ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 013/26

The Virtuous Cycle Between Skills and Technology

Sascha O. Becker, Christian Dustmann, Hyejin Ku

www.rfberlin.com JANUARY 2026

The Virtuous Cycle Between Skills and Technology

Authors

Sascha O. Becker, Christian Dustmann, Hyejin Ku

Reference

JEL Codes: I24, J24, O14, O33

Keywords: steam engine, technology adoption, diversity, innovation, human capital, productivity

Recommended Citation: Sascha O. Becker, Christian Dustmann, Hyejin Ku (2026): The Virtuous Cycle Between Skills and Technology. RFBerlin Discussion Paper No. 013/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

The Virtuous Cycle Between Skills and Technology*

Sascha O. Becker Christian Dustmann Hyejin Ku†

January 2026

Abstract

We examine the long-term labor market impact of the steam engine, an early general-purpose technology, by linking newly digitized 19th-century records from Prussia to modern German labor market data (1975–2019). Regions with a higher concentration of steam engines per worker in 1875 exhibit higher wages today, primarily because of higher firm productivity and a more skilled workforce. These regions also exhibited greater skill diversity in 1939 and generated more innovations between 1877 and 1918, a pattern that persists to this day. Our findings highlight a lasting, self-reinforcing cycle between technology and skills, set in motion by the steam engine, offering a novel explanation for regional income disparities and their persistence.

JEL Codes: I24, J24, O14, O33

Keywords: steam engine, technology adoption, diversity, innovation, human capital, productivity

---

* We thank participants in various seminars and workshops for helpful comments and suggestions, and the ROCKWOOL Foundation for funding this research. All errors are our own.
† Becker: University of Warwick, Monash University, and RFBerlin; s.o.becker@warwick.ac.uk. Dustmann: University College London and RFBerlin; c.dustmann@ucl.ac.uk. Ku: University College London and RFBerlin; h.ku@ucl.ac.uk.

1. Introduction

A central question in economics is how technological advancements shape the labor market and, conversely, how changes in the labor market influence innovation and technological progress (Acemoglu, 1998, 2002; Goldin and Katz, 1998, 2009; Autor et al., 2003, 2024; Katz and Margo, 2014). With the rapid rise of automation, digitalization, and AI, this question has gained renewed urgency, as it speaks directly to the forces that drive sustained economic growth and underpin persistent regional inequality. To contribute to this debate, we examine the long-term effects of the steam engine, a transformative technology that revolutionized industry in the 19th century (Mokyr, 1992; Allen, 2009). By combining labor market data from German social security records (1975–2019) with newly digitized historical census data dating back to 1875, we offer fresh insights into the dynamics linking steam engine adoption, skill development, and labor market outcomes over 140 years.

The steam engine, as an early general-purpose technology (Bresnahan and Trajtenberg, 1995; Crafts, 2004), provided distinct advantages over traditional energy sources like wind and water by offering a more reliable, scalable, and geographically flexible power solution for industrial processes. Unlike watermills or windmills, which depend on weather and location, steam engines delivered consistent year-round power, allowing factories to be established in previously unsuitable regions. Furthermore, the scalability of steam-powered energy production facilitated the growth of industrial clusters, fostering urbanization and agglomeration economies (Rosenberg and Trajtenberg, 2004; Gutberlet, 2014).

We argue that the steam engine triggered a self-reinforcing cycle of technological progress and skill development with lasting impacts on labor markets. By reducing reliance on traditional artisan skills and enabling the growth of industrial clusters, steam power generated rising demand for specialized knowledge in machine operation, maintenance, and repair (Goldin and Katz, 1998). Unlike hands-on craft work, these new roles required abstract scientific expertise, spurring the emergence of engineering professions (Hanlon, 2025) and the expansion of technical education (Semrad, 2015). As the supply of skilled workers increased, firms had stronger incentives to develop technologies that complemented these skills, further enhancing productivity, as formalized in Acemoglu's (1998, 2002) model of directed technical change.[1] Thus, regions with higher steam engine concentration developed a broader range of workforce skills, which accelerated innovation, attracted additional talent, and set in motion a virtuous cycle yielding lasting economic advantages.

Germany provides an ideal setting for exploring these hypotheses. As a large country with significant regional variation in labor market outcomes, it offers high-quality administrative data and a wealth of historical records. We digitize data on steam engines in 1875 from the business census (Preussische Statistik 1878) and merge it with other historical datasets, including population size, literacy rates, and industry structures. This historical data is combined with modern labor market outcomes from the social security records provided by the Institute for Employment Research (IAB), covering the years 1975 to 2019. To bridge the historical and modern periods, we also incorporate two additional data sources. First, the 1939 Occupational Census (Statistik des Deutschen Reichs, 1942) and the 1939 Non-agricultural Establishment Census (Statistik des Deutschen Reichs, 1943) allow us to assess occupational and industrial

[1] In this model, technological innovation is steered toward sectors with the highest expected returns. When the relative supply of skilled workers rises, two effects follow. First, a classical substitution effect puts downward pressure on their wages. Second, the "directed technology effect" emerges, where a larger skilled labor market encourages innovations that increase their relative productivity. If the market size effect outweighs the substitution effect, this leads to faster growth in skill-complementary technologies.

diversities, which reflect skill diversity. Second, data on high-value patents from 1877 to 1918, collected by Streb et al. (2006), serve as a proxy for regional innovation capacity.

Examining the impact of steam engine prevalence in 1875 on a range of regional outcomes in modern Germany (1975–2019), we find that regions with a higher concentration of steam engines per worker in 1875 exhibit higher wages, a more educated workforce, and more productive firms by 1975, effects that persist through 2019. These effects are substantial: between 1975 and 2019, workers in regions with a one standard deviation higher concentration of steam engines in 1875 earned wages 2.6 to 3.5 percent higher. The results remain robust after controlling for initial population size, initial literacy rates, and geographic characteristics predictive of alternative power sources like wind and water mills (e.g., wind speed at 10 meters and the quadratic of terrain slope), as well as market access (proxied by proximity to navigable water). Additionally, controlling for an extensive list of 1871 characteristics, such as the population growth rate from 1867 to 1871, the share of residents in urban areas, and household size, among others, barely changes our estimates of steam's impact on today's wages. Further conditioning on the 1882 industrial structure—including the shares of employment in mining, manufacturing, and services—if anything, strengthens steam's positive effect on wages.

We also find that a one standard deviation increase in steam engine concentration in 1875 is associated with an 11 percent higher share of high-skilled workers (defined as university graduates or white-collar apprenticeship graduates with high school degrees) between 1975 and 2019. Our analysis further shows that the higher prevalence of steam engines in 1875 is positively associated with firm productivity today, measured by pre-estimated AKM firm fixed effects (Abowd et al., 1999). Investigating education and firm productivity as potential channels through which early steam adoption impacts wages, we find that nearly half of the wage effect is driven by firm productivity, with another 20 percent explained by education. This aligns with the work of Card et al. (2013), which emphasizes differences in firm productivity as a key driver of wage inequality. Our findings suggest that today's regional disparities in firm productivity may, in part, be traced back to the differential adoption of the steam engine, a transformative technology of the 19th century.

We further show that regions with greater steam engine prevalence in 1875 exhibited higher skill diversity by 1939, as measured by both occupational and industrial diversity. These regions also produced more high-value patents between 1877 and 1918 (Streb et al., 2006) and demonstrated greater diversity in patent technology classes. These historical patterns persist to this day, both in terms of skill diversity and regional innovation capacity.

Despite extensive controls for initial economic, demographic, and geographic factors, regions with a higher concentration of steam engines in 1875 may have differed in unobserved ways that also influenced long-term labor market outcomes, potentially confounding the relationship between steam engine prevalence in 1875 and wages from 1975 to 2019. To address this concern, we implement an instrumental variables (IV) strategy using proximity to coal-rich Carboniferous rock strata as an instrument (de Pleijt et al., 2020; Asch, 2005). The IV results confirm our main findings, reinforcing the interpretation of a causal link between steam engine adoption and long-term labor market outcomes.

Overall, our findings document a long-term, self-reinforcing cycle between skills and technology—set in motion by the steam engine—by leveraging unique linkages between historical and modern labor market data spanning over 140 years. This provides a novel explanation for regional income disparities and their persistence.

We relate to several strands of the literature. A substantial body of research has examined the diffusion and impact of steam engines, including seminal contributions by Atack (1979), Atack et al. (1980, 2008), Crafts (2004), Rosenberg and Trajtenberg (2004), Kim (2005), Nuvolari et al. (2011), Gutberlet (2014), de Pleijt et al. (2020), Franck and Galor (2022), Ridolfi et al. (2023), among others.[2] More recently, Hornbeck et al. (2024) show that between 1850 and 1880, as steam power became cheaper, US manufacturing—particularly in lumber and flour milling—expanded more rapidly in counties with limited waterpower potential. Yet high fixed costs and switching barriers kept many incumbent firms reliant on waterpower, resulting in technological leapfrogging as new entrants adopted steam power more readily. Turning to the early twentieth century, Reichardt (2024) introduces a new mechanism—scale bias, the extent to which technological change disproportionately increases the productivity of large relative to small firms—and demonstrates how steam engines and electric motors had distinct effects on firm size and inequality. Building on this literature, we leverage newly linked data connecting the 1875 Prussian census with modern social security records (1975–2019) to examine the long-run relationship between technology and skill formation, providing evidence consistent with theories of technology-skill complementarity (Goldin and Katz, 1998, 2009) and directed technical change (Acemoglu, 1998, 2002).

We also contribute to research linking diversity to economic development (Rosenberg, 1963; Jacobs, 1969; Glaeser et al., 1992; Ciccone and Hall, 1996; Glaeser, 1999). In particular,

[2] Beyond the steam engine, a growing body of research examines how the adoption—or rejection—of new technologies shaped historical development paths. For instance, Caprettini and Voth (2020) analyze how the diffusion of threshing machines triggered riots in 1830s England, prompting institutional reforms (Aidt and Franck, 2015). Hornung (2014) investigates the impact of Huguenot settlement on the productivity of textile manufactories in Prussia. Juhász et al. (2024) study the adoption of mechanized cotton spinning during the First Industrial Revolution in France. Brey (2025) examines the long-term effects of early hydroelectricity adoption in late 19th-century Switzerland, showing how it accelerated the transition from agriculture to manufacturing, with lasting consequences for industrialization.

Fiszbein (2022) connects agricultural diversity in US counties in 1860 to long-run industrialization and income growth over a 140-year period. In our context, the characteristics of steam engine technology facilitated the formation of industrial clusters and attracted a diverse range of skills and industries. In such regions, entrepreneurs could draw on a broader pool of skills, ideas, and inputs, fostering more innovative and productive firms (see, e.g., Weitzman, 1998; Duranton and Puga, 2001; Helsley and Strange, 2002; Hausmann and Hidalgo, 2011). Notably, we identify regional differences in firm productivity—measured using AKM firm fixed effects (Abowd et al., 1999; Card et al., 2013)—as a key channel through which the prevalence of steam engines in 1875 continues to shape wages in the period 1975–2019. This perspective sheds new light on a previously underexplored dimension of the steam engine's long-run economic impact.

We also relate to studies examining the development trajectories of German regions in relation to early industrialization. Becker and Hornung (2025) study Prussian counties and find that proximity to Carboniferous strata increased the wages of ordinary day laborers between 1850 and 1914, with steam engine adoption acting as one of the key mediators. Berbée et al. (2025) show that German regions with a higher share of industrial employment in 1882 had a higher income rank in the early 20th century but a lower one by the late 20th century. Importantly, their analysis spans both Northern Germany (former Prussia) and Southern Germany (including Bavaria). The "reversal of fortune" they identify is largely driven by states like Bavaria, which transitioned from a predominantly agrarian economy before WWII into the richest German state by the late 20th century (Bury et al. 2024). Our study complements this work by focusing within Northern Germany—specifically the former Prussian territories overlapping with modern Germany—and isolating the role of the steam engine in 1875 on wages in 1975-2019.

Several studies document the long-term negative effects of coal proximity and the subsequent economic decline of former mining regions (see, e.g., Rosés and Wolf, 2021; Fritzsche and Wolf, 2023; Brey and Rueda, 2024). Our focus, instead, is on the role of steam technology. We find robust positive effects of steam engines per worker in 1875 on wages between 1975 and 2019. When we additionally control for the 1882 share of employment in mining, the positive effect of steam engine prevalence becomes even stronger, while the coefficient on mining employment itself remains negative—consistent with the existing literature.[3] By contrast, the 1882 manufacturing employment share is positively associated with present-day wages.[4] These results suggest that the long-run effects of steam technology—through the virtuous cycle it set in motion—differ from the effects of coal mining or manufacturing employment alone.

