is instrumented with the share of a county's population that lives in a city located within the 1.5km *SLC*. Standard errors, clustered at the district level, in parentheses. County-level controls include: population density, factory workers (share), agriculture (share), large farming (share), age composition, school enrollment rate, distance to node county, and a constant.
***Significant at 1%; **significant at 5%; *significant at 10%.
Source: See main text and Appendix A for data sources and details.

Within our framework, which uses a binary measure for local network access, it is however not possible to be definite about these issues. Donaldson (2014) and Donaldson and Hornbeck (2013) show that using measures for transportation infrastructure which capture the effect of the entire railroad network on a place leads to estimated effects which are larger than when using measures of local railroad access. These findings suggest that our results might be interpreted as underestimating the effect of railroads in Prussia. From our various robustness tests we can further infer that there seem to be heterogeneous effects over time, depending on geography, public provision of the infrastructure, or depending on the size of a city.

Concerns that the overall effect in equilibrium might be much smaller are associated with the choice of the city level as a unit of observation. We aim to attenuate such concerns by aggregating our data to the county level. When escalating to this higher level of aggregation, we use urbanization as an outcome variable which we regress on county-level indicators of railroad access. County-level data on urbanization and many other control variables are available for two cross-sections in 1849 and 1864 which allows us to estimate fixed-effects models using panel data.

Table 13 presents OLS and IV results from our regression. Here, county-level indicators of railroad access are the number of cities with railroad access and the share of a county's population living in cities with access.[53] The instrument is again based on the established *SLC* and counts the number of cities in a county located within 1.5 km range of the *SLC* or the share of a county's population that lives in a city located within this range, respectively. We find significant positive effects of increasing the number of stations as well as increasing the share of population with railroad access. The average change in the number of stations per county during the period 1849–1864 is 0.37 which we estimate to cause a change in urbanization of 0.44%—or roughly one-third of the overall increase in urbanization during the period. Similarly, a 6% change in the population with access to a railroad causes an estimated change in urbanization of 0.52%. It thus seems that we can exclude concerns of small or nonexisting general equilibrium effects—at least for the county level.

One limitation of using IV estimation approaches lies in the fact that we can only estimate the local average treatment effect (*LATE*) of railroad access for cities in the *SLC*. Since the IV approach is not informative about the noncompliers—that is, cities in the corridor that do not gain access or cities outside of the corridor that gain access regardless—we cannot be specific about average treatment effects (*ATE*). During the course of the paper, we have also successively reduced the sample by excluding nodes, excluding large cities, or applying a matching methodology. Judging from such reduced samples, it is difficult to assess the effect of railroad access for the entire population of Prussian cities. Due to these limitations, the results might not be generalizable to the full population without reservations. However, throughout all specifications we have consistently found effects ranging between 1 and 2 percentage points increases in annual population growth. This very stable result indicates that the heterogeneity of the effect is somewhat limited to a narrow range. This finding might also help to generalize from local average treatment effects to a statement about the entire population.

6. Conclusion

This paper tests the hypothesis that railroads induced economic growth at the city level for the historical German state of Prussia during a period of rapid railroad network expansion and indutrialization. We find that railroads had a significant causal effect on urban population growth over the period 1838–1871. Cities that gained railroad access during this period experienced additional *annual* growth of roughly 1% to 2%—a substantial amount. Across a range of different specifications, the effect is presumably best identified in the PSM and fixed-effects panel estimations using instrumental variables.

The paper adds to the literature by successfully establishing a time-variant instrumental variable to estimate the causal effect of transport infrastructure on growth. As such, we can plausibly introduce exogenous within-city variation in railroad access into

53. Unfortunately, this measure does not capture the few places that had railroad access but did not have city rights.

a panel using city-fixed effects. Excluding time-invariant differences between cities, this allows for a convincing assessment of the local consequences of railroad access.

This paper further provides evidence for one of the channels through which railroads affect growth, namely by increasing firm size. We further show that railroad access induced immigration, leading to city growth. The results show that cities with railroad access hosted factories that were more than twice as large, presumably triggering population growth through the demand for workers from the industrial sector.

We contribute additional evidence to the debate over whether railroads induced or followed economic growth. 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. The further inspection of nonlinearities in within-city variation of growth show that railroad access changed the pattern of growth for adopting cities.

Our analysis is, however, limited in providing evidence for the effects of the entire railroad network and cannot take spillovers into account. Furthermore, we suspect the existence of heterogeneous treatment effects which can ultimately not be estimated using a binary variable for railroad access. This leaves room for improvements in future research regarding the relative effect of railroads on growth.

Appendix A: Data Description

TABLE A.1. Control variables.

