# Reversing Fortunes of German Regions, 1926–2019: Boon and Bane of Early Industrialization?*

Paul Berbée † Sebastian Till Braun ‡ Richard Franke §

This paper shows that 19th-century industrialization is an essential determinant of the pronounced changes in economic prosperity across German regions over the last 100 years. Using novel data on economic activity in 163 labor market regions in West Germany, we find that nearly half of them experienced a reversal of fortune, moving from the lower to the upper median of the income distribution or vice versa, between 1926 and 2019. Exploiting plausibly exogenous variation in access to coal, we show that early industrialization led to a massive decline in the per capita income rank after World War II, as it turned from an asset to economic development into a liability. We present evidence consistent with the view that early industrialization created a lopsided economic structure dominated by large firms, which reduced adaptive capacity and local innovation. The (time-varying) effect of industrialization explains most of the decline in regional inequality observed in Germany in the 1960s and 1970s and more than half of the current North-South gap in economic development.

Keywords: Industrialization, economic development, regional inequality

JEL classification: N91, N92, O14, R12

* **Acknowledgements:** Elisa Poletto, Santiago Rey-Sanchez Parodi, Sarah Stricker, and Felix Winkelmayer provided excellent research assistance. We thank the editor and three anonymous referees as well as seminar and conference participants at Bocconi University, the Growth Lab at Harvard Kennedy School, IZA, LSE, MLU, the University of Bayreuth, UC3M, WUR, ZEW, the EHS annual conference 2022, the Congress for Economic and Social History 2023, the Growth, History and Development Workshop at SDU, the RES annual conference 2024, and the CEPR Economic History Annual Symposium 2024 for their helpful comments and suggestions. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–project number 471335227 is gratefully acknowledged. All remaining errors are our own.

† ZEW–Leibniz Centre for European Economic Research, Mannheim, Germany

‡ Corresponding author. University of Bayreuth and IZA. Contact: University of Bayreuth, Faculty of Law, Business and Economics, 95440 Bayreuth, Germany. Email: sebastian.braun@uni-bayreuth.de. Tel.: +49 (0) 921 55 6256.

§ Trinity College Dublin and CEPH, Ireland

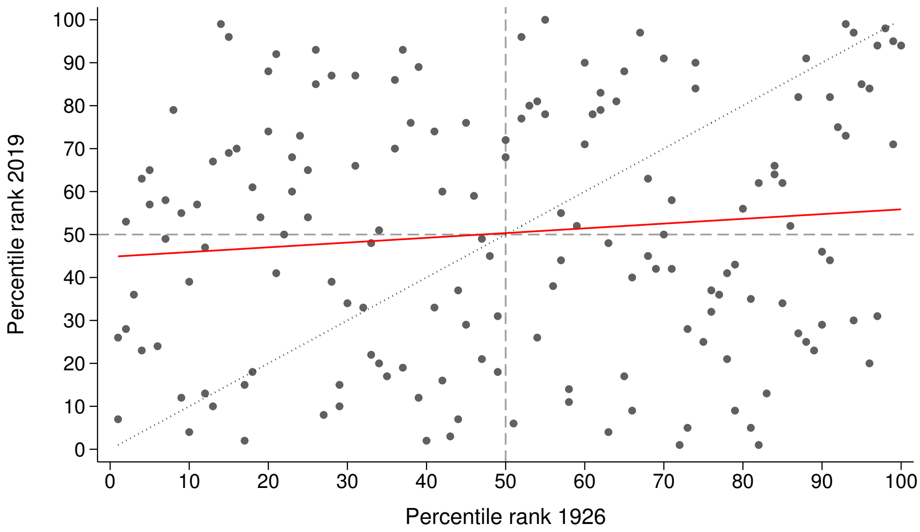
## 1 Introduction

There is growing evidence for many advanced economies that regional disparities have increased since about 1980 (e.g. Rosés & Wolf, 2019a; Gaubert, Kline, Vergara & Yagan, 2021). This "return of regional inequality" (Rosés & Wolf, 2019b) contributes to growing income inequality and could threaten social cohesion and political stability (Iammarino, Rodríguez-Pose & Storper, 2018; Floerkemeier, Spatafora & Venables, 2021). Of particular concern are declining regions, with their lack of economic opportunity, growing social problems, and rising political tensions (Austin, Glaeser & Summers, 2018; Rodríguez-Pose, 2018). What many of these declining regions have in common is that they had a high share of industrial jobs in the past and are now suffering from the dislocation of deindustrialization (Rosés & Wolf, 2021). Against this background, this paper examines the impact of 19th-century industrialization on changes in West Germany's economic geography over the past 100 years.

To this end, we first construct a novel dataset on regional economic activity for 163 West German labor markets at roughly 10-year intervals from 1926 to 2019. The dataset allows us to study the rise and decline of German regions over the last 100 years, as measured by their changing position in the GDP per capita distribution. Figure 1 plots the percentile rank in the distribution in 2019 against the rank in 1926 across regional labor markets. It shows that the correlation between income ranks in 1926 and 2019 is only weakly positive at 0.11 and not statistically significantly different from zero. Nearly half of the labor markets experienced a reversal of fortune between 1926 and 2019, moving from the upper to the lower median of the income distribution (lower-right quadrant) or vice versa (upper-left quadrant).

We then test the extent to which differences in early industrialization, measured as the industrial employment share in 1882, can explain these marked changes in West Germany's economic geography. Economic historians have long hypothesized that early industrialization in Germany contributed to economic development only until the mid-20th century, after which it led to regional decline (Abelshauser, 1984; Kiesewetter, 1986; Nonn, 2001). We test and quantify

Figure 1: Per capita income rank of West German labor markets in 1926 and 2019

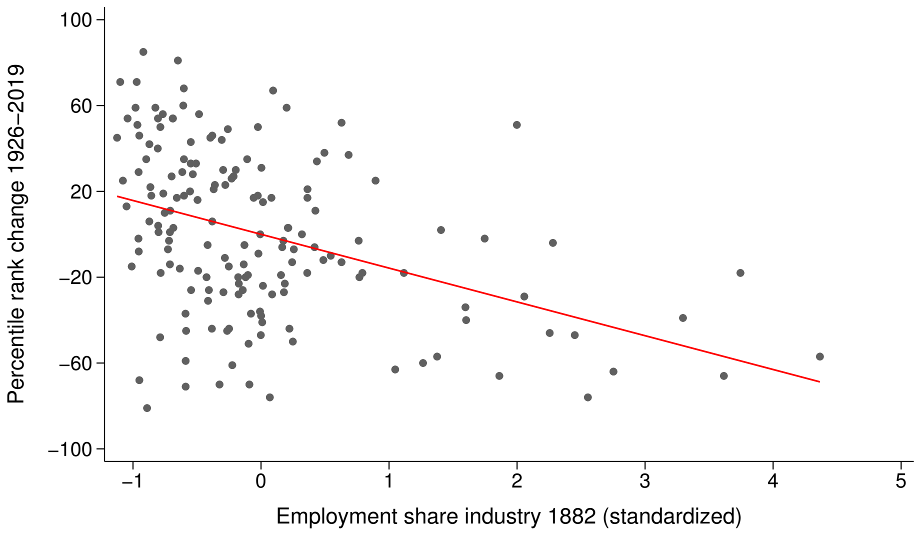



Notes: The figure plots the percentile rank in the 2019 income per capita distribution against the rank in 1926, along with the linear regression line in red. Each dot represents a labor market. Dashed horizontal and vertical lines indicate median percentile ranks. The dotted line indicates identical percentile ranks in 2019 and 1926.

this hypothesis. Consistent with the hypothesis, Figure 2 shows that early industrialization is strongly negatively associated with the change in the percentile rank of a labor market in the GDP per capita distribution over the period 1926-2019.

Of course, the observed relationship between industrialization and economic development is not necessarily causal. Unobserved institutional or geographic characteristics could jointly govern early industrialization and the development process. Moreover, the relationship may reflect reverse causality running from the development process to early industrialization (Franck & Galor, 2021). To establish causality, our empirical analysis instruments a region's industrial employment share in 1882 with its weighted least-cost distance to European coal fields, while controlling for the region's connectedness to other European markets. Our strategy exploits the fact that the heavy industries, characteristic of Germany's early industrialization process, were dependent on access to coal (Ziegler, 2012; Gutberlet, 2014). Coal, in turn, is often found in

Figure 2: Industrial employment in 1882 and per capita income rank change in 1926-2019




Notes: The figure plots the percentile change in the per capita income rank between 1926 and 2019 on the y-axis against the standardized industrial employment share in 1882, along with the linear regression line in red. Each dot represents a labor market.

carboniferous rock strata, formed hundreds of millions of years ago. Distance to coal fields is thus plausibly exogenous to economic development (Fernihough & O'Rourke, 2020).

Our 2SLS estimates indicate that a one standard deviation increase in the employment share of industry in 1882 resulted in a 30.0 point deterioration in the percentile rank in the income distribution between 1926 and 2019. Importantly, this decline is not only the result of the fading positive impact of early coal-based industrialization but also its detrimental impact in the long run. A one standard deviation increase in industrial employment in 1882 raised the rank of a labor market by 12.2 and 15.8 percentiles in 1926 and 1957, respectively, but lowered it by 17.8 percentiles in 2019. Thus, after World War II, early industrialization went from being an asset to economic development to a liability.

What explains the decline of early industrializing labor markets in the second half of the 20th century? Early industrializing regions in Germany were often dominated by large, capitalintensive firms, especially in heavy industries such as coal mining, iron and steel production, or shipbuilding (Abelshauser, 1984; Kiesewetter, 1986; Nonn, 2001). These large dominant firms limited regional adaptability when the old heavy industries went into crisis after World War II (e.g. Junkernheinrich, 1989; Hamm & Wienert, 1990; Grabher, 1993). These arguments are consistent with the view that the historical presence of large firms in heavy industry has crowded out entrepreneurial activity in U.S. cities (Chinitz, 1961). The lack of entrepreneurship in cities dominated by large firms then dampened urban growth (Glaeser, Kerr & Kerr, 2015).