2. Background and Data

2.1 Steam Engine Introduction in Germany

Steam engines were introduced to Germany in the 1770s by Prussian officials to drain mines, replacing horse-driven pumps (Redlich, 1944). After visiting Boulton and Watt in England, they secretly took notes and sketches of the improved atmospheric engine, which led to domestic production in Hettstedt, near Magdeburg. The first engine was used in 1785, but an attempt to replicate Watt's "double-acting" engine in 1786 failed, requiring imports from England along with

[3] This pattern is expected if steam engine adoption (or coal proximity) is positively correlated with mining activity, and mining has long-run adverse effects on regional development.

[4] This may seem contradictory to the negative long-term effect of the 1882 industrial employment share found in Berbée et al. (2025). As discussed above, their analysis covers both Northern and Southern Germany (including Bavaria), while our study focuses specifically on Northern Germany (i.e., the Prussian regions overlapping with modern Germany).

skilled artisans.[5] Initially state-driven for use in state-owned mines, by the early 1800s, private entrepreneurs began adopting steam engines, expanding their use.

In 1820, the first Watt-style "double-acting steam engines" were produced in Germany by Harkort & Company in Wetter, Westphalia. These engines were increasingly used outside the mining sector as power sources for machinery. By 1834, steamship engines were in use, and in December 1835, the first steam-powered train ran between Nuremberg and Fürth. However, overall use of steam engines was limited. A statistical investigation by the Prussian Ministry of Commerce in 1830 revealed that, across the hundreds of Prussian counties, there were only 231 steam engines in total, with a combined horsepower of just 3,670 P.S.—a stark contrast to the 165,000 P.S. of British steam engines in the same year (Banken, 1993).

By 1875, when we measure steam engine adoption, the number had grown to 28,783 steam engines with a total capacity of 632,067 P.S., while Britain had over 2,000,000 P.S. of steam power (Kanefsky, 1979). While in 1830, steam engines were predominantly used in mining (55.2%) and textiles (22.1%), by 1875, they had spread to a much broader range of industries (Rook, 1978).

2.2 Combining Historical and Contemporary Data from Germany: 1875-2019

2.2.1 Historical Data on Steam Engines

We digitized county-level data from the 1875 Business Census in Prussia (Preussische Statistik, 1878), which covers all establishments outside agriculture, with the exception of the public sector,

[5] Different from earlier steam engines, which used steam to push the piston in one direction and relied on atmospheric pressure or a counterweight to reset it, Watt's double-acting engine utilized steam alternately on both sides of the piston. Steam pushed the piston forward, and on the return stroke, it was applied to the opposite side, enabling continuous power generation in both directions. This innovation became a cornerstone of the Industrial Revolution.

and certain service-sector occupations.[6] The data report the stock of "motors" (engines), their types (i.e., steam-powered vs. other), and total horsepower. We also observe total employment by firm size, with firms employing six or more workers classified as "large," and those with five or fewer as "small." Throughout our analysis, we use Prussian county boundaries as of 1871 to link data across time periods.

2.2.2 Contemporary Social Security Data

To measure labor market outcomes in modern Germany, we use a 2% sample of the IAB Employment History (BeH) spanning from 1975 to 2019.[7] The data are cleaned and prepared according to the procedures outlined by Stüber et al. (2023). We capture a snapshot of the data each year on June 30, ensuring we include one spell for each worker. We focus on full-time, regularly employed workers aged between 18 and 65.

As key labor market outcomes, we use the log of daily wages (in 2015 Euros) and skill levels of workers. We define "high-skilled" as those with high school diplomas followed by apprenticeship training or college degrees (either from a university of applied sciences or a university). To measure firm productivity, we use the AKM firm fixed effects (Abowd et al., 1999). Building on the work of Card et al. (2015) and incorporating additional years, Bellmann et al. (2020) estimate AKM firm fixed effects for 1985–2017 across partially overlapping time windows: 1985–1992, 1993–1999, 1998–2004, 2003–2010, and 2010–2017.

[6] Excluded occupations comprise medical doctors, midwives, funeral homes, lawyers, notaries, and cultural enterprises (music, theatre etc.).

[7] Data for East Germany has been available only from 1992 onward.

2.2.3 Other Data Sources

Prussian county-level economic and demographic data. These are sourced from the ifo Prussian Economic History Database (iPEHD), which provides a comprehensive collection of variables for all Prussian counties during the 19th century (see Becker et al., 2014). To capture initial economic development at the time of steam engine adoption, we use data on the total population and literacy rates in 1871. We also consider an extensive list of economic and demographic controls including: population growth rate from 1867 to 1871, the share of residents in urban areas, household size, share of Protestants, share of Jews, share aged below 10, share of females, share born in municipality, and share of individuals of Prussian origin. Further, we use data from the 1882 Occupation Census (Preussische Statistik. 1884/85), from which we compute employment shares in mining, manufacturing, and services, respectively.

Data on county geography. We collect geographic data for each historical county. As a predictor for windmills, we gather data on county-level mean wind speed at a height of 10 meters from the Global Wind Atlas, see Quentel (2023).[8] Following Ashraf et al. (2025), who show a concave relationship between watermills and terrain slope, we also collect data on county-level mean terrain slope (roughness).[9] We compute the distance from each county's centroid to the nearest navigable river to proxy for market access (see Donaldson and Hornbeck, 2016).[10]

[8] See https://globalwindatlas.info/.

[9] Roughness is computed from elevation and shows the degree of irregularity of the surface. It's calculated by the largest inter-cell difference of a central pixel and its surrounding cell.

[10] The locations of navigable rivers are obtained from IEG / A. Kunz 2001: https://www.ieg-maps.de/mapsp/mapw874d.htm.

Historical patents. We utilize patent data from Streb et al. (2006) on 39,343 high-value patents granted in the German Empire between 1877 and 1918. The starting year of 1877 coincided with the establishment of the German patent law, which allowed inventors to apply for patent protection within individual states and across the entire German Empire. Rather than using data on all 311,019 patents granted during this period, Streb et al. (2006) convincingly argue that high-value patents are the most relevant for innovation, as over two-thirds of all German patents granted between 1891 and 1907 were canceled within five years.[11] Following Sullivan (1994), Streb et al. (2006) classify patents that survive at least ten years as high-value patents.

Historical occupations and industries. We digitize data from the 1939 Occupational Census (Statistik des Deutschen Reichs, 1942), which contains three-digit occupational information at the county level. It is important to note that this census took place in May 1939, before the outbreak of World War II (WWII). We also digitize data from the 1939 Non-agricultural Establishment Census (Statistik des Deutschen Reichs, 1943), which provides two-digit industry-level employment data.

Modern patents. We also use patent data covering modern Germany from 1980 to 2014, geocoded by de Rassenfosse et al. (2019). One patent is typically associated with multiple inventors. We assign each inventor's location (longitude and latitude) to a 1871 county and count the total number of inventors linked to each county and year. For each patent, we record the associated technology classes in four-digit IPC codes. If a patent is linked to multiple IPC codes, we use the modal IPC code.

[11] See Streb et al. (2006) for an extensive and enlightening discussion of low-value vs high-value patents.

2.3 Estimation Samples and Geographic Coverage

Data on steam motors in 1875 is available for Prussia, the largest state in the German Empire. **Figure 1** illustrates the number of steam motors per 1,000 workers across the 452 counties of Prussia, using county boundaries as of 1871. Although the map displays all counties of the German Empire (excluding Alsace-Lorraine), the data on steam motors is limited to Prussia. We link this data with information on high-value patents (1877 to 1918), occupation and industry statistics from 1939, and social security data from 1975 to 2019. The geographic coverage of the estimation sample depends on the data source, with the sample representing the intersection of the geographic areas covered by the respective datasets.

Before WWII, the data covered all of Prussia, as its borders remained consistent between 1871 and 1939. However, when linking data on steam motors from 1875 with modern social security data, we work with the intersection of modern-day Germany and historical Prussia. After WWII, Germany lost territories east of the Oder-Neisse line. For our analysis using social security data, we focus on West Germany for 1975-2019.[12]

Table 1 presents the summary statistics. Panel A shows county characteristics around 1875, with 452 counties representing all of Prussia, the same as in Becker and Woessmann (2009). On average, counties had 6.48 steam engines and 112 horsepower per 1,000 workers in 1875. Additionally, 26% of workers were employed in large firms (with six or more employees). It also shows the 1882 employment shares for mining, manufacturing, and services. Panel B presents individual-level data from social security records (1975–2019) for Prussian counties that overlap with modern-day West Germany. Panel C covers county characteristics measured in periods

[12] Our main findings on the long-term impacts of steam prevalence on wages remain invariant when we restrict the sample to 1992-2019 and include East Germany in the sample.

between 1875 and 1975, including high-value patents (1877–1918) and occupation and industry data from 1939. Finally, Panel D shows the modern-day counterparts of panel C.

3. Empirical Framework

3.1. Steam Engines in 1875 and Labor Market Outcomes in 1975-2019

We begin our analysis by relating the labor market outcomes in the period 1975 to 2019 to steam engine prevalence in 1875 based on the following equation:

$$y_{i,c,t} = \alpha + \beta Steam_{c,1875} + \gamma H_{c,1875} + \lambda G_c + \delta X_{i,t} + \phi_t + e_{i,c,t} \quad (1)$$

where $y_{i,c,t}$ is the outcome of interest, such as log wage or high-skilled status of individual i in historical county c (i.e., Prussian county in 1871) and year t. The variable $Steam_{c,1875}$ indicates the number of steam engines per 1,000 workers measured in 1875, while $H_{c,1875}$ is a vector of historical controls measured at a point in time as close as possible to 1875. In our baseline, we include log population and literacy in 1871, while considering progressively more controls, including industry structure. In G_c, we collect geographic variables that may predict power sources that predate steam engines. These include wind speed at 10 meters above ground, which may predict the availability of windmills. We also condition on a quadratic terrain slope, which may predict watermills. Further, we include the distance to the nearest navigable river to proxy for market access. The vector $X_{i,t}$ includes individual-year level controls such as gender and a cubic in age. When we pool multiple years in the same regression, we also control for year FE, ϕ_t. Standard errors are clustered at the level of historical (1871) Prussian counties.

3.2. Steam Engines in 1875 and County-level Outcomes between 1875 and 2019

We also conduct analysis at the historical county level, focusing on different points in time depending on the question addressed. We estimate an equation of the following form:

$$Y_{c,t} = \alpha_1 + \beta_1 Steam_{c,1875} + \gamma_1 H_{c,1875} + \lambda_1 G_c + e_{c,t} \quad (2)$$

where $Y_{c,t}$ is an outcome of interest in historical county c in year t. The variable $Steam_{c,1875}$ and the historical and geographic controls $H_{c,1875}$ and G_c are as defined above. Heteroskedasticity-robust standard errors are reported.

The steam engine, as an early general-purpose technology (Bresnahan and Trajtenberg, 1995; Crafts, 2004), offered a reliable, scalable, and location-independent power source compared to wind and water. Its consistent output enabled the establishment of factories in previously unsuitable areas and supported the growth of industrial clusters, urbanization, and agglomeration economies (Rosenberg and Trajtenberg, 2004; Gutberlet, 2014). We expect these effects to manifest in greater occupational and industrial diversity and higher innovation capacity, and we examine a variety of county-level outcomes, as described below.

Occupational and industrial diversity 1939. We draw on the 1939 Occupation Census, which provides employment data for three-digit occupations and was conducted in May 1939, prior to the outbreak of World War II. For each historical county, we compute the Shannon Diversity Index—commonly used in the ecological literature to measure species diversity (Spellerberg and Fedor, 2003)—and use it as our primary measure of occupational diversity.[13] We also use the 1939 Non-Agricultural Establishment Census, which provides employment data for two-digit industries

[13] The Shannon Index is given by $-\sum_{i=1}^{N} c_i ln c_i$ where c_i is the share of the ith occupation in the regional workforce, and N is the total number of occupations.

outside agriculture, to compute the Shannon Index for industrial diversity. For robustness, we additionally employ the inverse Herfindahl–Hirschman Index (1–HHI) as an alternative measure of diversity.[14]

High-value patents 1877-1918. We measure historical innovation activities using high-value patents (Streb et al., 2006). We examine both the extensive and intensive margins of patenting during this period. We also explore patent diversity using the Shannon index across patent technology classes.

Occupational and industrial diversity in modern data. Following the approach used to measure historical occupational and industrial diversity, we compute the Shannon Index for the modern period (1975–2019) based on three-digit occupation and industry classifications, respectively.

Modern patents. Similarly, using geocoded patent data from de Rassenfosse et al. (2019) for the period 1980 to 2014, we compute three outcomes at the county level: whether any patents were produced (extensive margin), the log number of patents (intensive margin), and patent diversity, measured using the Shannon Index across patent technology classes.