1. Distance to next node:	Linear distance to the closest node in 100 kilometers in 1848.
2. Street access:	Binary, 1 if 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.
3. Waterway access:	Binary, 1 if a city had at least one cargo ship for river navigation or one seagoing vessel in 1849.
4. Civilian population (log):	Natural logarithm of the resident civilian population in 1849.
5. Military population (log):	Natural logarithm of the resident military population in 1849.
6. Factory workers (share):	Share of total population employed in factories of all kinds in 1849.
7. Mining (county level):	Binary, 1 if the city is located in a county that has a least one steam engine in mining.
8. Large farming (county level):	Share of land holdings larger than 300 Prussian Morgen (roughly 75 hectare) as percentage of the total number of land holdings in 1849.
9. Age composition:	Population younger than 15 years as a percentage of the total population in 1849.
10. School enrollment rate:	Share of children at compulsory school age (6–14) that attended school in 1849.
11. Annual city growth 1831–1837:	Average annual growth of the civilian population as counted in the censuses of 1831 and 1837.
12. Incorporations:	Binary, 1 if a city changed its dimension through incorporation of surrounding parishes in a period.

Notes: Control variables 1 and 2 are coded using maps provided by IEG (2010), control variables 3–10 are digitized from the 1849 census (Statistisches Bureau zu Berlin, 1851–1855), and variables 11–12 are from data provided by Matzerath (1985). For additional information on our Prussian census data see Becker et al. (2014).

TABLE A.2. Matching variables.

13. City size 1837:	Natural logarithm of the total number of civilian inhabitants in 1837.
14. Annual city growth 1821-37:	Average annual growth of the civilian population as counted in the censuses of 1821 and 1837.
15. Merchants:	Number of merchants, hawkers and victual mongers per total population in 1819.
16. Looms:	Number of looms on different fabrics per total population in 1819.
17. Protestants:	Share of population that is Protestant in 1816.
18. Private dwellings:	Number of private dwellings per total population in 1821.
19. Commercial buildings:	Number of manufactories, mills and warehouses per total population in 1821.
20. Insurance value of buildings against fire:	Natural logarithm of the average insurance value in Thaler of buildings insured by the local fire insurance company (*Feuersocietät*) in 1821.

Notes: Matching variables 13–14 are calculated using the data provided by Matzerath (1985), matching variables 15–20 are digitized data from the 1816–1821 censuses (Mützell, 1823–1825). For additional information on our Prussian census data see Becker et al. (2014).

TABLE A.3. Alternative outcome variables.

21. Return on investment:	Ratio of profit to capital invested in the (operating) railroad line.
22. Number of firms 1849 (log):	Natural logarithm of the total number of factories located in the city in 1849.
23. Firm size 1849 (log):	Natural logarithm of the ratio of the number of workers to the number of factories located in the city in 1849.
24. Child–woman ratio 1849:	Ratio of the number of children under 5 to the number of women aged 15–45.
25. Born outside city 1871 (share):	Ratio of the number of city dwellers born outside of the city to the total number of inhabitants in 1871.

Notes: Alternative outcome 21 is digitized from Technisches Eisenbahn-Büreau (1855), alternative outcomes 22–24 are digitized data from the 1849 census Statistisches Bureau zu Berlin (1851–1855), and alternative outcome 25 is digitized from Königliches Statistisches Bureau (1874). For additional information on our Prussian census data see Becker et al. (2014).

References

Acemoglu, Daron, Simon Johnson, and James 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.

Alder, Simon (2014). "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development." Working paper, University of North Carolina at Chapel Hill.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert 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, Jeremy, Michael R. Haines, and Robert A. Margo (2011). "Railroads and the Rise of the Factory: Evidence for the United States, 1850–1870." In *Economic Evolution and Revolution in*