We provide suggestive evidence that early coal-based industrialization did indeed create a lopsided economic structure dominated by large firms, which ultimately proved detrimental to development. This lopsided structure, which we detect already in the early 20th century, mediates the adverse influence of early industrialization on economic change after 1957. Early industrialization is also associated with a rigid political system and lower innovation today, consistent with the hypothesis that the dominance of large firms, embedded in a supportive system of corporate relationships, has reduced local innovation and thus adaptive capacity (Junkernheinrich, 1989; Grabher, 1993).

Finally, we examine whether regional differences in 19th-century industrialization underlie two trends that have received much public attention: the widening economic gap between North and South Germany, and the decline and rise of regional inequality. We quantify the contribution of early industrialization to the North-South gap by predicting the gap for a counterfactual scenario in which regions differ only in their share of industrial employment in 1882. We find that early industrialization accounts for more than half of the current North-South gap in per capita GDP. To quantify the contribution of early industrialization to regional inequality, we measure the inequality of a counterfactual GDP distribution in which all regions are assigned the mean 1882 industrial employment share. We find that the diminishing positive impact of early industrialization explains well over half of the decline in regional inequality from 1957 to

1980, but cannot explain the increase in regional inequality since 1980.[1]

**Contribution to the literature.** Our paper contributes to several literature strands. Firstly, we add to the nascent literature, initiated by Franck & Galor (2021),[2] on the long-run economic effects of early industrialization by providing the first empirical evidence for Germany, a country where heavy industry played a central role in the development process. Heavy industrialization created a lopsided economic structure dominated by large firms, which reduced adaptive capacity and local innovation, and ultimately depressed contemporary economic development.

Our results are consistent with those of Franck & Galor (2021), who document negative effects of early industrialization on regional prosperity in contemporary France. Franck & Galor (2021) identify negative human capital effects as the key mediating channel. For Germany, we find that early industrialization reduces vocational training and increases school dropout rates. This is likely because Germany's old industrial regions lack the extensive vocational and technical training institutions associated with the small and medium firms that dominate elsewhere in Germany (Herrigel, 2000).

Secondly, we contribute to the growing literature on the long-term economic impact of natural resource abundance, showing that coal-based industrialization in Germany turned from an asset to economic development into a liability over the last 100 years.[3] This finding resonates with Esposito & Abramson (2021) who show that former coal mining regions in Europe now have lower GDP per capita than regions where coal was not previously mined. Fritzsche & Wolf (2023) demonstrate that local coal abundance in Western Europe turned into a curse with the rise of

[1]We focus on the period from 1957 to 2019 in our discussion of regional inequality, as we use firm sales as a proxy for GDP before 1957 (see Section 2 for details). Therefore, we cannot calculate regional inequality measures that are comparable over our entire sample period.

[2]Hitherto, the potentially detrimental long-run effects of the industrialization process in post-industrial economies have "neither been raised nor been explored in the modern economic growth literature" (Franck & Galor, 2021, p. 109).

[3]A large literature, reviewed for example by Van der Ploeg (2011) and Venables (2016), examines the impact of natural resource abundance on development, focusing mainly on the impact of oil in the post-World War II period. A common theme in this literature is the "resource curse," which describes the underperformance of resource-rich economies.

cheap oil imports in the early 1960s.[4] Both papers point to negative consequences for tertiary education as the key mediating channel, with fewer universities being built in historic coal mining regions (Esposito & Abramson, 2021). In contrast, we find no evidence that coal-based industrialization reduced tertiary education in West Germany, likely because new universities were established in Germany's industrial heartland in the 1960s and 1970s. Therefore, government intervention can counteract the adverse effects on tertiary education found elsewhere in Europe.

Instead, early industrialization became a constraint on economic development in Germany because of its negative impact on local entrepreneurship, innovation, and adaptability caused by the dominant position of large firms in heavy industries. This result aligns with previous findings for the U.S. and Great Britain, which have shown that the historical presence of large firms impedes entrepreneurship (Chinitz, 1961; Glaeser & Kerr, 2009; Glaeser et al., 2015; Stuetzer, Obschonka, Audretsch, Wyrwich, Rentfrow, Coombes, Shaw-Taylor & Satchell, 2016), and connects these earlier findings to the growing literature examining the patterns and causes of the decline of manufacturing in advanced economies.[5] In particular, our study suggests that a more diversified economic structure could help avoid the negative long-term effects of industrialization, adding to recent evidence emphasizing the importance of local human capital in adapting to deindustrialization (Fritzsche & Wolf, 2023; Gagliardi, Moretti & Serafinelli, 2023).

Finally, we provide new descriptive insights into the evolution of regional economic activity in Germany at a much more granular level than previous studies (e.g., Kaelble & Hohls, 1989; Frank, 1993; Kiesewetter, 2004) and quantify the contribution of early industrialization to the

[4] Moreover, Matheis (2016) documents the negative long-term effects of coal production on the population of US counties. In Europe, proximity to coal boosts urban growth between 1750 and 1900 (Fernihough & O'Rourke, 2020), but becomes a liability after the late 1970s (Rosés & Wolf, 2021). Prior work for Germany provided case studies for single regions and sectors, notably mining in the Ruhr region (Abelshauser, 1984; Nonn, 2001).

[5] An older literature strand documents the poor industrial performance of the U.K. in the 1960s and 1970s and explores potential causes for the country's relative decline (e.g., Kitson & Michie, 1996; Broadberry & Crafts, 2003). In contrast to our study, this literature focuses on the national level. At the regional level, Rice & Venables (2021) show that the decline in manufacturing employment in the 1970s still predicts low employment today. Similarly, Austin et al. (2018) links U.S. deindustrialization to current non-employment rates. Recent evidence suggests that trade with China has contributed to U.S. manufacturing decline since the 1990s (Autor, Dorn & Hanson, 2016) but has stabilized manufacturing employment in Germany (Dauth, Findeisen & Suedekum, 2017). The decline of the Rust Belt, the heavy manufacturing region bordering the Great Lakes, has been linked to prolonged labor market conflict in the region's key industries (Alder, Lagakos & Ohanian, 2023).

evolution of regional income inequality. In this respect, our work is most closely related to a recent study by Wolf (2019), which describes the evolution of regional GDP for German NUTS-2 regions over the period 1900-2010. We complement Wolf (2019) by presenting evidence for 163 labor markets within West Germany instead of 29 regions.[6] In addition, we focus on changes in the position in the income distribution (rather than on regional convergence as Wolf, 2019, does).

## 2 Data

This section describes our data. More details on the sources and the definition of all variables can be found in Section A.1 in the appendix.

**Unit of analysis.** Our unit of analysis is the 163 West German labor markets defined in Institut für Weltwirtschaft (IfW) (1974) based on commuting flows.[7] We aggregate our source data, collected at the level of counties (*Kreise*), to the level of labor markets using Geographical Information System (GIS) software.[8] The fact that we focus on local labor markets rather than counties has two advantages. First, where people live and work often differs at the county level, which poses problems in ranking counties based on their per capita income. In contrast, most people live and work within the same local labor market. Second, territorial reforms led to a sharp decrease in the number of West German counties, especially in the 1970s, making the conversion of data in historical to current county boundaries prone to error. A common reform was to merge urban counties (*Stadtkreise*) with their surrounding rural counties (*Landkreise*). Such reforms do not pose problems at the level of local labor markets, as the latter encompass interconnected rural and urban counties.

[6] Local labor markets are economic units defined by commuting patterns, so residents typically live and work in the same labor market. Nevertheless, the smaller spatial unit in our study may exacerbate potential regional spillovers, complicating our empirical analysis of the long-run effects of coal-based industrialization. However, we show in Section 3 that our main results hold at a higher level of aggregation, similar to that of Wolf (2019).

[7] To the best of our knowledge, the definition in IfW (1974) is the earliest available for West Germany. We exclude the Saarland from our sample, as it did not become part of postwar West Germany until 1957.

[8] The definition of local labor markets is based on county boundaries in 1966. For other years, we overlay maps of historical county boundaries with the base map of local labor markets. We then use the proportion of each historical county's area that belongs to a particular local labor market to aggregate the county-level data.

**GDP per capita, 1926-2019.** Our main outcome variable, ranging from 1 to 100, is the percentile rank of a labor market in the income distribution of West German labor markets. We have two main reasons for focusing on income ranks rather than levels. First, levels for 1926-1955 are not directly comparable to levels for 1957-2019. This is because we proxy regional GDP prior to 1957 with firm sales, as discussed below. More generally, the recurring revisions to the national accounts and the need to adjust for inflation make it difficult to compare income levels over time. Second, we consider percentile ranks a more intuitive measure for documenting reversals (or changes in position in the income per capita distribution more generally). Nevertheless, we also present estimates using income levels as the dependent variable, which shed additional light on the economic significance of our results.[9]