[14] The Shannon Index captures both the number of categories and their relative evenness, responding smoothly to changes in proportions without being overly influenced by large groups. In contrast, the Herfindahl–Hirschman Index (HHI) assigns greater weight to large shares, making it useful for detecting dominance or monopoly structures but less sensitive to smaller categories. We therefore focus on the Shannon Index as our preferred measure of diversity, as it better reflects gradual and continuous variation in diversity rather than concentration or dominance effects.

4. Steam Engines and Labor Market Developments: 1875-2019

4.1 Wages, Skills, and Firm Productivity in 1975-2019

Wages. In **Figure 2**, we plot mean log wages—averaged at the level of historical Prussian county boundaries—in 1975, 1995, and 2015 against the number of steam motors per 1,000 workers in 1875, with the red line indicating the fitted regression line. The graphs reveal a clear positive relationship, which appears strikingly persistent over time.

Table 2 reports the regression results. Column (1) relates individual log wages to the number of steam motors per 1,000 workers in 1875, standardized to have a mean of zero and a standard deviation of one. In column (2), we condition on each county's initial (1871) level of development, proxied by log population size and literacy rate. Column (3) additionally includes geographic characteristics—specifically, a quadratic in terrain slope and wind speed at a height of 10 meters, which may predict alternative power sources such as water and wind—as well as distance to the nearest navigable river (to proxy for market access). In column (4), we further control for a set of 1871 county characteristics, including population growth rate (1867–1871), urbanization rate (share living in urban areas), average household size, share born in the municipality, share of Prussian origin, share Protestant, share Jewish, share female, and share aged below 10. These additional controls do not materially affect the estimated impact of steam engines on log wages, which remains remarkably stable across specifications. Based on columns (1)–(4), we find that a one SD increase in steam engine concentration in 1875 is associated with 2.3–2.6 percent higher wages today (1975–2019).

Steam engines are likely correlated with industrial structure. For instance, they were heavily used in the mining sector. Several studies document long-term negative effects of coal proximity and the economic decline of former mining regions (see, e.g., Rosés and Wolf, 2021; Fritzsche and Wolf, 2023; Brey and Rueda, 2024). If steam engine adoption was positively correlated with mining activity—and mining has persistent adverse effects on regional development—we may underestimate the positive impact of steam engines on modern wages. To examine this possibility, column (5) adds the 1882 share of employment in mining. Consistent with prior research, mining is negatively associated with wages today. Once this variable is included, the estimated effect of steam engines on modern wages increases.

Column (6) adds a control for the 1882 manufacturing employment share. While the estimated effect of steam engines remains virtually unchanged, the positive coefficient on this variable indicates a positive association between historical manufacturing activity and modern wages.[15] At first glance, our results may appear at odds with the negative long-term effect of the 1882 industrial employment share documented by Berbée et al. (2025), who find that regions more industrialized in 1882 had higher income ranks in the early 20th century but lower ones by the late 20th century. Their analysis, however, spans both Northern and Southern Germany (including Bavaria), with the "reversal of fortune" largely driven by Bavaria's postwar transformation from an agrarian to a high-income economy (Bury et al., 2024). Focusing instead on Northern Germany (the former Prussian regions overlapping with modern Germany), our study complements their findings by isolating the long-run wage effects of early steam adoption in 1875 between 1975 and 2019. Column (7) introduces the 1882 service employment share as an additional control. The estimated relationship between steam engine prevalence and modern wages remains essentially unchanged.

[15] This pattern remains the same even when we include the 1882 manufacturing share alone in the wage regression, or alongside steam engines only (without the 1882 mining employment share).

Overall, Table 2 suggests that regions with a one SD higher share of steam engines in 1875 have wages 2.6 to 3.5 percent higher in 1975–2019, 100–140 years later.[16] As shown in **Table A1**, the results are persistent when looking at subperiods: 1975-1994, 1995-2010, and 2011-2019. Moreover, in **Table A2**, we find similar patterns when using other proxies of steam motors such as horsepower per worker or share of workers employed at large firms.

Education. The effects on wages could be mediated by education or firm productivity. We therefore next examine whether regions with a higher prevalence of steam engines in 1875 have more skilled workers and whether there are corresponding differences in firm productivity. In **Table 3**, we report the impact of a one SD higher share of steam engines in 1875 on the share of highly skilled workers over the period 1975-2019, where we control for the same set of variables as in column (7) of Table 2. High-skilled workers are, as described in Section 3, defined as those (i) with a high-school degree and an apprenticeship or (ii) with a college degree (university of applied sciences or university).[17] Our estimates in Table 3, column (1), suggest a positive association between steam engine concentration in 1875 and the skill level today, with a one SD higher steam engine prevalence associated with an 11 percent (0.020/0.187) higher share of skilled workers.

While no comprehensive historical datasets on educational attainment in Germany in the early 20th century exist, to understand the evolution of skill acquisition across high- and low-steam

[16] Here, standard errors are clustered by historical county. In Appendix **Table A3**, we show results using Conley (1999) standard errors, which account for potential spatial correlation. The results are very similar in magnitude to those reported in the main table.

[17] The share of high-skilled workers may seem low (mean of the dependent variable of 0.187 over the years 1975-2019). Note that Germany has a three-track secondary education system (Hauptschule, Realschule, Gymnasium), with a large share of pupils attending the two lower tracks, and some of those completing the upper track (Gymnasium) enter the labor market while neither completing an apprenticeship nor going to university. Our definition of high-skilled captures those completing education beyond Gymnasium.

regions over time, we exploit modern social security data, which allow us to compare different birth cohorts. As shown in columns (2)-(6), the share of individuals with high skills increases from 6.7% for pre-1940 birth cohorts to 31.4% for 1970+ birth cohorts. While the point estimates in columns (2)-(6) are increasing over time, relative to the mean of the dependent variable, the implied effects are quite stable over time: the effect of 1875 steam on skills ranges from 7.6% (0.024/0.314) for the 1970+ birth cohorts to 13.3% (0.015/0.113) for the 1940-1949 birth cohorts. Overall, a higher share of steam engines in 1875 is associated with a higher share of better-educated workers more than 100 years later, consistent with the virtuous cycle between technology and skills.

Firm Productivity. One reason for higher wages in areas with a higher share of steam engines per worker in 1875 is that workers are better educated, as shown above. Another potential explanation is employment at more productive firms. To measure firm productivity, we use firm fixed effects from AKM regressions (Abowd et al., 1999; Card et al., 2013), as estimated by Bellmann et al. (2020), which are available for the periods 1985–1992, 1993–1999, 1998–2004, 2003–2010, and 2010–2017 (1985 being the first year for which AKM firm fixed effects are available; see Section 2).

Table 4 reports estimates using data from 1975-2017 (column (1)) and splits the time window by the periods over which AKM effects were estimated (columns (2)-(6)). The estimates in column (1) suggest that a one SD higher share of steam engines is associated with a 1.5 percent increase in firm productivity, measured by the wage premiums paid to workers. This indicates that regions with a higher concentration of steam engines in 1875 have more productive firms (based on AKM effects) a century later. Estimates across the sub-periods are similar, ranging from 1.2 to 1.7 percent (columns (2)–(6)).

Sources of Regional Wage Differences. The results so far indicate that the higher wages observed in areas with greater steam engine concentration in 1875, as shown in Table 2, may be explained by a higher share of highly educated workers (Table 3) or by more productive firms (Table 4). To better understand the relative contribution of these two factors, we present estimates in **Table 5**, where we regress log wages on steam motors and include the share of high-skilled workers and AKM firm fixed effects as controls. As before, all regressions control for year fixed effects, demographic controls such as gender and a cubic in age, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882 (same as in column (7) of Table 2).

Table 5 reports estimates for the 1985-2017 period since AKM firm fixed effects are only available for this time window. Column (1) repeats the estimates from Table 2 but is restricted to 1985-2017, with estimates similar to those in column (7) of Table 2. Column (2) adds a dummy variable indicating whether the individual is highly skilled, as defined in Section 2. The coefficient on steam engines decreases by 22 percent (from 0.036 to 0.028), suggesting that education plays some role in explaining wage differences due to the adoption of steam engines in 1875. Column (3) controls for AKM firm fixed effects instead, with the coefficient on steam engines decreasing to 0.018, indicating that firm productivity is a key factor driving higher wages in regions with higher steam concentration. Finally, Column (4) incorporates both the high skilled status and AKM firm fixed effects. The estimates suggest that two thirds of the overall long-term impact of steam engine adoption is channeled through higher education and firm productivity. Firm productivity is the primary channel through which steam engine adoption in 1875 affects wages more than a century later, contributing about 47 percent, while the education channel accounts for 19 percent.[18] About one third of the impact of steam adoption on today's wages remains unexplained by these two channels.

4.2 Occupational and Industrial Diversity in 1939 and 1975-2019

As steam engines fostered agglomeration economies (Rosenberg and Trajtenberg, 2004; Gutberlet, 2014), regions with higher steam engine concentration in 1875 may have attracted a more diverse set of skills. To investigate this, we use measures of occupational and industrial diversity from 1939.

Figure 3 plots 1939 occupational diversity—measured by the Shannon Index based on employment shares in three-digit occupations—against steam engines per 1,000 workers in 1875. There is a clear positive relationship between later occupational diversity and steam engine concentration in 1875, with the red line representing the regression line. A similar pattern is observed for occupational diversity in 1975, 1995, and 2015, as shown in **Figure 4**.

Table 6 reports regression results looking at various aspects of diversity, where columns (1) to (3) focus on occupational diversity in 1939 while columns (4) to (6) focus on industrial

[18] To compute the relative contributions of the channels "skill (education)" and "firm productivity (AKM)", we rely on a simple omitted variable bias formula. Omitting skills and firm productivity from the regression as in column 1 of Table 5, the "contribution" in the estimate of the impact of steam on wages through education is $b_{ed} = \beta_{ed}\left[\rho_{S,ed}\frac{\sigma_{ed}}{\sigma_S} - \rho_{S,AKM}\rho_{ed,AKM}\frac{\sigma_{AKM}}{\sigma_S}\right]$, and through AKM is $b_{AKM} = \beta_{AKM}\left[\rho_{S,AKM}\frac{\sigma_{AKM}}{\sigma_S} - \rho_{S,ed}\rho_{ed,AKM}\frac{\sigma_{ed}}{\sigma_S}\right]$. Here β_{ed} and β_{AKM} are the partial regression coefficients on skill and firm productivity effects in Column 4, $\rho_{S,AKM}, \rho_{S,ed}$ and $\rho_{ed,AKM}$ are the correlation coefficients between steam and AKM, steam and education, and education and AKM, and $\sigma_{ed}, \sigma_{AKM}, \sigma_S$ are the standard deviations of education, AKM and steam. The contribution of steam through education and AKM effects are then $\frac{b_{ed}}{\beta_S}$ and $\frac{b_{AKM}}{\beta_S}$, respectively. Here β_S is the estimate of steam motors in column 1 (0.036).

diversity outside agriculture in 1939. Agriculture still played an important role in the German economy in 1939, with an employment share of ca. 40%. We thus first study how steam engine adoption relates to the share of employment in agriculture. Column (1) shows that a one SD higher number of steam engines per 1,000 workers in 1875 is associated with a 1.7 percentage point (4.2 percent) lower share of employment in agriculture in 1939, alongside greater occupational diversity. In column (2), a one SD higher steam engine concentration in 1875 is associated with a 3 percent increase in occupational diversity in 1939. When occupational diversity is measured outside agriculture only, the effect size is approximately 1 percent (column (3)). This suggests that regions with a higher share of steam engines in 1875 moved away faster from agricultural production in the first half of the 20th century and exhibited greater occupational diversity.

In columns (4)–(6), we use two-digit industry data (excluding agriculture) as an alternative way to examine skill diversity in 1939 in regions with higher steam engine adoption in 1875. Column (4) shows a larger share of jobs in manufacturing sub-sectors, as opposed to service or government-related sectors. Column (5) indicates that areas with a one SD higher steam engine concentration in 1875 have 1.1 percent higher diversity of two-digit sectors outside agriculture. When focusing on industrial diversity within manufacturing, the effect size rises to 2.5 percent (column (6)).[19]

Table 7 shows similar results using modern data. Between 1975 and 2019, the manufacturing share of employment is 2.9 percentage points (8.5 percent) higher in regions that had a one SD higher steam engine prevalence in 1875 (column (1)). These regions also exhibit greater occupational diversity, measured using three-digit occupations, both overall and within manufacturing (columns (2)–(3)). In columns (4) and (5), we measure industrial diversity at the

[19] The results remain robust when using the inverse Herfindahl-Hirschman Index (HHI) as an alternative measure of diversity (see **Table A4**).

four-digit sector level. A one SD higher number of steam engines per 1,000 workers in 1875 is associated with 2.1 percent higher industrial diversity for overall industries (column (4)) and 3.7 percent higher diversity within manufacturing (column (5)). Together, these results strongly support the notion that steam engine adoption fostered a more diverse industry and occupational structure decades later.