Historical Time, edited by P. Rhode, J. Rosenbloom, and D. Weiman. Stanford University Press, pp. 162–179.
Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2004). "The Railroad to Success: The Effect of Access to Transportation Infrastructure on Economic Growth in China." MIT Working paper.
Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). "On the Road: Access to Transportation Infrastructure and Economic Growth in China." NBER Working Paper No. 17897.
Baum-Snow, Nathaniel (2007). "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122, 775–805.
Becker, Sascha O., Francesco Cinnirella, Erik Hornung, and Ludger Woessmann (2014). "iPEHD—The ifo Prussian Economic History Database." *Historical Methods*, 47, 57–66.
Becker, Sascha O., Erik Hornung, and Ludger Woessmann (2011). "Education and Catch-up in the Industrial Revolution." *American Economic Journal: Macroeconomics*, 3, 92–126.
Bodenhorn, Howard and David, Cuberes (2010). "Financial Development and City Growth: Evidence from Northeastern American Cities, 1790–1870." NBER Working Paper No. 15997.
Boelcke, Willi A. (1996). "Städtewachstum und Eisenbahnentwicklung in deutschen Bundesstaaten." In *Stadt und Verkehr im Industriezeitalter*, edited by Horst Matzerath. Böhlau Verlag, Köln, pp. 23–40.
Borchardt, Knut (1972). *Die Industrielle Revolution in Deutschland*. Piper, München.
Cantoni, Davide (2014). "The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands." *Journal of the European Economic Association*, 13, 561–598.
Cinnirella, Francesco and Erik Hornung (2013). "Landownership Concentration and the Expansion of Education." CEPR Discussion Paper No. 9730.
Dittmar, Jeremiah E. (2011). "Information Technology and Economic Change: The Impact of The Printing Press." *Quarterly Journal of Economics*, 126, 1133–1172.
Donaldson, Dave (2014). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, forthcoming.
Donaldson, Dave and Richard Hornbeck (2013). "Railroads and American Economic Growth: A 'Market Access' Approach." NBER Working Paper No. 19213.
Duranton, Gilles and Matthew A. Turner (2012). "Urban Growth and Transportation." *Review of Economic Studies*, 79, 1407–1440.
Faber, Benjamin (2014). "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *Review of Economic Studies*, 81, 1046–1070.
Fishlow, Albert (1965). *American Railroads and the Transformation of the Antebellum Economy*. Harvard University Press.
Fogel, Robert William (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, Rainer (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.
Fremdling, Rainer (1983). "Germany." In *Railways and the Economic Development of Western Europe, 1830–1914*, edited by Patrick O'Brien. St. Martin's Press, New York, pp. 121–147.
Fremdling, Rainer (1985). *Eisenbahnen und deutsches Wirtschaftswachstum 1840–1879: Ein Beitrag zur Entwicklungstheorie und zur Theorie der Infrastruktur*, 2nd ed. Gesellschaft für Westfälische Wirtschaftsgeschichte, Dortmund.
Fremdling, Rainer and Günther Knieps (1997). "Competition, Regulation and Nationalization: The Prussian Railway System in the Nineteenth Century." In *Selected Cliometric Studies on German Economic History*, edited by John Komlos and Scott Eddie. Steiner, Stuttgart, pp. 136–150.
Gutberlet, Theresa (2014). "Transport Improvements and the Distribution of Industry: The Railroad Boom in Germany's Regions." Working paper, Rensselaer Polytechnic Institute, Troy.
Hahn, Hans-Werner (2005). *Die Industrielle Revolution in Deutschland*, 2nd ed. Oldenbourg, Munich.
Henderson, J. Vernon (2005). "Urbanization and Growth." In *Handbook of Economic Growth, Vol. 1B*, edited by Philippe Aghion and Steven Durlauf, Handbook of Economic Growth, pp. 1543–1591. Amsterdam: Elsevier.

Henning, Friedrich-Wilhelm (1995). *Die Industrialisierung in Deutschland 1800 bis 1914*, 9th ed. UTB, Stuttgart.

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

Keller, Wolfgang and Carol H. Shiue (2013). "The Link Between Fundamentals and Proximate Factors in Development." NBER Working Paper No. 18808.

Keller, Wolfgang and Carol H. Shiue (2014). "Endogenous Formation of Free Trade Agreements: Evidence from the Zollverein's Impact on Market Integration." *Journal of Economic History*, 74, 1168–1204.

Keyser, Erich (1939–1974). *Deutsches Städtebuch-Handbuch städtischer Geschichte*, Vols. 1–5. Kohlhammer, Stuttgart.

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

Königliches Statistisches Bureau (1874). *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871*, Vols. 1–11. Verlag des Königlichen Statistischen Bureaus, Berlin.

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

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

Malanima, Paolo (2010). "Urbanization." In *The Cambridge Economic History of Modern Europe, Vol. 1: 1700–1870*, edited by Stephen Broadberry and Kevin H. O'Rourke. Cambridge University Press, pp. 235–263.

Matzerath, Horst (1985). *Urbanisierung in Preußen 1815–1914*. Kohlhammer/Deutscher Gemeindeverlag, Stuttgart.

Matzerath, Horst (1996). "Verkehr und Stadtentwicklung. Stand und Möglichkeiten historischer Forschung." In *Stadt und Verkehr im Industriezeitalter*, edited by Horst Matzerath. Böhlau Verlag, Köln, pp. VII–XX.

Michaels, Guy (2008). "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System." *Review of Economics and Statistics*, 90, 683–701.

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

Mützell, Alexander August (1823–1825). *Neues Topographisch-Statistisch-Geographisches Wörterbuch des Preussischen Staats*, Vols. 5 and 6. Karl August Kümmel, Halle.

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

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

Pierenkemper, Toni and Richard H. Tilly (2004). *The German Economy during the Nineteenth Century*. Berghahn Books.

Ploeckl, Florian (2012). "It's all in the Mail: The Economic Geography of the German Empire." Working paper, University of Oxford, Department of Economics.

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

Sombart, Werner (1921). *Die deutsche Volkswirtschaft im neunzehnten Jahrhundert und im Anfang des 20. Jahrhunderts*, 5th ed. Georg Bondi, Berlin.

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

Technisches Eisenbahn-Büreau (1855). *Statistische Nachrichten von den Preussischen Eisenbahnen*, Vols. 1–11. Verlag von Ernst and Korn, Berlin.

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

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Online Appendix for “Railroads and Growth in Prussia”