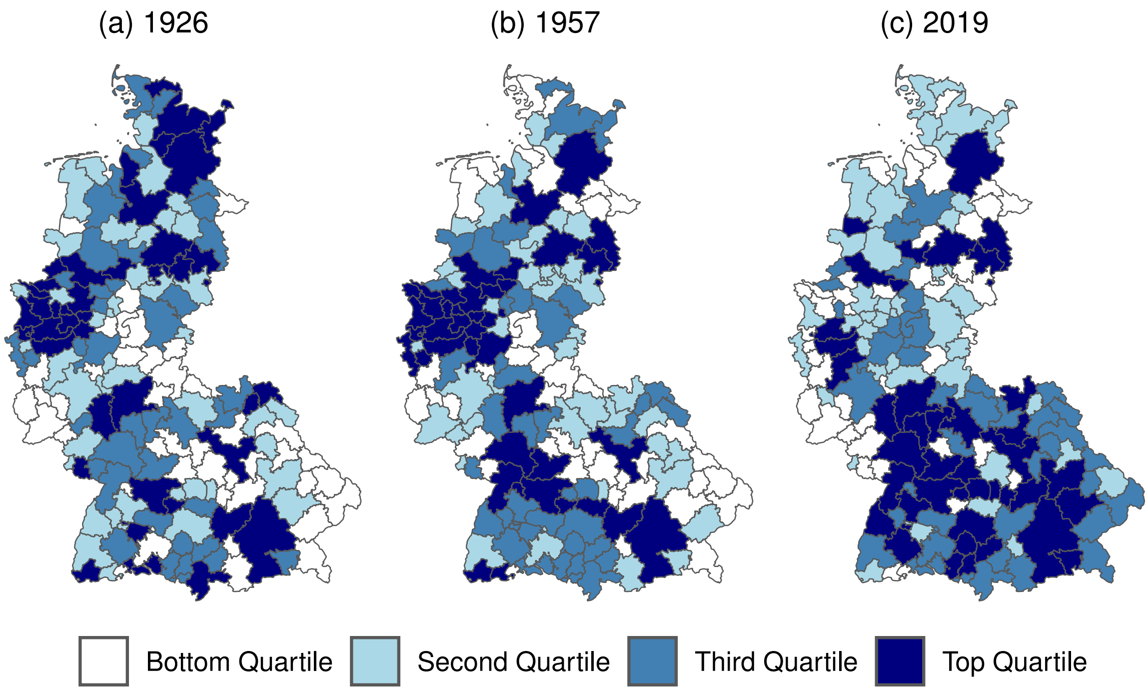
West Germany's federal statistical office began publishing disaggregated GDP per capita data at the county level in 1957. We digitized the data for 1957-1992 from printed sources. Data are available for eleven years in this period, namely for 1957, 1961, 1964, 1966, 1968, 1970, 1972, 1974, 1978, 1980, and 1992. GDP data for 1992-2019 are available at an annual frequency in electronic form (Arbeitskreis VGR der Länder, 2021). As in, e.g., Vonyó (2012) or Peters (2022), we proxy regional GDP before 1957 by firm sales, which we collected for 1926, 1935, 1950, and 1955. Although firm sales are not a direct measure of the production value, they correlate strongly with local GDP and deliver similar income rankings.[10]

In total, our dataset contains regional income statistics for 41 years between 1926 and 2019. Since we are interested in long-term changes rather than short-term fluctuations, our analysis focuses on data points at roughly 10-year intervals, namely 1926, 1935, 1950, 1957, 1961, 1970,

[9]Appendix Figure A-1 shows the distribution of (log) GDP per capita in 1957, 1980, and 2019. The distribution became less unequal between 1957 and 1980, making a labor market's position in the distribution less consequential for relative GDP per capita. In contrast, the distribution became more unequal again between 1980 and 2019. We return to the issue of regional inequality in Section 5.

[10]The correlation between a labor market's percentile rank in the 1955 sales per capita and the 1957 GDP per capita distribution is 0.870. The corresponding scatterplot in Appendix Figure A-2 confirms the close correspondence between GDP per capita and sales across labor markets. Additional comparisons between sales and GDP are possible for the nine West German states in 1950, when the correlation coefficient is 0.99. In data from before World War II, the correlation between per capita sales in 1935 and per capita national income in 1936 is 0.92 across 19 German regions (Braun & Kvasnicka, 2014).

Figure 3: Per capita income rank of West German labor markets, 1926-2019 (quartiles)




Notes: Each map shows the quartile rank in the per capita income distribution of West German labor markets in 1926 (Panel (a)), 1957 (Panel (b)), and 2019 (Panel (c)).

1980, 1992, 2000, 2010, and 2019 (where we include 1957 as the first year for which GDP per capita data are available). The years chosen are broadly comparable to those in Wolf (2019).[11]

The maps in Figure 3 illustrate how the economic weights within West Germany have shifted over time. For each West German labor market, the maps show its quartile rank in the income distribution in 1926, 1957, and 2019. In 1926, the economic powerhouses are scattered across the country. They are mainly concentrated in the metropolitan areas in the west (Rhineland, Ruhr region) and north (Bremen, Hamburg). But there are also clusters of rich labor markets in the south, e.g., around Munich or Stuttgart. Poorer labor markets are concentrated in the southeast

[11] Wolf (2019) reports GDP per capita data at the NUTS-2 level for 1900, 1910, 1925, 1938, 1950, 1960, 1970, 1980, 1990, 2000, and 2010. We do not interpolate to obtain data for years ending with 0 in case they are not readily available. For instance, we do not interpolate our data for 1957 and 1961 to obtain a value for 1960, but rather stick to 1961.

of West Germany. Despite the potentially asymmetric effects of World War II and the rise of the Iron Curtain, West Germany's economic geography changed little between 1926 and 1957.[12]

In 2019, the regional distribution of incomes has changed significantly. Labor markets in the Ruhr area in particular have slipped in the income ranking. The same applies to some of the historically rich regions in northern Germany, such as Bremerhaven or Itzehoe. By contrast, only a few of today's poorest labor markets are still to be found in the southeast. Instead, the poorest regions are now concentrated in the far west and northwest of West Germany. Several major centers of the automotive industry (Ingolstadt, Munich, Sindelfingen, Stuttgart, and Wolfsburg) top the ranking, accompanied by large cities such as Cologne, Düsseldorf, and Frankfurt.

**Industrial employment share 1882.** Our main explanatory variable of interest is the share of the local workforce working in industrial occupations in 1882. We thus measure industrialization after Germany's take-off phase, typically dated to 1840-1870s, but before the rise of new industries during Germany's *Hochindustrialisierung* (Ziegler, 2012; Tilly & Kopsidis, 2020). Our measure comes from the first German-wide occupation census that contains results at the county level (Kaiserliches Statistisches Amt, 1884) and is standardized to have a mean of zero and a standard deviation of one.

**Distance to European coal fields.** Our empirical analysis uses an instrumental variable strategy to identify the causal effect of early industrialization on development. We use the weighted least-cost distance to European coal fields as an instrument for the 1882 employment share in industrial occupations. Access to coal is widely acknowledged as a key factor behind the success of Germany's early industrializing regions, which relied mainly on heavy industries (Fremdling, 1977; Ziegler, 2012). Fernihough & O'Rourke (2020) have recently demonstrated the importance of coal for the European Industrial Revolution in general.

[12]The correlation between the position in the 1957 GDP per capita distribution and the 1926 sales per capita distribution is 0.712, despite the change in measurement. Correlating the ranks in the 1926 and 1955 sales per capita distributions even yields a coefficient of 0.797.

Previous studies have used the proximity to the nearest coal-bearing rock strata as an instrument for the historical use of steam engines (de Pleijt, Nuvolari & Weisdorf, 2020) and coal mining (Esposito & Abramson, 2021). In contrast to these papers, we use the sum of least-cost distances to all European coalfields, weighted by their area, to account for the fact that the closest coalfield is not necessarily the one that can be reached with the lowest transportation costs. This modification is important for the German context. In particular, regions in northern and northeastern Germany initially relied primarily on British coal, rather than coal from closer German mines, because of the low cost of river and sea transportation (Fremdling, 1979).[13] By accounting for the size of a coalfield, our measure also distinguishes between smaller and larger coal fields with potentially very different yields.

To calculate the instrumental variable, we first divide Europe in a one-by-one kilometer grid. Based on the local geography, we assign each cell a specific pre-industrial (1790) transportation cost, which we take from Daudin (2010).[14] We normalize the cost to one for cells that have access to the sea. Cells with access to a major river are assigned a cost value of 1.018, all other cells are assigned Daudin's value for road transport of 2.963.[15] We then calculate the least-cost distance from each labor market to all European coalfields, using Dijkstra's algorithm and the grid as cost surface. The algorithm finds the least-cost path from a region to a coalfield, adding cell-specific costs along the way. The instrument for a given labor market $i$, $C_i$, is the logarithm of the weighted sum of the least-cost distances to all coalfields:

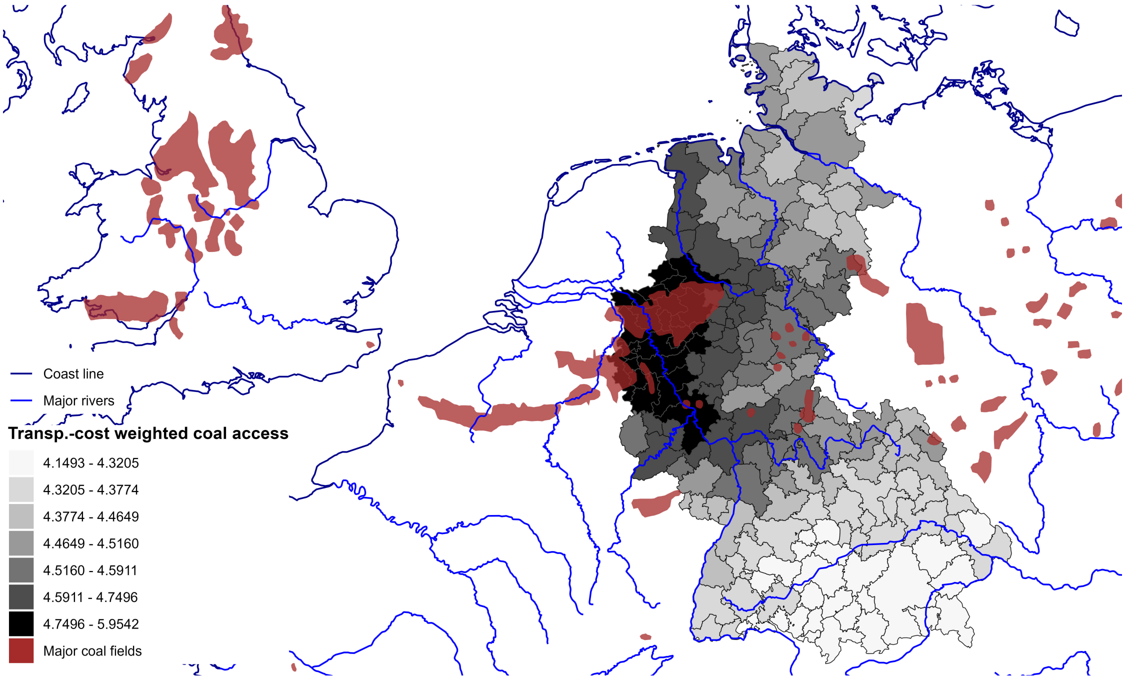
$$C_i = \log \left( \sum_{k=1}^{K} \frac{area_k}{cost_{ik}} \right), \tag{1}$$

[13] The relatively low price of English coal in northern Germany was also due to lower production costs in English mines, but this is not accounted for by our instrument.