4.3 Innovations in 1877-1918 and 1980-2014

Patent data 1877-1918. We hypothesized in Section 2 that the steam engine concentration in 1875 created agglomeration economies with higher demand for skills and skill diversity, which in turn spurned innovation activity (Rosenberg, 1963; Jacobs, 1969; Glaeser et al., 1992; Ciccone and Hall, 1996; Glaeser, 1999). Drawing on high-value patent data collected by Streb et al. (2006), we regress total patents per county over the period from 1877 to 1918 on steam engines per 1,000 workers.

The estimates are presented in **Table 8**. Using the same set of controls as in earlier tables (column (7) of Table 2), including employment shares by industry in 1875, we first examine the extensive margin of having any patents (panel A). It is important to note that in the early years of the German patenting system, high-value patents were far from ubiquitous. Over the entire period 1877–1918, only 74.3% of the 452 Prussian counties had any high-value patents. Nevertheless, the share of counties with any high-value patents increased from 52.0% in 1877–1899 to 66.8% in 1900–1918. In the first sub-period (1877–1899), a one SD higher steam engine concentration is associated with an 8.4 percentage point higher probability that a county has any high-value patents. This effect on the extensive margin is smaller in the later period (1900–1918), as shown in column (3).

We then examine the intensive margin in panel B for the sub-sample of counties with any high-value patents. We find that a higher concentration of steam engines in 1875 predicts more patents over 1877–1918 and during the first sub-period (1877–1899), but not in the second sub-period (1900–1918). Our estimates indicate that a one SD higher share of steam engines per 1,000 workers is associated with 13.5% more patents—a substantial increase. This suggests that innovation activity is a key mechanism linking historical steam engine concentration to future economic advantage. In panel C, we also examine patent diversity across categories and find that higher steam engine concentration in 1875 predicts greater patent diversity.

Modern patents data (1980-2014). Using geocoded patent data from de Rassenfosse et al. (2019), we also analyze modern innovation capacity (1980–2014). In **Table 9**, we follow the same structure as Table 8: panel A examines the extensive margin of patenting, panel B the intensive margin, and panel C patent diversity across four-digit technology classes. Since nearly all counties exhibit patenting activity during 1980–2014, the estimates in panel A are close to zero. Yet, at the intensive margin (panel B), steam engine concentration in 1875 continues to predict higher patenting activity, with a one SD higher share of steam engines per 1,000 workers associated with 29% more patents—a substantial effect. Finally, steam engine concentration in 1875 also predicts greater patent diversity (panel C) in 1980–2014 and across our three sub-periods. Overall, historical steam engine adoption predicts significantly higher patenting activity throughout the 20th century and beyond.

5. Sources of Regional Variation in Steam Engine Adoption

We condition on extensive controls for initial economic, demographic, and geographic factors, variables that might otherwise induce a spurious correlation between modern-day wages and steam engine adoption when omitted. As Table 2 shows, including these controls successively affects the estimated coefficients of steam adoption on modern wages only marginally. Nevertheless, to address remaining concerns that regions with a higher concentration of steam engines in 1875 may have differed in *unobserved* characteristics that also shaped long-term labor market outcomes, potentially confounding the relationship between steam engine prevalence in 1875 and wages from 1975 to 2019, we implement an instrumental variables (IV) strategy. Following de Pleijt et al. (2020), we use proximity to Carboniferous rock strata (Asch, 2005) as an instrument for the number of steam engines per 1,000 workers installed in 1875.[20] De Pleijt et al. (2020) demonstrate that while the share of Carboniferous rock strata in a county is strongly correlated with the number of steam engines per person, the concentration of these rock strata is independent of pre-industrial indicators of development and human capital formation in the county. Similarly, Fernihough and O'Rourke (2021) show that Carboniferous rock strata predict active coal fields, which in turn provided the fuel necessary for steam engines.[21] **Figure A1** shows the locations of coal deposits and of Carboniferous rock strata. It is worth noting that coal deposits can exist without

[20] de Pleijt et al. (2020) explain that "[c]oal is found in rock layers from the Carboniferous age, which were created more than 300 million years ago. During this era, large forests covered the areas that later formed the earth's coal layers. The coalfields that supplied the emerging industries during the early phases of the Industrial Revolution therefore appeared near to rock strata from the Carboniferous epoch."

[21] Alternatively, since proximity to Carboniferous rock strata predicts proximity to active coal fields (Fernihough and O'Rourke, 2021), the latter constitutes another potential instrument. Yet, active coalfields are more likely to suffer from endogeneity concerns, whereas Carboniferous rock strata are geological features determined millions of years ago.

Carboniferous rock strata, for example, in former swampy lowlands where coal occurs as lignite and can be extracted at the surface.

Figure 5 shows the relationship between distance to coal and steam engine concentration in 1875 (panel A) and between distance to Carboniferous rock strata and steam engine concentration in 1875 (panel B). As expected, both relationships are negative. The slope (SE) of the regression line in panel A is -0.015 (0.002), indicating that proximity to coal deposits was a significant factor facilitating steam engine adoption.

Table 10 presents the corresponding regression results. Column (1) reproduces the OLS estimates from Table 2, column (3). Columns (2)–(4) use distance to Carboniferous rock strata as an instrument, while columns (5)–(7) use distance to coal. Column (2) reports the first-stage estimates, column (3) the reduced-form estimates, and column (4) the IV estimates. The IV coefficients in columns (4) and (7) are similar, with magnitudes roughly 1.8 times larger than the OLS estimate in column (1).

One interpretation of these IV coefficients is that the local average treatment effect (LATE) reflects particularly high returns to steam engine adoption in areas that adopted steam engines *because of* their proximity to Carboniferous rock strata, or to coal deposits.[22] Alternatively, the larger IV estimates may reflect measurement error in the steam engine variable. Regardless, these results suggest that the OLS estimates are, if anything, downward biased, providing a lower bound for the effects of steam engine adoption on log wages.

[22] See Imbens and Angrist (1994) for discussion on treatment effects for "compliers" to instruments in the LATE framework.

6. Discussion and Conclusion

In this paper, we merge newly digitized 1875 German census data with contemporary social security records (1975–2019) to examine the long-term links between steam engine adoption, skill accumulation, and labor market outcomes over 140 years. As an early general-purpose technology, the steam engine provided reliable, scalable, and geographically flexible energy that fostered industrial clusters, urbanization, and agglomeration economies (Rosenberg & Trajtenberg, 2004; Gutberlet, 2014).

We argue that steam engines triggered a self-reinforcing cycle of technological progress and skill formation. Fully exploiting steam power and the industrial clusters it supported required skilled labor and skill diversity, which in turn fostered further innovation, reflecting technology-skill complementarities (Goldin & Katz, 1998, 2009) and directed technical change (Acemoglu, 1998, 2002). Therefore, regions with higher early concentrations of steam engines likely developed a broader range of workforce skills, which in turn accelerated innovation, attracted additional talent, and reinforced a self-sustaining cycle between technology and skills.

Our empirical results support this mechanism. Regions with higher steam engine shares per worker in 1875 continue to have higher wages today, driven by both higher educational attainment and greater firm productivity—the latter playing a particularly important role in explaining wage disparities between regions with higher versus lower steam engine concentrations. This is consistent with the key role of firm productivity in determining wages in modern data (Card et al., 2013).

Bridging the long interval between historical steam engine adoption (1875) and modern labor market outcomes (1975–2019), we show that regions with more steam engines also had higher patent activity and diversity around the turn of the 20th century. These innovation patterns persist today: areas with historically higher steam engine prevalence exhibit greater patent intensity and diversity. Using the 1939 German census as a midpoint, we find that these regions also had higher occupational and industrial diversity and a smaller agricultural sector—patterns that remain observable today.

Importantly, the persistence of steam engine effects on modern wages does not contradict standard convergence theories (Barro & Sala-i-Martin, 1992), which focus on relative wage growth. Instead, we highlight the long-term impact of a key general-purpose technology on regional development, with particular attention to the dynamics between technology and skill development. Interestingly, while wage convergence occurred between 1975 and 1995 (see **Figure A2**), regional wages diverged from 1995 to 2015, echoing patterns observed in the United States (Blanchard & Katz, 1992; Ganong & Shoag, 2017), suggesting avenues for future research.

Finally, the enduring effects of 19th-century steam adoption on today's wages suggest that the same mechanisms—linking technology, skills, and innovation—may shape the long-term trajectories of regions leading in digital and artificial intelligence technologies.

References

Abowd, John, Francis Kramarz and David Margolis. 1999. “High Wage Workers and High Wage Firms.” *Econometrica* 67(2): 251-334.

Acemoglu, Daron. 1998. “Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality.” *Quarterly Journal of Economics* 113(4): 1055–1089.

Acemoglu, Daron. 2002. “Directed Technical Change.” *Review of Economic Studies* 69(4): 781–809.

Aidt, Toke S., and Raphaël Franck. 2015. “Democratization under the Threat of Revolution: Evidence from the Great Reform Act of 1832.” *Econometrica* 83 (2): 505–547.

Allen, Robert C. 2009. *The British Industrial Revolution in Global Perspective*, Cambridge and New York: Cambridge University Press.

Asch, Kristine. 2005. *The 1:5 Million International Geological Map of Europe and Adjacent Areas*, Hannover: Bundesanstalt für Geowissenschaften und Rohstoffe.

Ashraf, Quamrul, Francesco Cinnirella, Oded Galor, Boris Gershman, Erik Hornung. 2025. “Structural Change, Elite Capitalism, and the Emergence of Labour Emancipation.” *Review of Economic Studies* 92(2): 808-836.

Atack, Jeremy. 1979. “Fact in Fiction? The Relative Costs of Steam and Water Power: A Simulation Approach.” *Explorations in Economic History* 16(4): 409-437.

Atack, Jeremy, Fred Bateman, and Thomas Weiss. 1980. “The Regional Diffusion and Adoption of the Steam Engine in American Manufacturing.” *Journal of Economic History* 40(2): 281-308.

Atack, Jeremy, Fred Bateman, and Robert A. Margo. 2008. "Steam Power, Establishment Size, and Labor Productivity Growth in Nineteenth Century American Manufacturing." *Explorations in Economic History* 45(2): 185-198.

Autor, David H., Frank Levy, Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, 118(4): 1279–1333.

Autor, David, Caroline Chin, Anna Salomons, and Bryan Seegmiller. 2024. "New Frontiers: The Origins and Content of New Work, 1940–2018." *Quarterly Journal of Economics* 139(3): 1399-1465.

Banken, Ralf. 1993. "Die Diffusion der Dampfmaschine in Preußen um 1830." *Jahrbuch für Wirtschaftsgeschichte* 34(2): 219-248.

Barro, Robert J., and Xavier Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100(2): 223-251.

Becker, Ann-Kristin, and Erik Hornung. 2025. "Industrialization and the Return to Labor: Evidence from Prussia." Working Paper.

Becker, Sascha O., Francesco Cinnirella, Erik Hornung, and Ludger Woessmann. 2014. "iPEHD—The ifo Prussian Economic History Database." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 47:2: 57-66

Bellmann, Lisa, Ben Lochner, Stefan Seth, and Stefanie Wolter. 2020. AKM Effects for German Labour Market Data. *FDZ-Methodenreport* 01/2020, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Berbée, Paul, Sebastian Till Braun, and Richard Franke. 2025. "Reversing Fortunes of German Regions, 1926–2019: Boon and Bane of Early Industrialization?." *Journal of Economic Growth* 30(2): 307-337.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 1992(1): 1-75.

Bresnahan, Timothy F., and Manuel Trajtenberg. 1995. "General Purpose Technologies: 'Engines of Growth'?." *Journal of Econometrics* 65(1): 83-108.

Brey, Björn. 2025. "Sparking Knowledge: Early Technology Adoption, Innovation Ability and Long-Run Growth." *Journal of the European Economic Association*, forthcoming.

Brey, Björn, and Valeria Rueda. 2024. "The Death of King Coal and the Scars of Deindustrialization." CEPR Discussion Paper 19082.

Bury, Yannik, Lars P. Feld and Heiko T. Burret. 2024. "Skimming the Achieved? Quantifying the Fiscal Incentives of the German Fiscal Equalization Scheme and Its Reforms Since 1970." *Economic Governance* 25: 291–334.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128(3): 967–1015.

Card, David, Jörg Heining, and Patrick Kline. 2015. "CHK Effects." *FDZ-Methodenreport* 6: 2015.

Caprettini, Bruno, and Hans-Joachim Voth. 2020. "Rage against the Machines: Labor-saving Technology and Unrest in Industrializing England." *American Economic Review: Insights* 2(3): 305–320.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86(1): 54-70.