[14] By relying on a pre-industrial cost vector that does not include transportation via paved roads, canals, or railroads, we avoid that our instrumental variable captures the fact that early industrialized regions are better connected to the transportation network for endogenous reasons. Since mode-specific transport cost vectors are not available for Germany, we adopt a cost vector for France, which has a similar topography.

[15] We take shape files of major European rivers from Fernihough & O'Rourke (2020). The value for rivers is the average of upstream and downstream river transport. We also probe the robustness of our results to alternative costs vectors. In particular, we use squared transport costs and assign higher costs of 2.476 to river and 9.75 to road transport, respectively, following Bairoch (1990). Moreover, we restrict the set of rivers to those that are at least 20 meters wide and two meters deep in an additional robustness check.

Figure 4: Weighted least-cost distance to European coalfields




Note: See the main text for details on the construction of the weighted least-cost distance. The transport cost vector is taken from Daudin (2010).

where $cost_{ik}$ is the least cumulative costs from labor market $i$ to coalfield $k$ and $area_k$ is the area of the coalfield polygon in square kilometers. We take the location and extent of European coalfields from Fernihough & O'Rourke (2020).

Figure 4 illustrates the regional variation in the instrument as well as the location of the most important coalfields, coast lines, and major rivers. Higher values indicate more favorable access to coal. Not surprisingly, access is most favorable in the Ruhr region and in regions connected to the Ruhr by rivers. Regions in northern Germany also have relatively favorable access to coal because they can obtain British coal via the North Sea.[16] To ensure that the instrument does not just pick up the connectedness of a labor market within Europe, our empirical analysis controls for a labor market's sum of least-cost distances to all European grid cells on land using the same

[16]Unfortunately, we lack 19th-century coal prices for regional labor markets, which are arguably a more direct measure of access to coal. However, using data on coal prices for 45 German cities in 1906/7, we find that our instrument predicts the cross-city variation in coal prices well (the correlation coefficient is -0.722).

cost vector as in equation (1).

**Economic structure in 1907.** To test whether early industrialization led to a lopsided economic structure, we digitized the 1907 Census of Manufactures (*Gewerberbliche Betriebsstatistik*) (Kaiserliches Statistisches Amt, 1909a,b). The data provide county-level information on the number of establishments and their total employment, disaggregated by 3-digit industries. We use the data to calculate the average firm size in industry and mining, the employment share in industries dominated by large firms, and sectoral employment concentration. We define industries dominated by large firms as those with a Germany-wide employment share of at least 50% in firms with at least 501 or 1000 employees. These industries include, for example, coal mining, arms foundries, and shipbuilding (see Appendix Table A-2 for a complete list). We measure concentration using the Herfindahl-Hirschman Index (HHI).[17]

# 3 Early industrialization and economic development, 1926-2019

This section tests the hypothesis that early industrialization was initially beneficial and later detrimental to economic development, leading to a relative deterioration in the position in the income distribution over the period 1926-2019.

**Empirical specification.** We estimate the effect of early industrialization on subsequent development using 2SLS. The second stage regression quantifies the effect of the standardized industrial employment share in 1882, $I_{i,1882}$, on a labor market's percentile rank in the income per capita distribution, $y_{i,t}$ (or changes thereof):

$$y_{i,t} = \alpha + \beta_t \hat{I}_{i,1882} + \mathbf{X}_i' \gamma_t + \epsilon_{i,t}, \tag{2}$$

where $\mathbf{X}_i$ is a set of control variables, which in our baseline regression includes land accessibility and the number of towns per area in 1700 (see next paragraph for details), and $\epsilon_{i,t}$ is an error

[17]The index is calculated as $HHI_i = \sum_{l=1}^{L} (b_{il})^2$, where $b_{il}$ is the employment share of labor market $i$ in (3-digit) industry $l$ in total industrial employment. The HHI ranges from $1/L$ (when all sectors have equal employment) to 1 (when all employment is concentrated in one sector). We use data for $L = 300$ sectors.

term.[18] We obtain the predicted values of $I_{i,1882}$, $\hat{I}_{i,1882}$, from a first stage regression using the (log) weighted least-cost distance to European coal fields, $C_i$, as an instrument for early industrialization (see Section 2 for details on the construction of the instrument):

$$I_{i,1882} = \delta + \zeta C_i + \mathbf{X}_i^{'}\eta_t + u_i, \tag{3}$$

where $\mathbf{X}_i$ contains the same control variables as in equation (2). Along with 2SLS estimates, we also report the results from conditional OLS regressions (relating $y_{i,t}$ to $I_{i,1882}$) and reduced-form regressions (relating $y_{i,t}$ to $C_i$).[19] To account for spatial correlation, we always report Conley (1999) standard errors and standard errors clustered at the level of 36 districts (*Regierungsbezirke*).

The key identifying assumption for the 2SLS regression to yield a consistent estimate of our coefficients of interest, $\beta_t$, is $Cov(C_i, \epsilon_{i,t}) = 0$. The assumption states that (i) there is no unobserved factor that drives economic development and is correlated with $C_i$ and that (ii) $C_i$ affects economic development only through its effect on early industrialization. An obvious challenge to the identifying assumption is that coal access might capture favorable location and thus better market access in general, as, e.g., regions along rivers tend to have better access to coal. To address this concern, we include the weighted least-cost distance to all European land cells as control variable. Intuitively, the control measures a region's geographic isolation within Europe.[20] We thus exploit only the residual variation in coal access, which is not driven by a region's favorable location in general. Furthermore, we control for the number of towns per square kilometer in 1700, reported in Cantoni, Mohr & Weigand (2020), to capture differences in pre-industrial development (and locational fundamentals that lead to these differences).

Another challenge to the identifying assumption is shocks during our sample period that

[18]Our regression model is similar in spirit to Franck & Galor (2021), who also regress indicators of regional development at different points in time on a measure of early industrialization (total horsepower of steam engines in 1860-65), exploiting the diffusion of steam engines in their IV strategy.

[19]In addition, we estimate an ordered probit model in which the explanatory variable of interest, the industrial employment share in 1882, is treated as endogenous. This allows us to account for the fact that $y_{i,t}$ is a limited dependent variable. The results of the ordered probit model, which are available from the authors upon request, confirm the pattern found in our linear model.

[20]Ashraf, Özak & Galor (2010) establish that, in contrast to conventional wisdom, prehistoric geographical isolation has positive long-run effects on cross-country differences in economic development.

exhibit spatial patterns, which the instrument picks up. Several shocks come to mind. First, the German division in 1945 and re-unification in 1990 led to an asymmetric loss and gain of market access for regions at the former inner-German border (Redding & Sturm, 2008). These regions, in turn, are predominately located in northern Germany, which shared a long border with the German Democratic Republic (GDR). Second, about eight million displaced Germans from Eastern Europe arrived in West Germany after World War II, initially gathering mainly in the eastern parts of the country. The inflow accelerated structural change away from low-productivity agriculture (Braun & Kvasnicka, 2014) and increased income per capita in the long run (Ciccone & Nimczik, 2022; Peters, 2022). Third, between 1949 and 1961, nearly 3 million people fled the GDR to West Germany, many highly skilled (Becker, Mergele & Woessmann, 2020). Fourth, the immigration of so-called "guest workers" in the late 1950s and 1960s, mainly from southern and south-eastern Europe, also followed a spatial pattern. Migrants came first to southern Germany's industrial centers and later to western and, finally, northern Germany. Fifth, the influx of refugees in 2015-16 was also concentrated in regions in western Germany, despite the existence of allocation quotas (Gehrsitz & Ungerer, 2022). Sixth, bombing damage during World War II was higher along the northern coastline due to its proximity to Great Britain (Vonyó, 2012) and reduced long-term private wealth in Germany (Halbmeier & Schröder, 2024). Finally, Germany's western integration after 1945 may have disproportionately benefited regions in western Germany (Redding & Sturm, 2008). We show in different checks that adding controls for these shocks leaves our results unchanged.

**Baseline results.** Table 1 reports OLS (Panel A), 2SLS (Panel B), and reduced form (Panel C) estimates of the effect of early industrialization on economic development in 1926-2019. The first two columns report the impact on changes in percentile rank in the per capita income distribution between 1926 and 2019 (Column (1)) and between 1957 and 2019 (Column (2)). Columns (3) through (5) report the (underlying) effect on the position in the income distribution in 1926, 1957, and 2019. These years include the starting and ending points of our sample period and the first year for which we have GDP per capita data.

Columns (1) and (2) show that early (coal-based) industrialization has an economically and statistically significant negative impact on the change in a labor market's position in the per capita income distribution. The OLS estimate in Column (1) implies that a one standard deviation increase in the industrial employment share in 1882 led to a 19.3 percentile deterioration in income position between 1926 and 2019. The 2SLS estimate implies an even larger decline of 30.0 percentiles. Notably, the effect sizes for 1926-2019 and 1957-2019 are very similar. Thus, the decline of the early industrialized regions did not begin until after 1957, just before the coal crisis began with the collapse of demand in the winter of 1957/58.

Table 1: Early industrialization and regional economic development, 1926-2019

| | 1926-2019<br>(1) | 1957-2019<br>(2) | 1926<br>(3) | 1957<br>(4) | 2019<br>(5) |
|---|---|---|---|---|---|
| **Panel A. OLS** | | | | | |
| Employment share industry 1882 | -19.33*** | -22.26*** | 15.31*** | 18.24*** | -4.02*** |
| | (3.06) | (2.83) | (2.79) | (2.75) | (1.51) |
| | [2.76] | [2.93] | [2.31] | [2.68] | [1.73] |
| **Panel B. 2SLS** | | | | | |
| Employment share industry 1882 | -30.04*** | -33.62*** | 12.22*** | 15.80*** | -17.82*** |
| | (7.52) | (7.24) | (3.25) | (3.14) | (6.89) |
| | [5.75] | [6.18] | [2.96] | [3.14] | [6.03] |
| **Panel C. Reduced form** | | | | | |
| Log access to coalfields | -65.12*** | -72.87*** | 26.50*** | 34.25*** | -38.62*** |
| | (10.20) | (8.77) | (9.24) | (9.90) | (8.64) |
| | [12.76] | [11.03] | [11.26] | [11.18] | [7.58] |

Notes: The table shows results from OLS (Panel A) and 2SLS (Panel B) regressions of the effect of early industrialization on (changes in) regional economic development as well as reduced form regressions of the effect of coal access on development (Panel C). The ranking in 1926 is based on sales per capita, the rankings in 1957 and 2019 are based on GDP per capita. The 1882 employment share in industry, our explanatory variable of interest, is standardized with a mean of zero and a standard deviation of one. The instrument used in the 2SLS regressions is the log weighted average distance to European coal fields where coalfield sizes serve as weights (see Section 2). All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors. The first-stage Kleibergen-Paap F-statistics for the 2SLS estimates are 19.72 with Conley standard errors and 11.82 with clustered standard errors.

Columns (3) to (5) show that the decline of the early industrialized regions is not only due to the waning positive effects of early industrialization but also to its long-run adverse effects. The 2SLS estimate in panel B of Column (3) implies that a one standard deviation increase in the 1882 industrial employment share improves the 1926 rank in the income distribution by 12.2 percentiles. The 2SLS estimate is very similar to the OLS estimate. Thus, early industrialization was still conducive to economic development in 1926. The same is true for 1957 (see Column (4)). In 2019, on the other hand, early industrialization harmed the position in the income distribution (Column (5)). While the OLS estimate in Panel A is relatively small, the 2SLS estimate is sizable (Panel B), mirroring the significant negative reduced-form estimate in Panel C. The 2SLS estimate implies that a one standard deviation increase in the 1882 industrial employment share reduces a labor market's rank in the 2019 income distribution by 17.8 percentiles.

Overall, our analysis shows that early coal-based industrialization had a long-lasting positive effect on economic development that eventually turned negative. Figure 5 plots the coefficient estimates of $\beta_t$ from equation (2) for 1926, 1935, 1950, 1957, 1961, 1970, 1980, 1992, 2000, 2010, and 2019. The positive effect of early industrialization persisted (and even increased slightly over time) in 1926-1957.[21] The positive effect of industrialization in the 19th century began to shrink in 1957, paralleling the onset of the coal crisis. The point estimate turns negative in 1992 and has been declining ever since.

Table 1 shows that the 2SLS estimates are comparable to the OLS estimates in 1926 and 1957 but are markedly more negative in 2019. What is the reason for this pattern? In this context, it is worth highlighting that the IV estimate in Franck & Galor (2021) of the effect of early industrialization on regional GDP per capita in France "reverses the OLS estimates [...] from positive to negative" in 2001-2005.[22] In contrast, for earlier periods, the positive effect of past

[21]The estimated coefficient $\beta_t$ increases from 8.9 in 1935 to 15.3 in 1950 and 15.8 in 1957. This increase does not reflect a change in the variable used to construct the percentile ranks because we use per capita sales for 1935 and 1950.

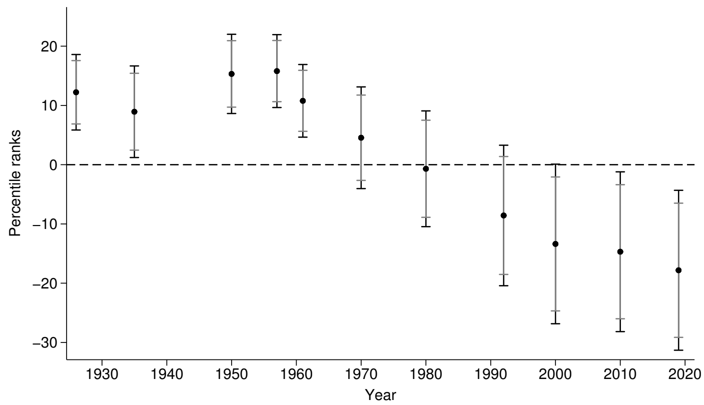
[22]Also Esposito & Abramson (2021) find that the negative effect of historical coal mining on current income is 2-3 times larger (in absolute magnitudes) than the OLS estimates in their IV regressions.

industrialization on economic development in Franck & Galor (2021) is much larger in the IV regressions than the corresponding OLS estimates. The authors argue that omitted characteristics may have harmed development in earlier periods but may be conducive to development today. As an example, they cite government capacity, which may have promoted industrialization but also reduced per capita income in the past by facilitating the protection of the agricultural sector. Thus, a first explanation for our result is that early industrialization is correlated with omitted local characteristics that promote economic development today but had little impact in 1926 and 1957.

A second related explanation is that the exogenous aspects of early industrialization—isolated in the 2SLS regressions by coal access—have a more negative effect than the OLS estimate suggests, while the endogenous aspects—which are independent of coal—favor economic development today. In this explanation, the differences between the two aspects of early industrialization emerge only after the crisis of old German heavy industry in the late 1950s and 1960s (because only then do the two aspects differ in their effects). A third, less favorable explanation is that access to coal picks up local characteristics associated with longer-term decline after 1957 (but not earlier). We address this concern in several robustness checks that consider the effects of other large shocks that asymmetrically affected German regions in the postwar period.

**Robustness checks.** We run several tests, listed in Appendix Table A-3, to assess the robustness of our 2SLS results. As mentioned earlier, a major concern is that our instrumental variable is correlated with "spatial shocks" that could explain differences in regional economic development after 1957. To address this concern, we add control variables that proxy for these shocks in an initial set of checks in Panel B. First, we add control variables that account for the consequences of German division and re-unification (see Panel B1). Specifically, we add distance to the former inner-German border that separated West and East Germany between 1949 and 1990, a direct measure for the market access lost (and gained) due to the division (and later reunification) of Germany based on Redding & Sturm (2008), and distance to Germany's eastern border. We also

Figure 5: Impact of early industrialization on the per capita income rank, 1926-2019




Notes: The figure plots the $\beta_t$ coefficients from 2SLS estimations of equation (2). Point estimates are marked by a dot. The vertical bands in gray and black indicate 90% and 95% confidence intervals, respectively. The dependent variable is the percentile rank in the income per capita distribution. The 1882 employment share in industry, our explanatory variable of interest, is standardized with a mean of zero and a standard deviation of one.

add an indicator variable for redevelopment regions that after the war received governmental aid for reconstruction and industry promotion. Many of these regions were heavily war damaged or suffered from being close to the inner-German border. Second, we add control variables for different immigration episodes (see Panel B2).[23] We add the population share of displaced persons in 1950 and of GDR refugees in 1961, the share of foreigners in 1970 and 1987 as a measure of the local influx of guest workers,[24] and the share of refugees in 2016. Third, we proxy war damage by the amount of rubble per capita in 1946 and by the fraction of dwellings built before 1945 that were damaged in the war (see Panel B3). Finally, we account for Germany's integration into (Western) Europe by adding distance to Germany's western border and to the Schengen area in 2008 (see Panel B4). Our results are robust to all of these checks.

A related concern is that our instrumental variable picks up omitted geographic characteristics that affected economic development differently in 1926, 1957, and 2019. We consider this less of

[23]The regional distribution of migrants might be endogenous, as it follows regional economic development. For example, the regional quotas for the distribution of refugees in 2015/16 are two-thirds based on local tax revenues.

[24]By far the most foreigners in West Germany at the time came to the country as part of the guest worker program. The recruitment of guest workers ended in 1973.

a threat to identification because our baseline regression already controls for a region's favorable location. Nevertheless, in our second set of robustness checks in Panel C, we add several controls for local geographic and climatic conditions to our baseline specification, including distance to rivers and oceans, soil quality, remoteness within Germany, and mean sunshine hours. None of these controls has a marked impact on our baseline estimate.

Our instrumental variable may also be correlated with state-level policies that have affected economic development in the postwar period. Therefore, in a third set of checks in Panel D of Appendix Table A-3, we add state-level fixed effects to our controls (with and without additional geographic controls). We are reluctant to include state-level fixed effects in our baseline specification because they eliminate most of the variation we are interested in (much of the change in relative prosperity occurs between distant regions in different states). Nevertheless, our main estimates are only slightly smaller when we add fixed effects. We find that a one standard deviation increase in the share of industrial employment in 1882 led to a worsening of the rank in income distribution by 22.4 (without additional geographic controls) and 25.5 (with geographic controls) percentiles between 1957 and 2019. Adding fixed effects for states in pre-industrial Germany yields very similar results. Moreover, accounting for early membership in the German Customs Unions of 1834, which promoted market integration among member states (Keller & Shiue, 2014), leaves our baseline results unchanged.

Appendix Table A-4 reports additional robustness checks. In Panel B we measure early industrialization in the subsequent 1895 and 1907 occupation censuses and add the tertiary employment share in 1882 as an additional control. Panel C constructs the instrument using alternative cost vectors for the least-cost paths to the coal fields.[25] It also uses alternative measures of coal deposits to address the concern that the location of historical coal fields, measured in our baseline regression as of 1931, may be endogenous to exploration efforts (as

[25]Among other things, we construct the instrument using squared transport costs. This alternative instrument gives less weight to more distant coal fields, reflecting the fact that immediate access to coal may be more important than proximity to multiple coal fields.

noted by Fernihough & O'Rourke, 2020).[26] The miscellaneous checks in Panel D exclude the Ruhr from our sample, estimate population-weighted regressions, and use log GDP per capita as the dependent variable instead of percentile ranks. Our main result is robust to all of these checks: early industrialization led to a sharp decline in per capita income after 1957, driven in large part by a negative income effect in the long run.