Crafts, Nicholas. 2004. "Steam as a General Purpose Technology: a Growth Accounting Perspective." *The Economic Journal* 114(495): 338–351.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92(1): 1-45.

de Pleijt, Alexandra, Alessandro Nuvolari, and Jacob Weisdorf. 2020. "Human Capital Formation During the First Industrial Revolution: Evidence from the Use of Steam Engines." *Journal of the European Economic Association* 18(2): 829–889.

de Rassenfosse, Gaétan, Jan Kozak, and Florian Seliger. 2019. "Geocoding of Worldwide Patent Data." *Scientific Data* 6: article 260.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *Quarterly Journal of Economics* 131(2): 799-858.

Duranton, Gilles, and Diego Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Review* 91(5): 1454-1477.

Fernihough, Alan, and Kevin Hjortshøj O'Rourke. 2021. "Coal and the European Industrial Revolution." *The Economic Journal* 131(635): 1135–1149.

Fiszbein, Martin. 2022. "Agricultural Diversity, Structural Change, and Long-run Development: Evidence from the United States." *American Economic Journal: Macroeconomics* 14(2): 1-43.

Franck, Raphaël, and Oded Galor. 2022. "Technology-skill Complementarity in Early Phases of Industrialisation." *The Economic Journal* 132(642): 618-643.

Fritzsche, Miriam, and Nikolaus Wolf. 2023. "Fickle Fossils. Economic Growth, Coal and the European Oil Invasion, 1900-2015." CESifo Working Paper No. 10805.

Ganong, Peter, and Daniel Shoag. 2017. “Why Has Regional Income Convergence in the US Declined?.” *Journal of Urban Economics* 102: 76-90.

Glaeser, Edward L. 1999. “Learning in Cities.” *Journal of Urban Economics* 46(2): 254-277.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. “Growth in Cities.” *Journal of Political Economy* 100(6): 1126-1152.

Goldin, Claudia, and Lawrence F. Katz. 1998. “The Origins of Technology-Skill Complementarity.” *Quarterly Journal of Economics* 113(3): 693-732.

Goldin, Claudia, and Lawrence F. Katz. 2009. *The Race between Education and Technology*. Harvard University Press.

Gutberlet, Theresa. 2014. “Mechanization and the spatial distribution of industries in the German Empire, 1875 to 1907.” *Economic History Review* 67(2): 463-491.

Hanlon, W. Walker. 2025. “The Rise of the Engineer: Inventing the Professional Inventor during the Industrial Revolution.” *The Economic Journal* 135(670): 1749–1781.

Hausmann, Ricardo, and César A. Hidalgo. 2011. “The Network Structure of Economic Output.” *Journal of Economic Growth* 16(4): 309-342.

Helsley, Robert W., and William C. Strange. 2002. “Innovation and Input Sharing.” *Journal of Urban Economics* 51(1): 25-45.

Hornbeck, Richard, Anders Humlum, Shanon Hsuan-Ming Hsu and Martin Rotemberg. 2024. “Gaining Steam: Incumbent Lock-in and Entrant Leapfrogging.” *NBER Working Paper* 32384.

Hornung, Erik. 2014. “Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia.” *American Economic Review* 104(1): 84-122.

Imbens, Guido W., and Joshua D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62(2): 467-475.

Jacobs, Jane. 1969. *The Economy of Cities*. New York: Vintage.

Juhász, Réka, Mara P. Squicciarini, and Nico Voigtländer. 2024. “Technology Adoption and Productivity Growth: Evidence from Industrialization in France.” *Journal of Political Economy* 132(10): 3215-3259.

Kanefsky, John W. 1979. “The Diffusion of Power Technology in British Industry.” *unpublished Ph.D. thesis*, University of Exeter.

Katz, Lawrence F. and Margo, Robert A. 2014 “Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective.” *Human Capital in History: The American Record*, edited by Leah Platt Boustan, Carola Frydman and Robert A. Margo, Chicago: University of Chicago Press, pp. 15-58.

Kim, Sukkoo. 2005. “Industrialization and Urbanization: Did the Steam Engine Contribute to the Growth of Cities in the United States?.” *Explorations in Economic History* 42(4): 586-598.

Mokyr, Joel. 1992. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press.

Nuvolari, Alessandro, Bart Verspagen, and Nick von Tunzelmann. 2011. “The Early Diffusion of the Steam Engine in Britain, 1700–1800: A Reappraisal.” *Cliometrica* 5: 291-321.

Preussische Statistik. 1878. “Die definitiven Ergebnisse der Gewerbezählung vom 1. December 1875 im preussischen Staate. I. Theil. Die Gewerbebetriebe in den einzelnen Verwaltungsbezirken, Kreisen und grösseren Städten des preussischen Staates. ” *Preussische Statistik* Band 40. Berlin: Verlag des Königlichen Statistischen Bureaus.

Preussische Statistik. 1884/85. „Die Ergebnisse der Berufsstatistik vom 5. Juni 1882 im preussischen Staat.“ *Preussische Statistik* Vol. 76 a-c. Berlin: Verlag des Königlichen Statistischen Bureaus.

Quentel, Milan. 2023. “Gone with the Wind: Renewable Energy Infrastructure, Welfare, and Redistribution.” UPF Working Paper.

Redlich, Fritz. 1944. “The Leaders of the German Steam-Engine Industry During the First Hundred Years.” *Journal of Economic History* 4(2): 121–148.

Reichardt, Hugo. 2024. “Scale-Biased Technical Change and Inequality.” LSE Working Paper.

Ridolfi, Leonardo, Carla Salvo, and Jacob Weisdorf. 2023. “The Effect of Mechanisation on Labour: Evidence from the Diffusion of Steam.” CAGE Working Paper 689.

Rook, Hans-Joachim. 1978. “Tendenzen der räumlichen Verteilung und Entwicklung der auf dem Gebiet des Deutschen Reiches in der gewerblichen Produktion eingesetzten Dampfmaschinen im Vergleich der Jahre 1846/47 und 1875.” *Jahrbuch für Wirtschaftsgeschichte* 19(1): 119-146.

Rosenberg, Nathan. 1963. “Technological Change in the Machine Tool Industry, 1840–1910.” *Journal of Economic History* 23(4): 414-443.

Rosenberg, Nathan and Manuel Trajtenberg. 2004. “A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States.” *Journal of Economic History* 64(1): 61-99.

Rosés, Joan R., and Nikolaus Wolf. 2021. “Regional Growth and Inequality in the Long-run: Europe, 1900–2015.” *Oxford Review of Economic Policy* 37(1): 17-48.

Semrad, Alexandra. 2015. “Modern Secondary Education and Economic Performance: The Introduction of the Gewerbeschule and Realschule in Nineteenth-Century Bavaria.” *Economic History Review* 68(4): 1306–1338.

Spellerberg, Ian F., and Peter J. Fedor. 2003. “A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous use of Species Richness, Species Diversity and the ‘Shannon–Wiener’ Index.” *Global Ecology and Biogeography* 12(3): 177-179.

Statistik des Deutschen Reichs. 1942 “Volks-, Berufs- und Betriebszählung vom 17. Mai 1939: Die Berufstätigkeit der Bevölkerung in den Reichsteilen.” *Statistik des Deutschen Reichs* Band 557, Verlag für Sozialpolitik, Wirtschaft und Statistik, Paul Schmidt.

Statistik des Deutschen Reichs. 1943 “ Volks-, Berufs- und Betriebszählung vom 17. Mai 1939: Die nichtlandwirtschaftlichen Arbeitsstätten in den Reichsteilen und Verwaltungsbezirken.” *Statistik des Deutschen Reichs* Band 568, Verlag für Sozialpolitik, Wirtschaft und Statistik, Paul Schmidt.

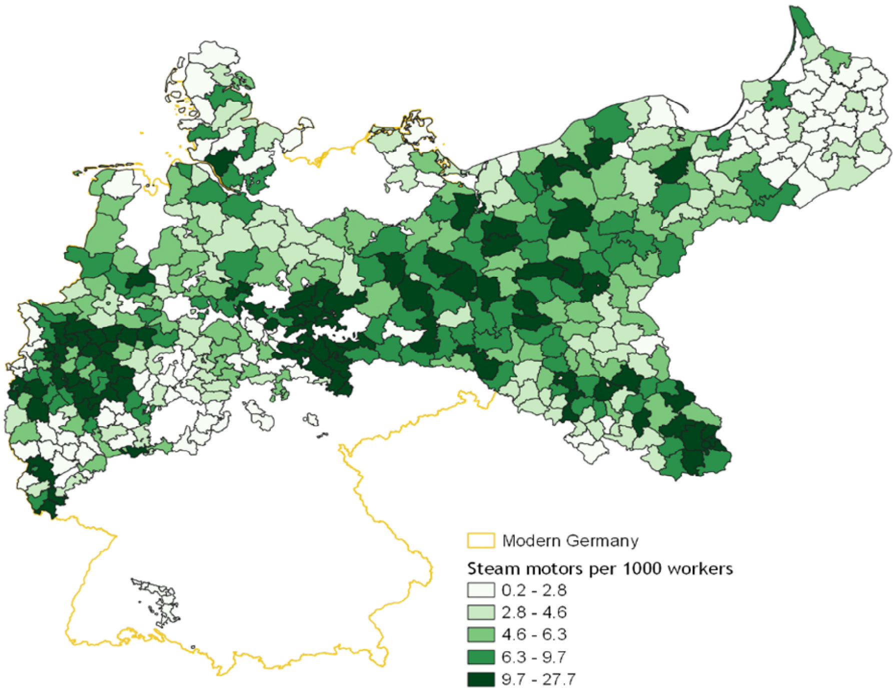
Streb, Jochen, Jörg Baten, and Shuxi Yin. 2006. “Technological and Geographical Knowledge Spillover in the German Empire 1877–1918.” *Economic History Review*, 59(2): 347-373.

Stüber, Heiko, Wolfgang Dauth, and Johann Eppelsheimer. 2023. “A Guide to Preparing the Sample of Integrated Labour Market Biographies (SIAB, Version 7519 v1) for Scientific Analysis.” *Journal for Labour Market Research* 57: 1-7.

Sullivan, Richard J. 1994. “Estimates of the Value of Patent Rights in Great Britain and Ireland, 1852–1976.” *Economica* 61(241): 37–58.

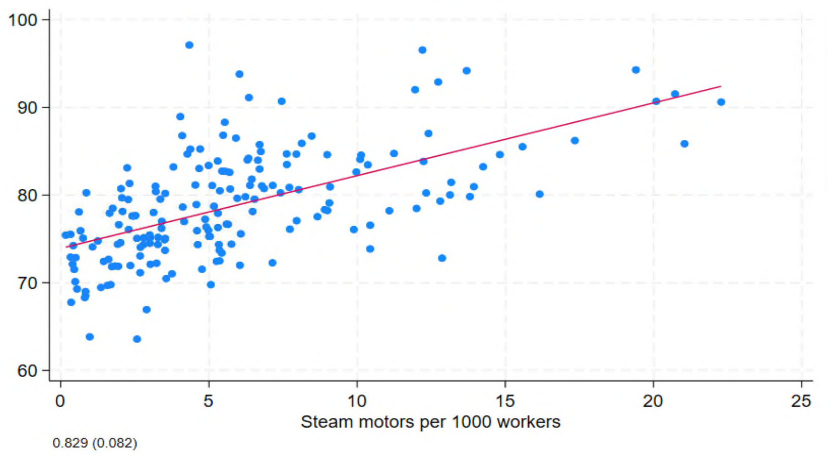
Weitzman, Martin L. 1998. “Recombinant Growth.” *Quarterly Journal of Economics* 113(2): 331-360.

Figure 1. Steam engines in Prussian counties in 1875

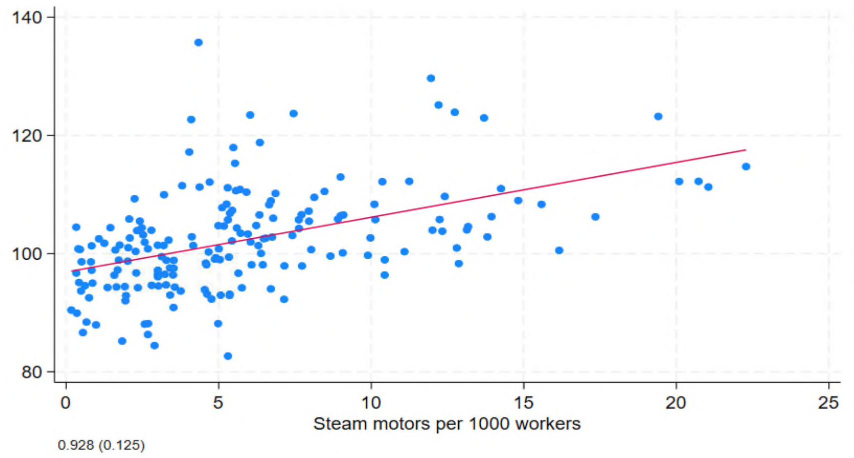


Notes: This figures shows the spatial distribution of steam engines per 1000 workers in 1875. Data on steam engines are available at the level of year 1871 Prussian counties. The figure also shows the boundary of modern Germany in 2019.