**Effect size.** Using GDP per capita as the dependent variable (instead of ranks) also sheds additional light on the effect size. According to the estimates in Panel D of Appendix Table A-4, a one standard deviation increase in the share of industrial workers in 1882 raised GDP per capita by 0.16 and 0.13 log points in 1926 and 1957, respectively. However, today's GDP per capita falls by 0.11 log points. Therefore, real GDP per capita increased by about 0.24 log points or 27% less in 1957-2019 (regional GDP per capita increased on average by 313% or almost 40,000 DM in 1957-2019).

**Spillover effects.** Finally, we re-estimate our regression at the more aggregate level of districts (*Regierungsbezirke*), of which there are 36, to test for spillover effects across labor markets (which our baseline specification ignores). Positive spillovers could result from agglomeration effects that radiate across labor market regions. Negative spillovers could arise from the relocation of economic activity to early industrializing regions (or away from them when they decline). Differences between labor market and district level estimates would hint at such regional spillovers. However, the results in Panel E of Appendix Table A-4 are relatively similar to those at the labor market level and we continue to find negative long-run effects of early industrialization.[27]

[26] Our focus on distance to all European coal fields should mitigate this concern, since, for example, coal mining in England, crucial to northern Germany, is unlikely to have been driven by German industrialization. Nevertheless, the robustness checks in Panel C address the concern of endogenous coal location in two ways: First, we use the distance to carboniferous surface strata rather than the actual coal fields, though this results in a weaker first stage, with a Kleibergen-Paap F-statistic of 10.20. While many coal fields are located in areas with carboniferous strata at the surface, others lie below the surface and are covered by other strata. In our data, only about half of the 1931 coal fields are located in areas with carboniferous surface strata. Second, we supplement the 1931 data with later data from the Pergamon World Atlas of 1967 (Pergamon, 1968). The combined list should give a reasonably accurate picture of where coal could have been discovered in the 19th century.

[27] Comparing estimates at different levels of aggregation is a common strategy for detecting regional spillovers (see, for example, Criscuolo, Martin, Overman & Van Reenen, 2019; Franck & Galor, 2021). The approach assumes

## 4 Lopsided economic structure and limited adjustment capacity

What explains the slow and protracted decline of the early industrialized regions in Germany? And why were these originally rich regions unable to revitalize their economic base? Existing hypotheses point to the regions' one-sided economic structure. Old industrial regions, the argument goes, were historically characterized by monostructural agglomerations, typically in heavy and extractive industries, dominated by large corporations (Hu & Hassink, 2016). This lopsided structure led to high adjustment pressure after Germany's old industries fell into crisis in the late 1950s and 1960s. At the same time, it limited the regional ability to adapt and innovate (e.g. Junkernheinrich, 1989; Hamm & Wienert, 1990; Grabher, 1993). This section provides tentative evidence consistent with these hypotheses.

**Economic structure in 1907.** Using detailed three-digit sectoral employment data from the 1907 manufacturing census, Table 2 provides evidence that (coal-based) early industrialization indeed produced a large-firm-dominated and highly concentrated economic structure. Column (1) shows that early industrialization led to a larger average firm size in industry. A one standard deviation increase in the share of employment in industry in 1882 increased the average firm size by 2.9 employees per firm in 1907 (compared to a mean of 4.6). Similarly, Columns (2) and (3) show that early industrialized regions had higher employment shares in sectors dominated by large firms. We focus on sectors with a Germany-wide employment share of at least 50% in firms with at least 501 (Column (2)) or 1000 (Column (3)) employees. The effect sizes are considerable: a one standard deviation increase in the industrial employment share in 1882 increased the 1907 employment share in sectors dominated by firms with at least 501 employees by 7.1 percentage points (relative to a mean of 2.3%) or more than one standard deviation. Finally, Column (4)

that spillovers occur only across labor markets within the same district. While it is possible that spillovers also affect more distant areas, it is likely that they are strongest between nearby labor markets. The lower district-level coefficient for 1926 and 1957 suggests negative spillovers, but the large standard errors caution against drawing firm conclusions. Nevertheless, in unreported regressions we find that after the onset of the coal crisis, late industrializing regions experienced higher net migration in 1961-1970 (although we cannot distinguish between migration from other German regions and from abroad).

shows that early industrialized regions were much more specialized within the industrial sector in 1907. A one standard deviation increase in early industrial employment increased the HHI of industry concentration by 0.05 (relative to a mean of 0.06).[28]

Table 2: Early industrialization and industry structure in 1907 (2SLS estimates)

| | | Employment share in sectors dominated by large firms | | | |
|---|---|---|---|---|---|
| | Average firm size | $\geq$ 501 workers | $\geq$ 1000 workers | HHI index of industry concentration | First principal component of (1)-(4) |
| | (1) | (2) | (3) | (4) | (5) |
| Employment share industry 1882 | 2.875*** | 0.071*** | 0.060*** | 0.049** | 2.249*** |
| | (0.554) | (0.019) | (0.020) | (0.022) | (0.654) |
| | [0.542] | [0.019] | [0.021] | [0.027] | [0.691] |
| **Outcome statistics** | | | | | |
| Mean | 4.592 | 0.023 | 0.012 | 0.058 | -0.000 |
| Standard deviation | 2.394 | 0.057 | 0.049 | 0.062 | 1.771 |
| **Causal mediation analysis** | | | | | |
| Total effect | | | | | -33.62*** [6.18] |
| Direct effect | | | | | -8.49 [6.03] |
| Indirect effect | | | | | -25.13** [12.81] |

Notes: The table shows results from 2SLS regressions of the effect of early industrialization on the regional economic structure in 1907. The dependent variables are the average number of employees per firm in industry (Column (1)), the employment share in sectors in which at least 50% of employees work in firms with at least 501 workers (Column (2)) or 1000 workers (Column (3)), the HHI-Index of industry concentration with $\alpha = 2$ (Column (4)), and the first principal component of the four indicators in Columns (1) to (4) (Column (5)). The 1882 employment share in industry, our explanatory variable of interest, is standardized with a mean of zero and a standard deviation of one. The lower panel also presents second stage results of the causal mediation framework for linear IV models introduced in Dippel, Gold, Heblich & Pinto (2020a). The mediation analysis decomposes the total effect of early industrialization on the change in the GDP per capita rank between 1957 and 2019 into a direct effect and an indirect effect running through lopsided economic structure. All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

Existing hypotheses suggest that the negative impact of early industrialization on long-term economic development is mediated by the lopsided economic structure it created. Unfortunately, quantifying this mediation effect is only possible under strong assumptions. Typically, causal

[28] We still observe a positive effect on sectoral concentration in 1950, shortly before the coal crisis began (see Column (1) of Appendix Table A-8).

mediation analysis focuses on settings where the treatment is exogenous given the pretreatment covariates (Celli, 2022).[29] Since this assumption does not hold in our context, we rely instead on recent results in Dippel et al. (2020a). The authors present a framework for estimating causal mediation effects in IV settings with endogenous treatment using a single instrument.[30] The key additional identifying assumption is that unobserved confounding variables that jointly cause the treatment (early industrialization in our case) and the mediator (lopsided economic structure) are independent of confounding variables that jointly cause the mediator (lopsided economic structure) and the outcome (change in GDP per capita ranking). This assumption is violated if a common confounder jointly causes the treatment, the mediator, and the outcome.

The method developed by Dippel et al. (2020a) can only identify the mediating effect of a single mediator (but omitting other mediators does not bias the results). Thus, following Dippel et al. (2021), we focus on the first principal component of our four economic structure variables as the mediating variable.[31] Column (5) of Table 2 shows that a one standard deviation increase in the industrial employment share in 1882 increases the principal component by more than one standard deviation. Applying the mediation analysis of Dippel et al. (2020a) suggests that about three-quarters of the total effect of early industrialization on the change in GDP per capita ranking between 1957 and 2019 works through the effect of industrialization on economic structure.[32] We find this value plausible and consistent with the strong focus in the historical

[29] In addition, standard causal mediation analysis requires that the mediator is exogenous, or independent of potential outcomes, given the actual treatment and the pretreatment covariates (Imai, Keele & Yamamoto, 2010). This sequential ignorability assumption is not guaranteed to hold even when treatment and mediator are randomized.

[30] See Dippel, Ferrara & Heblich (2020b) for implementation details and Dippel, Gold, Heblich & Pinto (2021) for an application. The framework complements alternative IV approaches that require two separate instruments, one for the treatment and one for the mediator (Frölich & Huber, 2017).

[31] This first principal component explains 78% of the total variation and is the only component with an eigenvalue greater than one.

[32] To get at the mediation effect, we regress the change in GDP per capita ranking on our measure of lopsided economic structure as instrumented by access to coal, controlling for covariates and early industrialization. The effect of economic structure on the GDP per capita ranking is then multiplied by the effect of early industrialization on economic structure. Appendix Table A-5 shows all the underlying estimates. In addition, the table also reports the results of mediation analyses that directly use our four measures of lopsided economic structure as mediating variables, one at a time. In these alternative specifications, lopsided economic structure explains 56-94% of the total effect of early industrialization.

literature on the negative effect of the lopsided economic structure of early industrializing regions dominated by large firms, but reiterate the strong identifying assumptions required.

**Adjustment pressure and capacity.** What were the consequences of the lopsided economic structure that led to the decline of Germany's early industrialized regions in the second half of the 20th century? It has been hypothesized that the dominance of large corporations in heavy industries created high adjustment pressure post-1957 and limited adjustment and innovation capacity (Junkernheinrich, 1989; Hamm & Wienert, 1990; Grabher, 1993).