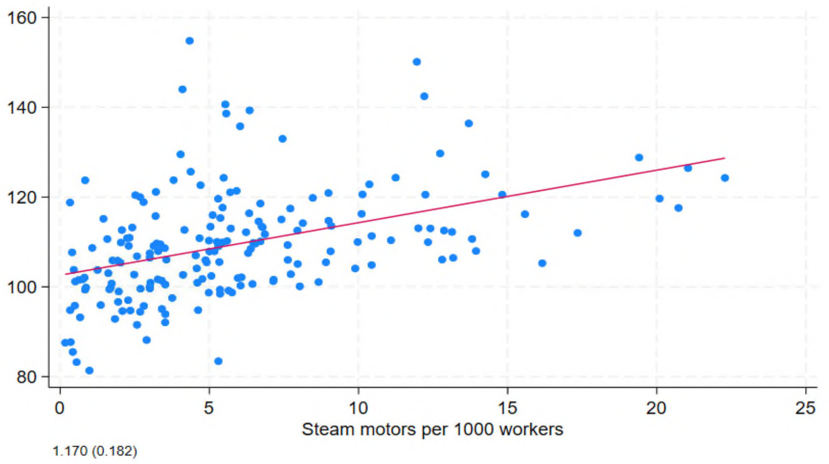
Figure 2. Steam engines in 1875 and mean wages in West Germany: 1975, 1995 and 2015



(a) 1975



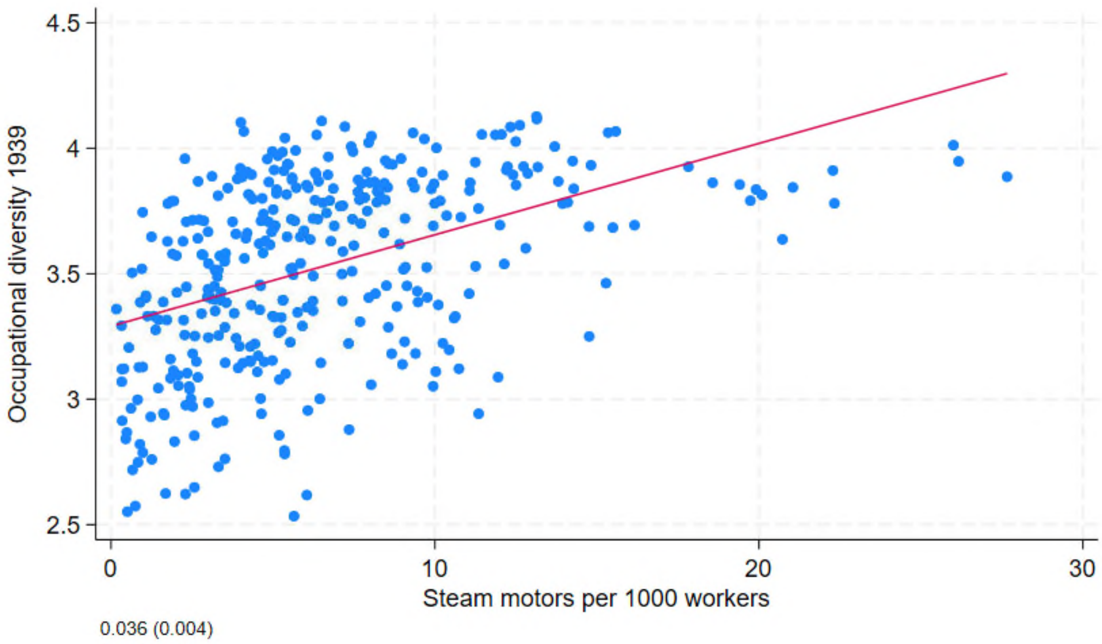
(b) 1995



(c) 2015

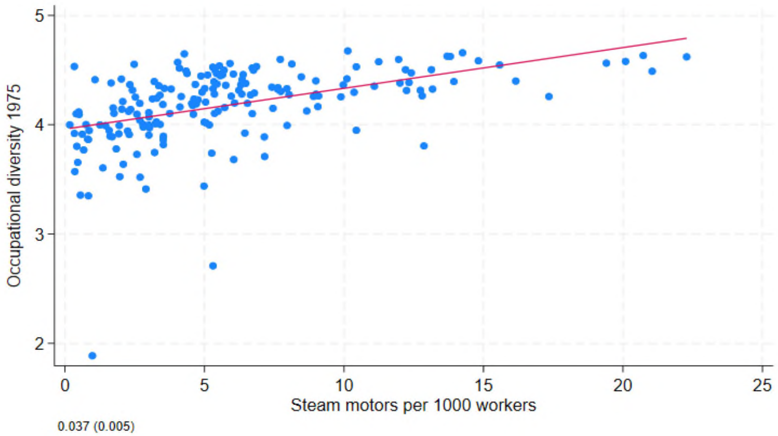
Notes: This figure shows the relation between steam engines per 1000 workers in 1875 Prussia and the mean wages (in 2015 euros) in West Germany in 1975, 1995, and 2015. The regression coefficient (SE) for the fitted line is presented in each panel.

Figure 3. Steam engines 1875 and occupational diversity 1939

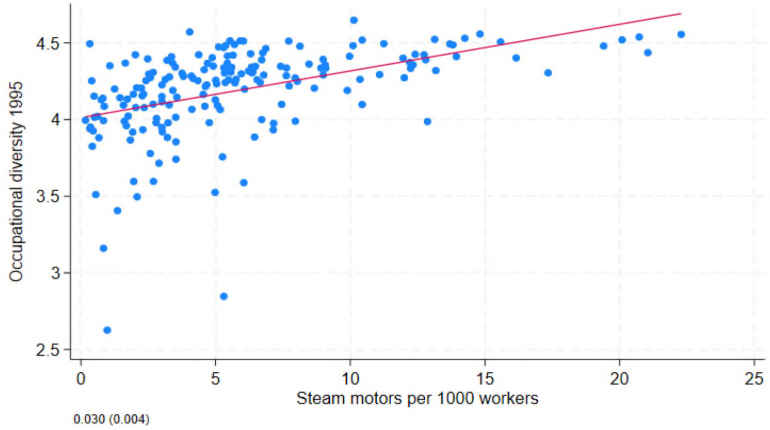


Notes: For each Prussian (1871) county, we relate steam engine concentration in 1875 to occupational diversity in 1939. Occupational diversity is measured by the Shannon index based on employment shares at 3-digit occupation level. The regression coefficient (SE) for the fitted line is presented in the panel.

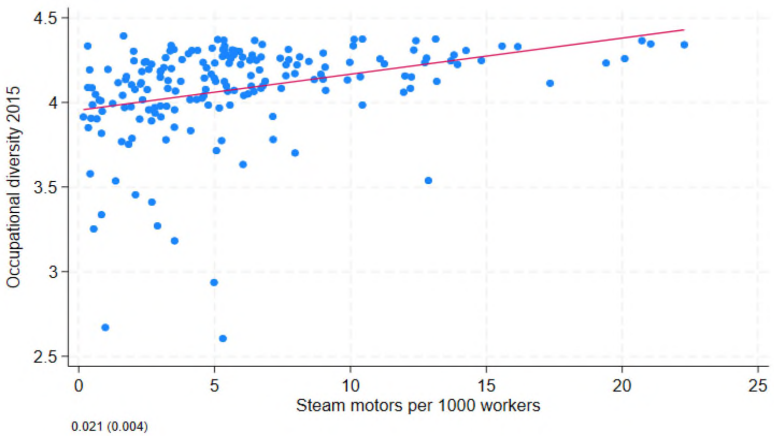
Figure 4. Steam engines in 1875 and occupational diversity: 1975, 1995 and 2015



(a) 1975



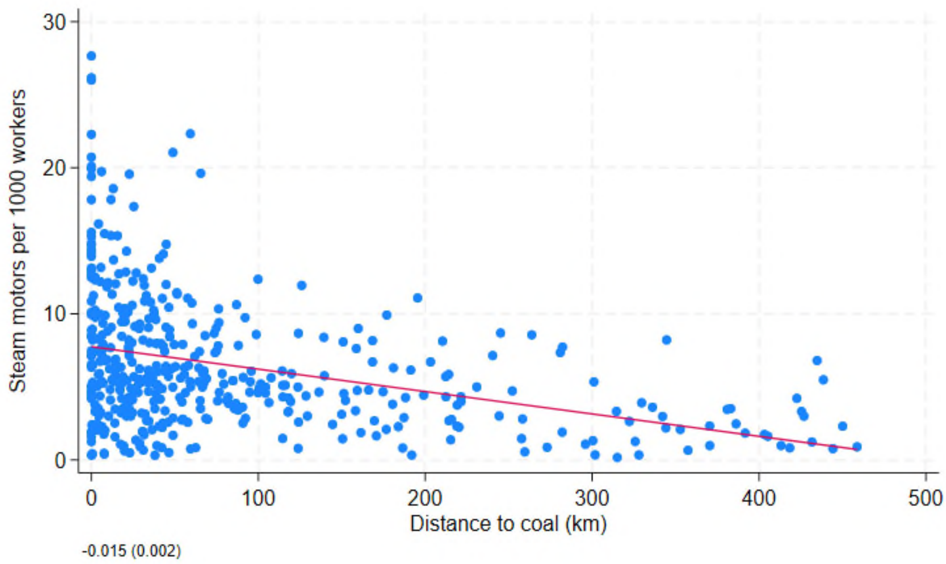
(b) 1995



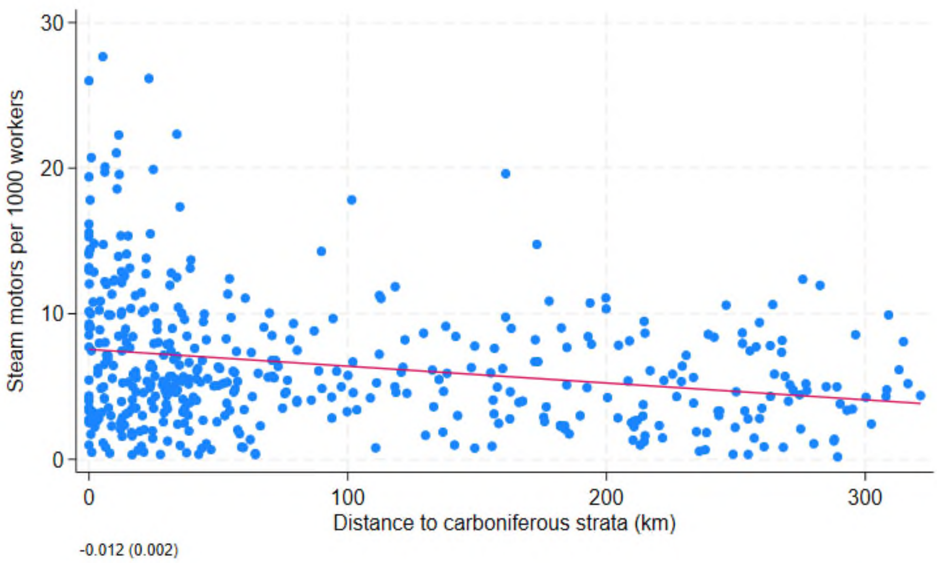
(c) 2015

Notes: This figure shows the relation between steam engines per 1000 workers in 1875 Prussia and occupational diversity in West Germany in 1975, 1995, and 2015. Occupational diversity is measured by the Shannon index based on employment shares at 3-digit occupation level. The regression coefficient (SE) for the fitted line is presented in each panel.

Figure 5. Distances to Caboniferous rock strata and steam engines



(a) Distance to coal



(b) Distance to Carboniferous rock strata

Notes: Data on the locations of coal and Carboniferous rock strata are from Fernihough and O'Rourke (2021). Observations are at the Prussian (1871) county level. In panel (a), we compute the nearest distance between the centroid of a county and the polygons corresponding to coal deposits. The coefficient (SE) of the regression line is -0.015 (0.002). In panel (b), we compute the nearest distance between the centroid of a county and the polygons corresponding to Carboniferous rock strata. The coefficient (SE) of the regression line is -0.012 (0.002).