Large corporations and close interregional business ties dominated local economies in early industrializing regions. These tightly knit industrial networks created a "cognitive lock-in" that prevented the regional economy from adapting (Grabher, 1993). Large corporations created a tradition of dependent employment and a corresponding lack of entrepreneurial role models (Stuetzer et al., 2016). As a result, they crowded out entrepreneurial activity and dampened growth in the long run (Chinitz, 1961; Glaeser et al., 2015). Large corporations were also closely tied to local governments and unions that supported the old mining industry structures. Extensive subsidies to the coal, iron, and steel complex maintained outdated structures (Hamm & Wienert, 1990; Hassink & Kiese, 2021).

The dominance of large corporations in Germany's industrial heartland has been contrasted with the "decentralized industrial order" prevalent especially in southwestern Germany (Herrigel, 2000). The small and medium-sized enterprises that dominate there[33] often emphasize customized quality production, requiring high skill and flexibility in production (Herrigel, 1999, 2000). Flexible, decentralized production, in turn, has been cited as a key factor for adapting to rapidly changing global market environments and thus for sustained competitive success during the globalization process (Streeck, 1991; Simon, 1992; Herrigel, 2000; Simon, 2009).

[33] In addition to the lack of coal, the emergence of a decentralized industrial order has been fostered historically by equal division inheritance rules that divided property equally among all children (Herrigel, 2000; Bartels, Jäger & Obergruber, 2024). Farmers with small landholdings in equal division areas often supplemented their agricultural incomes with income from secondary industrial employment. In the long run, this industrial by-employment formed the breeding ground for adaptive small- and medium-sized firms (Bartels et al., 2024).

Undisputed, the old industrial regions were exposed to high adjustment pressure after 1957. Table 3 shows that in 1950, shortly before the coal crisis outbreak, early industrialization's positive effect on industrial employment was exclusively attributable to the coal, iron, and steel industries. A one standard deviation increase in early industrialization raised the industry employment share by 7.4 percentage points in 1950 (Column (1)). However, the effect on iron, coal, and steel industries was even larger at 8.0 percentage points (Column (2)). It was precisely these industries that fell into crisis in the 1960s and 1970s. In contrast, early industrialization slightly reduced employment in "modern industries", such as electrical engineering, the automobile industry, and the chemical industry, already in 1950 (Column (3)). These modern industries continued to flourish after 1945 and are still associated with Germany's economic strength today. Column (5) of Table 3 shows that Germany's former industrial heartland has failed to revitalize its industrial base in new growth industries and has suffered from deindustrialization. We find that a one standard deviation increase in early industrialization lowers the industrial employment share by 5.9 percentage points in 2019.[34]

Do we also find evidence of limited regional adaptive capacity, possibly reflected in lower human capital, lower entrepreneurial activity, a rigid policy environment, and ultimately lower innovation? Previous work on the long-term effects of industrialization suggests adverse effects on human capital. Franck & Galor (2021) show that early industrialization has a negative impact on the share of the population in France with a post-secondary degree (from a vocational school or university). They also show that second-generation migrants whose parents originally came from historically industrial departments have lower human capital aspirations today and are more likely to leave the school system at the end of middle school. Esposito & Abramson (2021)

[34]Appendix Figure A-3 shows that early industrialization positively affected employment in industry only until 1970. After that, the effect becomes increasingly negative, consistent with earlier results for France (Franck & Galor, 2021). The figure also shows that the negative effect on industrialization is accompanied by positive effects on employment in services. Appendix Figure A-4 shows that industrial employment peaks around 1970 for both regions with above- and below-median 1882 industrial employment shares. In fact, more than 70% of all labor market regions record their peak industrial employment share in 1970, and another 20% record their peak in 1961 or 1980 (the data points before or after 1970). However, the early industrializing regions experience much greater deindustrialization and have lower industrial employment shares today.

Table 3: Early industrialization and adjustment pressure (2SLS estimates)

| | Industrial employment share (%) | | | | |
|---|---|---|---|---|---|
| | 1950 | | | | 2019 |
| | Total<br>(1) | Coal, iron and steel<br>(2) | Modern industries<br>(3) | Other industries<br>(4) | Total<br>(5) |
| Employment share industry 1882 | 0.074*** | 0.080*** | -0.004 | -0.002 | -0.059*** |
| | (0.012) | (0.016) | (0.007) | (0.012) | (0.018) |
| | [0.013] | [0.016] | [0.005] | [0.010] | [0.016] |
| **Outcome statistics** | | | | | |
| Mean | 0.310 | 0.032 | 0.060 | 0.218 | 0.287 |
| Standard deviation | 0.110 | 0.065 | 0.037 | 0.077 | 0.074 |

Notes: The table shows results from 2SLS regressions of the effect of early industrialization on employment shares in industry. The dependent variable is the total industrial employment shares in 1950 (Column (1)) and 2019 (Column (5)), and the 1950 shares in coal, iron, and steel (Column (2)), in "modern industries" (Column (3)), and in all other industries (Column (4)). Modern industries encompass mechanical engineering, road vehicle and aircraft construction, electrical engineering, precision mechanics, optics, the chemical industry, and plastics. The 1882 employment share in industry is standardized with a mean of zero and a standard deviation of one. All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

find that former coal mining regions in Europe invested less in tertiary education.

In our context, it is unlikely that a lack of investment in tertiary education, which in Germany is mainly financed by the federal states rather than the local governments,[35] explains the adverse long-term effects of early industrialization. In the Ruhr region, for example, universities were opened in Bochum (1965), Dortmund (1968), Duisburg (1972), and Essen (1972). Today, the Ruhr area is the largest university region in Germany (Kriegesmann, Böttcher, & Lippmann, 2016). Appendix Figure A-5 shows that early industrializing regions indeed tend to have more universities today, especially larger ones with 7,500 or more students. Consequently, we find that the number of students enrolled as a share of the population is higher in early industrialized

[35]The same holds for spending on schools. In 2020, 53.2% of total education spending was financed by the states, 11.8% by the federal government, 18.4% by local governments, and 16.4% by the private sector. The state's share of funding was particularly high at 78.4% for general education schools and 69.4% for university programs (Statistisches Bundesamt, 2023).

regions in 2019, a finding that holds in all subjects (see Table A-6 in the appendix).

Appendix Table A-7 confirms that early industrialized regions do not have a lower share of individuals with a university degree today (Columns (1)-(2)).[36] However, we find that early industrialization led to a lower share of individuals with secondary or vocational degrees (Columns (3)-(4)) and increased the share of school dropouts without a school-leaving qualification (Columns (5)-(6)). The effect magnitudes are economically relevant: A one standard deviation increase in early industrialization increased the share of high school dropouts in 2011 by 1.1 percentage points (compared to a mean of 4.5%), or almost one standard deviation.

Outside the old industrial regions, with their large-scale heavy industry, an extensive vocational and technical training system ensured a steady supply of skilled labor for small and medium-sized enterprises (Herrigel, 2000). Consistent with this argument, we find that the employment share of industrial apprentices is much lower in early industrialized regions. A one standard deviation increase in early industrialization reduced the share of apprentices in industrial employment in 1970 by 1.9 percentage points relative to an average of 5.4% (see Column (1) of Appendix Table A-8). Vocational education and training, in turn, is an essential institutional prerequisite for the flexible production of diversified, high-quality products, often seen as the key to Germany's ability to adapt to rapid product and technological change (Streeck, 1991; Herrigel, 2000).

Large firms and the lack of smaller suppliers are seen as key factors dampening local entrepreneurship (Chinitz, 1961; Glaeser & Kerr, 2009; Glaeser et al., 2015; Stuetzer et al., 2016).[37] Consistent with this argument, Appendix Table A-8 shows that early industrializing regions had low entrepreneurial activity, as measured by self-employment rates, in 1950, just before the onset of the coal and later steel crisis. A one standard deviation increase in early industrialization led to a 3.9 percentage point decline in the self-employment rate in 1950 (Column

[36]We use data on individuals with a university degree from the 1970 and 2011 censuses. Earlier censuses did not ask about educational attainment. The coefficient for 1970, during the university expansion, is negative and borderline significant. The coefficient estimate for 2011, after the expansion, is very close to zero. Note that the null effect of early industrialization on tertiary education may reflect two opposing forces: lower preference for, but better supply of, university education (we thank an anonymous reviewer for this suggestion).

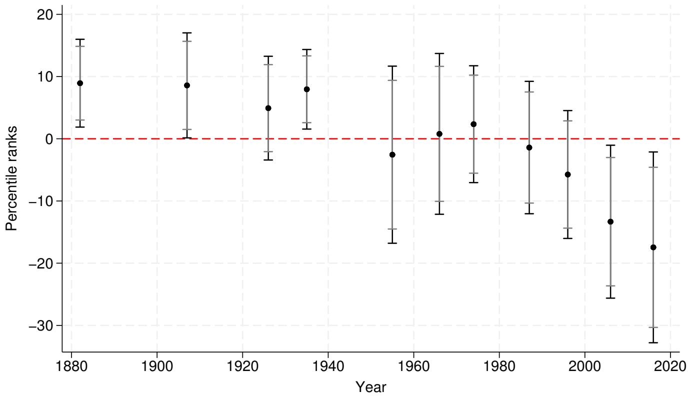
[37]See, for instance, Glaeser & Kerr (2009) and more recently Barrios, Hochberg & Macciocchi (2021) for studies comparing different potential determinants of local entrepreneurial activity.

(3)), relative to a mean of 16.8%. This result is consistent with earlier evidence in Fritsch & Wyrwich (2014) who show that already in 1925, self-employment rates were much lower in the Ruhr region, historically characterized by large-scale heavy industry, than in southwestern Germany. Such local differences in entrepreneurship were very persistent in Germany between 1925 and 2005, possibly testifying to the long-lasting impact of a local entrepreneurial culture.