Table 1. Summary statistics

	Mean	SD	Obs
A. County level economic and geographical characteristics			
Establishment survey 1875			
Steam engines per 1000 workers 1875	6.48	4.60	452
Horsepower per 1000 workers 1875	111.59	150.78	452
Share employed at large (6 or more employees) firms 1875	0.26	0.18	452
Population census 1871			
Log population 1871	10.80	0.41	452
Literacy 1871	0.88	0.13	452
Household size	4.79	0.34	452
Population growth rate 1867-1871	0.02	0.05	452
Share Protestants	0.64	0.38	452
Share Jews	0.01	0.01	452
Share aged below 10	0.25	0.02	452
Share female	0.51	0.02	452
Share born in municipality	0.59	0.12	452
Share Prussian origin	0.99	0.02	452
Share of residents in urban areas	0.28	0.22	452
Geographic characteristics			
Terrain slope	6.32	6.45	452
Wind speed at 10 meters	3.75	0.51	452
Distance to navigable river (km)	29.18	27.20	452
Distance to coal (km)	82.62	106.61	452
Distance to Carboniferous rock strata (km)	92.69	94.76	452
Occupation census 1882			
Share of workers in mining 1882	0.03	0.08	452
Share of workers in manufacturing sectors 1882	0.28	0.13	452
Share of workers in service sectors 1882	0.06	0.04	452
B. Individual level labor market outcomes 1975-2019			
Age	39.71	11.37	7778346
Female	0.32	0.47	7778346
Daily wage (in 2015 euros)	106.65	52.19	7778346
High skilled	0.19	0.39	7778346
Manufacturing	0.32	0.47	7778346
AKM firm FE for 1985-2017	0.13	0.28	5494277
C. County level outcomes between 1875 and 1975			
Occupation census 1939			
Share agriculture 1939	0.41	0.20	360
Shannon diversity index: occupations 1939	3.53	0.37	360
Shannon diversity index: occupations outside agriculture 1939	3.83	0.13	360
Non-agricultural establishment census 1939			
Share manufacturing in non-agricultural establishments	0.59	0.10	360
Shannon diversity index: industry	2.75	0.18	360
Shannon diversity index: industry within manufacturing	2.11	0.23	360

Table 1. Summary statistics (continued)

	Mean	SD	Obs
Patents 1877-1918			
Any high-value patents 1877-1918	0.74	0.44	452
High-value patents 1877-1918	36.97	255.48	452
Shannon diversity index: patents 1877-1918	1.21	1.01	336
D. County level outcomes 1975-2019			
Occupations and industries 1975-2019			
Share manufacturing	0.34	0.13	8300
Shannon diversity index: occupations	4.15	0.32	8300
Shannon diversity index: occupations in manufacturing	3.58	0.56	8300
Shannon diversity index: industries	4.35	0.53	8300
Shannon diversity index: industries in manufacturing	3.02	0.65	8300
Patents 1980-2014			
Any patents 1980-2014	0.98	0.14	6510
Patents 1980-2014	102.66	164.71	6510
Shannon diversity index: patents 1980-2014	2.86	0.00	6387

Notes: Data in panel A are Prussian (1871) county level characteristics. We digitize data on steam engines from the Preussische Statistik (1878). Panel B includes 1871 Prussian counties that overlap with 2019 West German municipalities. Individual-level data for 1975-2019 come from the BeH 2% sample from Institute of Employment Research (IAB). The AKM firm FE are provided by Bellmann et al. (2020) and are available for 1985-2017 only. The sample includes full-time regularly employed workers aged 18-65 whose workplaces are in West Germany. In panel C, occupation data for 1939 come from the 1939 Occupation Census while industry data for 1939 come from the 1939 Non-agricultural Establishment Census. Patent data for 1877-1918 come from Streb et al. (2006). In Panel D, occupation data are compiled from the same sources as in Panel B. Patent data for 1980-2014 come from de Rassenfosse et al. (2019).

Table 2. Steam engine prevalence in 1875 and wages in 1975-2019

	Dependent var: Log wage						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Steam engines (standardized)	0.026 **	0.023 *	0.024 ***	0.023 ***	0.037 ***	0.034 ***	0.035 ***
	(0.011)	(0.012)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Share of mining employment (standardized)					-0.040 ***	-0.042 ***	-0.043 ***
					(0.008)	(0.007)	(0.007)
Share of manufacturing employment (standardized)						0.027 ***	0.027 ***
						(0.008)	(0.008)
Share of services employment (standardized)							-0.004
							(0.017)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log population 1871		Yes	Yes	Yes	Yes	Yes	Yes
Literacy rate 1871		Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls			Yes	Yes	Yes	Yes	Yes
Additional controls 1871				Yes	Yes	Yes	Yes
Obs	7778346	7778346	7778346	7778346	7778346	7778346	7778346
Adjusted R-squared	0.23	0.23	0.24	0.25	0.25	0.25	0.25

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. Shares of mining, manufacturing and services employment are from 1882 occupation census. Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table 3. Educational attainment by birth cohorts

	Dependent var: High skilled					
Birth cohorts	All	<1940	1940-1949	1950-1959	1960-1969	1970+
	(1)	(2)	(3)	(4)	(5)	(6)
Steam motors (standardized)	0.020 ***	0.007 ***	0.015 ***	0.017 ***	0.025 ***	0.024 ***
	(0.005)	(0.002)	(0.004)	(0.005)	(0.006)	(0.008)
Number of observations	7778346	1247134	1200838	1849333	1975607	1505434
Adjusted R-squared	0.10	0.02	0.03	0.05	0.08	0.12
Mean of DV	0.187	0.067	0.113	0.170	0.227	0.314

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. High skilled includes invididuals with a high school diploma (Abitur) followed by apprenticeship training, or college graduates (university of applied sciences or university). Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include the full set of controls as in column (7) of Table 2, namely, year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table 4. Steam engines in 1875 and firm productivity 1985-2017

	Dependent var: Firm productivity					
Windows for AKM estimation	1985-2017	1985-1992	1993-1999	1998-2004	2003-2010	2010-2017
	(1)	(2)	(3)	(4)	(5)	(6)
Steam motors (standardized)	0.015 ***	0.014 ***	0.012 ***	0.015 ***	0.017 ***	0.015 ***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
Number of observations	5494277	1477473	1201094	1153297	1244096	1223487
Adjusted R-squared	0.52	0.10	0.08	0.07	0.07	0.05

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Firm productivity indicates the AKM firm FE as estimated by Bellmann et al. (2020) and are availble for 1985-2017. They use partially overlapping estimation windows 1985-1992; 1993-1999; 1998-2004; 2003-2010; and 2010-2017. For columns (2)-(6), the sample corresponds to the windows used in the Bellmann et al. (2020) estimation. For column (1), we pool all years between 1985 and 2017, where for years 1993-1997, we use estimates from the window 1993-1999; for years 1998-2002, we use estimates from the window 1998-2004; for years 2003-2009, we use estimates from the window 2003-2010, to deal with the overlapping windows for the AKM estimation. The number of observations from (2)-(6) do not add up to that in column (1). Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include the full set of controls as in column (7) of Table 2, namely, year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table 5. Role of skills and firm productivity in explaining wages 1985-2017

	Dependent var: Log wage			
	(1)	(2)	(3)	(4)
Steam motors (standardized)	0.036 ***	0.028 ***	0.018 ***	0.012 ***
	(0.008)	(0.007)	(0.004)	(0.003)
High skilled		0.358 ***		0.310 ***
		(0.005)		(0.003)
AKM firm FE			1.161 ***	1.103 ***
			(0.018)	(0.014)
Number of observations	5494277	5494277	5494277	5494277
Adjusted R-squared	0.22	0.31	0.45	0.51

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. High skilled includes invididuals with a high school diploma (Abitur) followed by apprenticeship training, or college graduates (university of applied sciences or university). AKM firm FE come from Bellmann et al. (2020) and are availble for 1985-2017. They use partially overlapping estimation windows 1985-1992; 1993-1999; 1998-2004; 2003-2010; and 2010-2017. We pool all years between 1985 and 2017, where for years 1993-1997, we use estimates from the window 1993-1999; for years 1998-2002, we use estimates from the window 1998-2004; for years 2003-2009, we use estimates from the window 2003-2010, to deal with the overlapping windows for the AKM estimation. All regressions include the full set of controls as in column (7) of Table 2, namely, year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table 6. Steam engines and occupational and industrial diversity 1939

	Dependent variable:					
	Occupational diversity (3-digit)			Industrial diversity (2-digit)		
	Share agriculture	All	Non-agriculture	Share manufacturing	All	Manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Occupation			**B. Industry (Non-agricultural establishments)**		
Steam motors (standardized)	-0.017 **	0.105 ***	0.043 ***	0.015 ***	0.029 **	0.053 ***
	(0.007)	(0.017)	(0.008)	(0.005)	(0.013)	(0.016)
Mean of DV	0.406	3.528	3.829	0.592	2.757	2.112
Adjusted R-squared	0.77	061	0.36	0.61	0.32	0.29
Number of observations	360	360	360	360	360	360

Notes: This table relates steam motors in 1875 to occupational and industrial diversity in 1939. Panel A uses occupation data (in three digits) from the 1939 occupation census. Panel B uses industry data (in two digits) from the 1939 non-agricultural establishment survey. Column (1) shows the share of agriculture in the overall workforce. Column (4) shows the share of manufacturing employment in non-agricultural establishments. In columns (2)-(3) and (5)-(6), diversity is measured by Shannon index. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors are reported. *** p<.01, ** p<.05, * p<.1

Table 7. Steam engines and occupational and industrial diversity 1975-2019

	Dependent variable:				
		Occupational diversity (3-digit)		Industrial diversity (4-digit)	
	Share manufacturing	All	Manufacturing	All	Manufacturing
	(1)	(2)	(3)	(5)	(6)
Steam motors (standardized)	0.029 ***	0.075 ***	0.172 ***	0.091 **	0.111 **
	(0.010)	(0.024)	(0.046)	(0.040)	(0.056)
Mean of DV	0.342	4.154	3.583	4.348	3.023
Adjusted R-squared	0.46	0.41	0.48	0.51	0.39
Number of observations	8300	8300	8300	8300	8300

Notes: This table relates steam motors in 1875 to occupational and industrial diversity in 1975-2019. The sample includes individuals working in areas of West Germany that overlap with the 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Column (1) shows the share of manufacturing. In columns (2)-(3), occupational diversity for each year is measured by the Shannon index at the 3-digit occupation level. In columns (4) and (5), industry diversity for each year is measured by the Shannon index at the 4-digit industry level. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by Prussian (1871) counties are reported. *** p<.01, ** p<.05, * p<.1

Table 8. Steam engines 1875 and high-value patents 1877-1918

	Dependent variable: patents		
	All	1877-1899	1900-1918
	(1)	(2)	(3)
		Panel A. Any patents	
Steam motors (standardized)	0.072 ***	0.084 ***	0.062 **
	(0.026)	(0.028)	(0.028)
Mean of DV	0.743	0.520	0.668
Obs	452	452	452
		Panel B. Log patents	
Steam motors (standardized)	0.135 *	0.147 *	0.080
	(0.079)	(0.083)	(0.083)
Obs	336	235	302
		Panel C. Patent diversity	
Steam motors (standardized)	0.077 *	0.079 *	0.065
	(0.04)	(0.048)	(0.053)
Mean of DV	1.211	0.857	1.124
Obs	336	235	302

Notes: The sample includes 452 Prussian counties matched to patent data from Streb et al. (2006) covering high-value patents from 1877-1918. Panel A focuses on the extensive margin of patents using a dummy indicating non-zero patents as dependent variable. Panel B examines the intensive margin, using log patents (conditional on non-zero patents) as the dependent variable. Panel C analyzes patent diversity, where diversity is measured by the Shannon index based on patent shares in 2-digit technology classes. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors are reported. *** p<.01, ** p<.05, * p<.1

Table 9. Steam engines 1875 and innovation 1980-2014

	All	1980-1994	1995-2004	2005-2014
	(1)	(2)	(3)	(4)
		Panel A. Any patents		
Steam motors (standardized)	0.003	0.007	0.001	0.001
	(0.008)	(0.012)	(0.013)	(0.002)
Obs	6510	2790	1860	1860
Mean of DV	0.981	0.971	0.980	0.998
		Panel B. Log patents		
Steam motors (standardized)	0.292 ***	0.259 **	0.371 ***	0.262**
	(0.109)	(0.114)	(0.123)	(0.115)
Obs	6387	2708	1823	1856
		Panel C. Patent diversity		
Steam motors (standardized)	0.206 ***	0.198 **	0.246 ***	0.177 ***
	(0.076)	(0.081)	(0.086)	(0.075)
Obs	6387	2708	1823	1856
Mean of DV	2.861	2.526	2.946	3.266

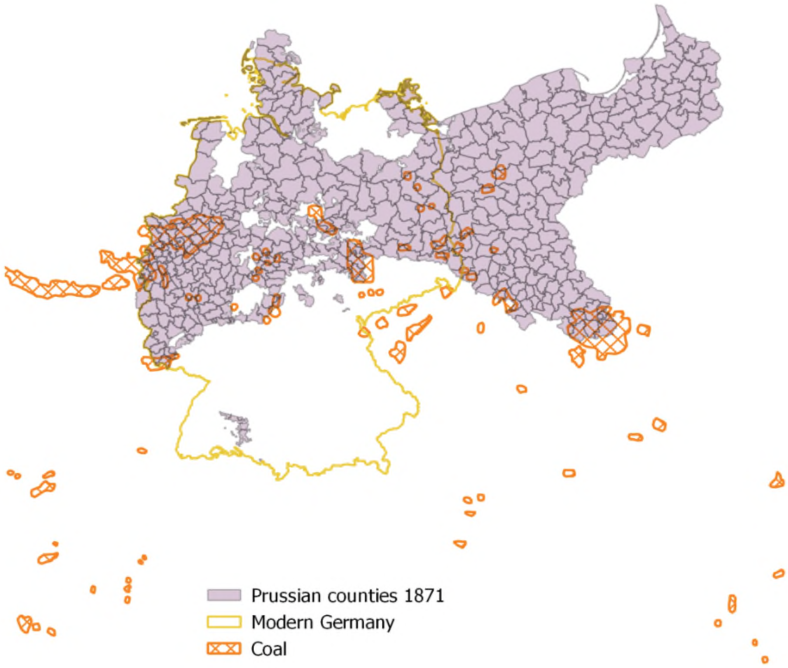
Notes: The sample includes 186 Prussian counties at the intersection between 1871 Prussia and West Germany. Geo-coded patent data covering 1980-2014 (West) Germany are provided by de Rassenfosse et al. (2019). Panel A focuses on the extensive margin of patents using a dummy indicating non-zero patents as dependent variable. Panel B examines the intensive margin, using log patents (conditional on non-zero patents) as the dependent variable. Panel C uses the diversity across patent classes, where diversity is measured by the Shannon index based on patent shares in 4-digit technology classes. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian countis are reported. *** p<.01, ** p<.05, * p<.1

Table 10. Effects of steam engine prevalence 1875 on wages 1975-2019: IV appraoch

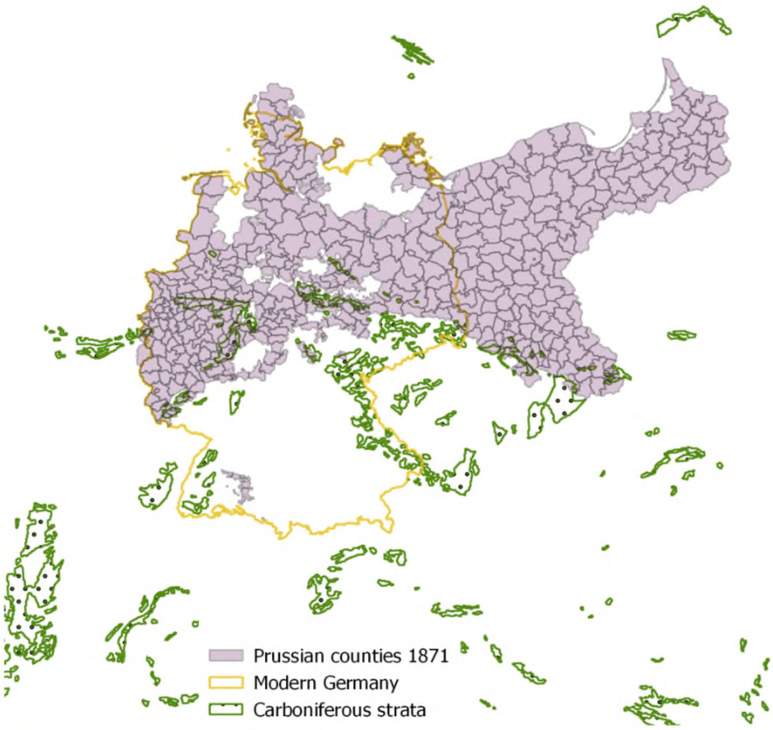
	Dependent variable						
	Log wage	Steam	Log wage	Log wage	Steam	Log wag	Log wage
	OLS	1st stage	RF	IV	1st stage	RF	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Steam motors: standardized	0.024 ***			0.043 ***			0.045 **
	(0.007)			(0.016)			(0.023)
Distance to Carboniferous strata: standardized		-0.442 ***	-0.019 **				
		(0.116)	(0.008)				
Distance to coal: standardized					-0.294 ***	-0.013 **	
					(0.088)	(0.006)	
F-stat		14.45			11.14		
[p-value]		[0.0002]			[0.0010]		
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log population 1871	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Literacy rate 1871	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	7778346	7778346	7778346	7778346	7778346	7778346	7778346
Adjusted R-squared	0.24	0.28	0.24	0.24	0.22	0.24	0.24

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Steam engines are per 1000 workers and measured in 1875. Column (1) shows the OLS estimates, using the specification in column (3) of Table 2. Columns (2)-(4) show the first stage, reduced-form, and IV estimates using distance to Carboniferous rock strata as an instrument for steam engines per 1000 workers. Columns (5)-(7) show the first stage, reduced-form, and IV estimates using distance to coal as an instrument for steam engines per 1000 workers. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. In columns (2) and (5), F-stat [p-value] refers to the respective instruments. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Figure A1. Locations of coal and Carboniferous rock strata



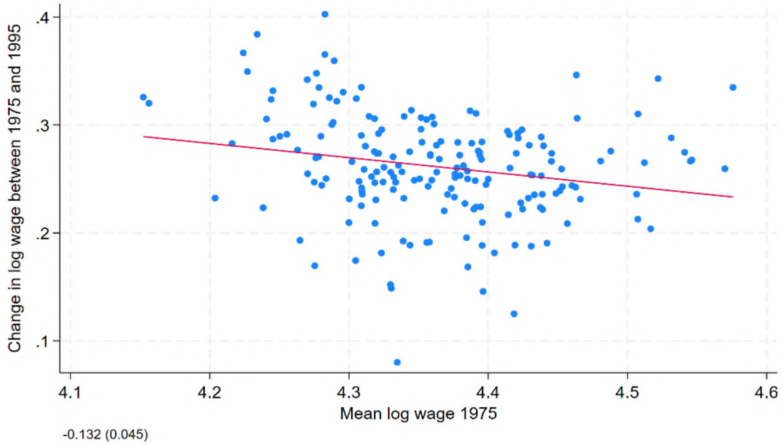
(a) Locations of coal



(b) Locations of Carboniferous rock stata

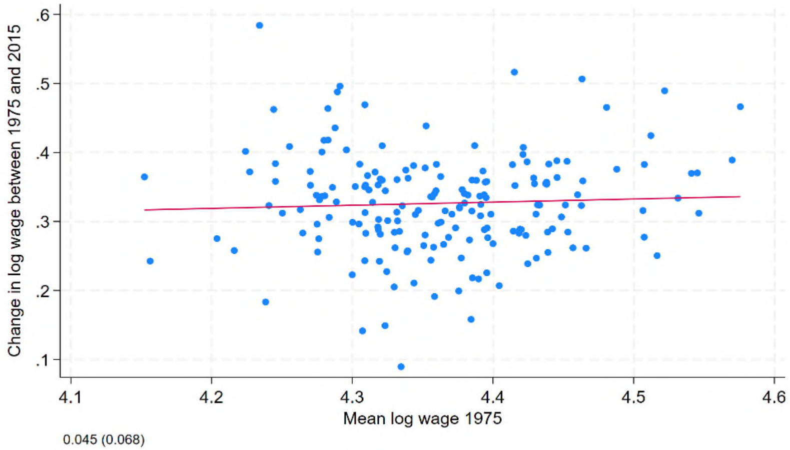
Notes: This figure overlays the locations of coal and Carboniferous rock strata on the map of Prussian counties (1871). Data on the locations of coal and Carboniferous rock strata are from Fernihough and O'Rourke (2021).

Figure A2. Regional convergence 1975-1995 and divergence 1995-2015



(a) Wage growth 1975-1995 vs. wages 1975

(b) Wage growth 1995-2015 vs. wages 1995



(c) Wage growth 1975-2015 vs. wages 1975

Notes: This figure shows the relation between initial mean wages and wage growth over different periods: 1975-1995 (panel (a)), 1995-2015 (panel (b)), and 1975-2015 (panel (c)). The regression coefficient (SE) for the fitted line is presented in each panel.

Table A1. Steam engine prevalence in 1875 and wages in 1975-2019: By time priods

	Dependent var: Log wage			
Sample	All	1975-1994	1995-2010	2011-2019
	(1)	(2)	(3)	(4)
Steam motors (standardized)	0.035 ***	0.029 ***	0.032 ***	0.033 ***
	(0.007)	(0.006)	(0.008)	(0.010)
Obs	7778346	3517520	2657841	1258985
Adjusted R-squared	0.25	0.32	0.19	0.15

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Column (1) includes the full sample. Columns (2)-(4) restrict sample to years 1975-1994; 1995-2010; and 2011-2019, respectively. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882 (same as in column (7) of Table 2). Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table A2. Alternative measures of steam engine prevalence in 1875 and wages in 1975-2019

	Dependent var: Log wage			
	(1)	(2)	(3)	(4)
Steam motors (standardized)	0.035 ***			0.032 ***
	(0.007)			(0.010)
Horsepower per worker (standardized)		0.034 ***		0.007
		(0.011)		(0.012)
Share of workers at large firms (standardized)			0.014 *	0.000
			(0.008)	(0.009)
Obs	7778346	7778346	7778346	7778346
Adjusted R-squared	0.25	0.25	0.25	0.25

Notes: The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Steam engines are per 1000 workers, horsepower per worker and share of workers in large (6 or more employees) firms are all measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882 (same as in column (7) of Table 2). Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors clustered by 1871 Prussian counties are reported. *** p<.01, ** p<.05, * p<.1

Table A3. Conley (1999) standard errors

	Dependent var: Log wage					
		Distance (km) for Conley SE				
	Baseline	60	90	120	150	180
	(1)	(2)	(3)	(4)	(5)	(6)
Steam motors (standardized)	0.035	0.035	0.035	0.035	0.035	0.035
	(0.0074)	(0.0069)	(0.0063)	(0.0074)	(0.0080)	(0.0088)
Number of observations	7778346	7778346	7778346	7778346	7778346	7778346

Notes: Column (1) reports our baseline estimates where standard errors are clustered by 1871 county levels. Columns (2)-(6) report Conley (1999) standard errors to allow for spatial correlations. Each column uses different distance cutoffs (from 60 to 180 km) where beyond the distance cutoff the correlation between the error terms of two observations is assumed to be zero. The sample includes individuals working in areas of West Germany that overlap with year 1871 Prussian counties. Individual labor market data come from the BeH 2% sample covering 1975-2019. We restrict attention to full-time regularly employed workers aged between 18 and 65. Firm productivity indicates the AKM firm FE as estimated by Bellmann et al. (2020) and are availble for 1985-2017. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include year FE, demographic controls, log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882 (same as in column (7) of Table 2). Demographic controls include gender and a cubic in age. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share Protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. *** p<.01, ** p<.05, * p<.1

Table A4. Steam engines and occupational and industrial diversity 1939: Alternative measures of diversity

	Dependent variable:					
	Occupational diversity (3-digit)			Industrial diversity (2-digit)		
	Share agriculture	All	Non-agriculture	Share manufacturing	All	Manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Occupation			**B. Industry (Non-agricultural establishments)**		
Steam motors (standardized)	-0.017 **	0.009 ***	0.003 ***	0.015 ***	0.029 **	0.053 ***
	(0.007)	(0.002)	(0.001)	(0.005)	(0.013)	(0.016)
Mean of DV	0.406	0.932	0.965	0.592	0.907	0.827
Adjusted R-squared	0.77	0.53	0.30	0.61	0.33	0.28
Number of observations	360	360	360	360	360	360

Notes: This table relates steam motors in 1875 to occupational and industrial diversity in 1939. Panel A uses occupation data (in three digits) from the 1939 occupation census. Panel B uses industry data (in two digits) from the 1939 non-agricultural establishment survey. Column (1) shows the share of agriculture in the overall workforce. Column (4) shows the share of manufacturing employment in non-agricultural establishments. In columns (2)-(3) and (5)-(6), diversity is measured by (1-HHI), where HHI is Hirfindahl-Hirschman Inex. Steam engines are per 1000 workers measured in 1875 and standardized to have a mean zero and SD of 1. All regressions include log population 1871, literacy rate 1871, geographic controls, additional controls 1871, and industry structure 1882. Geographic controls include a quadratic in terrain slope, mean wind speed at a height of 10 meters, and distance to a navigable river. Additional controls 1871 include historical county characteristics from the 1871 population census of Prussia: population growth rate 1867-1871, share living in urban areas, household size, share born in municipality, share Prussian origin, share protestants, share Jews, share females, and share aged below 10. Industry structure 1882 includes shares of mining, manufacturing and services employment from 1882 occupation census. Robust standard errors are reported. *** p<.01, ** p<.05, * p<.1