Moreover, we also find that early industrialization led to a rigid policy environment with little political change (see Columns (4)-(6) of Appendix Table A-8). We find that in the old industrial regions, it was less likely that the mayoralty was held by a party other than the dominant one. Of the major parties, the social democratic SPD benefited most from early industrialization, both in local and federal elections. The SPD has traditionally represented the interests of the working class, especially unionized workers. Maintaining the heavily unionized coal and steel sector was in the interest of politicians and unions struggling to defend their political base. Overall, adherence to the existing economic structure was the consensual denominator of local action by companies, politicians, and unions (Junkernheinrich, 1989; Grabher, 1993).

Finally, Figure 6 shows the impact of early industrialization on per capita patenting activity over time (taken from Bergeaud & Verluise, 2024). Although the early industrialized regions were much more prosperous and more industrialized immediately after World War II, they did not have a head start in patenting activity. They then gradually fell behind the later industrialized regions. Today, a one standard deviation increase in industrial employment in 1882 is associated with a 17.5 percentiles lower rank in the per capita patent distribution. This result goes hand in hand with the deindustrialization effect reported earlier. Deindustrialization in historically industrial regions hinders local innovation, as patent applications are mainly filed in the industrial sector (Kiese, 2019). Low levels of innovation, in turn, are likely to slow regional economic growth (e.g., Akcigit, Grigsby & Nicholas, 2017). Today, the value added per employee in the industrial sector is almost 20 percent higher than in the German economy as a whole.

Figure 6: Impact of early industrialization on the rank in the patent per capita distribution, 1882-2016




Notes: The figure plots slope coefficients from 2SLS estimations of patents per capita on early industrialization. The dependent variable is the percentile rank in the patent per capita distribution. Point estimates are marked by a dot. The vertical bands in gray and black indicate 90% and 95% confidence intervals, respectively. The 1882 employment share in industry, our explanatory variable of interest, is standardized with a mean of zero and a standard deviation of one.

## 5 North-South reversal and changing inequality

We conclude our analysis by exploring whether regional differences in 19th-century industrialization can explain two empirical patterns that have received much public attention: the widening economic gap between northern and southern Germany and the return of regional inequality in recent decades.

**Germany's reversal of fortune.** First, consider the growing North-South divide. The regions that lost ground in the income distribution between 1926 and 2019 are predominantly in the West and North, while the winning regions are predominantly in the South.[38] The North-South divide, which is the subject of current political debates (e.g., Schrader & Laaser, 2019), has thus

[38] In fact, the declining labor markets in Figure 1 (dots below the dotted bisector) are concentrated in northern Germany, while the rising labor markets (dots above the bisector) are concentrated in the south. Appendix Figure A-6 illustrates this fact by distinguishing between northern and southern labor markets in a plot of percentile rank in 2019 versus rank in 1926.

emerged over the last 100 years or so. How much of this divide can be attributed to differences in 19th-century industrialization?

Table 4 compares the actual gap with the predicted gap resulting from our main 2SLS estimates (in Table 1). Panel A measures the North-South gap as the average difference in percentile ranks between northern and southern regions.[39] The predicted gap is the average difference that would have resulted if the regions had differed only in their 1882 industrial employment share.[40] As shown, the actual mean difference evolves from +20.7 percentiles in 1926 to −18.2 percentiles in 2019. The mean difference in predicted income ranks follows a similar, though less pronounced, trajectory. Panel B reports average differences in (log) income per capita. While northern labor markets had, on average, 0.141 log points higher GDP per capita in 1957, they trailed southern labor markets by 0.098 log points in 2019 (recall that the 1926 value is not directly comparable to the 1957 and 2019 values when considering levels).

Our estimates imply that if the regions had differed only in the 1882 share of employment in industrial occupations, per capita income in 1957 would have been 10.0 percentile ranks or 0.085 log points higher in the North than in the South. This northern advantage in predicted income is due to the fact that the average industrial employment share in 1882 was much larger in the North than in the South (0.223 versus 0.158) and that early industrialization still had a positive effect on economic development in 1957. By 2019, the average differences in predicted income have become negative at −11.2 percentile ranks or −0.067 log points, reflecting the adverse effect of early industrialization on current development. A "back-of-the-envelope" calculation suggests that early industrialization explains about 62% (=11.2/18.2) of the current North-South gap

[39] Our baseline definition of northern labor markets adheres to state boundaries. It classifies as northern the labor markets located in Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, and Schleswig-Holstein. Southern labor markets are those in Bavaria, Baden-Württemberg, Hesse, and Rhineland-Palatinate (see Appendix Figure A-7 for an illustration). Appendix Figure A-8 shows that the reversal of fortune does not hinge on this specific classification. It remains visible if we also assign the northern parts of Hesse and Rhineland-Palatinate to the North or if we use latitude as a continuous measure to divide labor markets into northern and southern ones.

[40] We first use the 2SLS estimate of $\beta_t$ in equation (2) to predict each labor market's income rank, given its 1882 industrial employment share. We then estimate the slope coefficient $\tilde{\beta}_t$ from a regression of the *predicted* income rank on $N_i$, where $N_i$ equals one if labor market $i$ is in the North and zero otherwise. The estimate yields the mean difference in the predicted GDP per capita rank between the northern and southern labor markets.

Table 4: Mean differences in per capita income between northern and southern labor markets

| | 1926<br>(1) | 1957<br>(2) | 2019<br>(3) |
|---|---|---|---|
| **Panel A. Income per capita rank** | | | |
| Actual mean difference | 20.67 | 15.31 | -18.24 |
| | (5.23) | (7.33) | (5.47) |
| Predicted mean difference | 7.70 | 9.95 | -11.23 |
| | (3.97) | (5.13) | (5.78) |
| **Panel B. Log income per capita** | | | |
| Actual mean difference | 0.300 | 0.141 | -0.098 |
| | (0.074) | (0.058) | (0.034) |
| Predicted mean difference | 0.100 | 0.085 | -0.067 |
| | (0.052) | (0.044) | (0.035) |

Notes: The table reports actual and predicted mean differences in per capita income ranks (Panel A) and log income (Panel B) between northern and southern labor markets in 1926 (Column (1)), 1957 (Column (2)), and 2019 (Column (3)). The actual and predicted mean differences are the slope coefficients from a regression of actual and predicted income, respectively, on an indicator variable for northern labor markets. We calculate predicted income as $\hat{\beta}_t I_{i,1882}$. Income in 1926 is proxied by sales per capita, income in 1957 and 2019 are based on GDP per capita. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets.

when the latter is measured in income ranks, and about 68% (=0.067/0.098) when we use income levels instead. Thus, the blessing and curse of early industrialization contributed significantly to the North-South reversal. Today, after decades of industrial leadership, the North has a much smaller industrial base than the South (see Appendix Figure A-9 for trends in industrial employment shares over the period 1882-2019).

**Regional inequality.** Next, consider the decline and rise in regional inequality. How much of the overall change in inequality can be attributed to regional differences in early industrialization and its differential impact on economic development over time? To answer this question, we decompose the total change in inequality into an "industrialization effect" and a residual. We calculate the industrialization effect as the difference between observed changes in inequality and the counterfactual change in inequality that would have occurred in the absence of differences in early industrialization. The industrialization effect thus measures the contribution of differences in early industrialization to the change in regional inequality. Following the literature on sigma-convergence within countries (e.g. Sala-i Martin, 1996; Persson, 1997; Young, Higgins & Levy, 2008), we use the unweighted standard deviation of log GDP per capita as our baseline measure of income dispersion. Yet, our results are robust to alternative measures, as we show below.[41]

Let $y_{it}^c$ denote the counterfactual log income per capita of a labor market $i$ in year $t$, i.e., the income that would result if the 1882 share of industrial employment had been equal to the mean across all labor markets.[42] For the period $t-1$ to $t$, the contribution of the industrialization effect to the total change in $\sigma_{y_t}$, the standard deviation of log income per capita, is given by:

$$\begin{aligned} \Delta IND_{t,t-1} &= \overbrace{\left[\sigma_{y_t} - \sigma_{y_{t-1}}\right]}^{\text{Actual change}} - \overbrace{\left[\sigma_{y_t^c} - \sigma_{y_{t-1}^c}\right]}^{\text{Counterfactual change}} \\ &= \underbrace{\left[\sigma_{y_t} - \sigma_{y_t^c}\right]}_{\text{Effect on } \sigma \text{ in } t} - \underbrace{\left[\sigma_{y_{t-1}} - \sigma_{y_{t-1}^c}\right]}_{\text{Effect on } \sigma \text{ in } t-1}. \end{aligned} \tag{4}$$

The second line of equation (4) shows that $\Delta IND_{t,t-1}$ also reflects the difference in the within-year effect of early industrialization on inequality.

Table 5 reports $\sigma_{y_t}$ and $\sigma_{y_t^c}$ for 1957, 1980, and 2019 and the changes in 1957-1980, 1980-2019, and 1957-2019. The first key observation is that the waning effect of early industrialization explains much of the decline in $\sigma$ between 1957 and 1980. In 1957, regional differences in early industrialization increased the dispersion of real per capita income by 0.054 log points (the difference between $\sigma_{y_{1957}}$ and $\sigma_{y_{1957}^c}$ reported in the last row). By 1980, however, this effect had vanished,[43] since early industrialization no longer has a sizeable effect on economic disparity in this year (see Figure 5). Overall, the industrialization effect explains $-0.054$ log points—or

[41]We use unweighted inequality measures because we are interested in inequalities between spatial units (within a country). Instead, population-weighted measures can be interpreted as measures of inequality between groups, where people are grouped according to where they live (Achten & Lessmann, 2020). First proposed by Williamson (1965), population-weighted measures are often used to compare inequality between countries with regions of different sizes.

[42]The standardized industrial employment share, $I_{i,1882}$, is then zero. We calculate the counterfactual log income per capita as $y_{it}^c = y_{it} - \hat{\beta}_t I_{i,1882}$ where $\hat{\beta}_t$ is the 2SLS estimate from the regression model (2) (with log income per capita as the dependent variable).

[43]Earlier work for Germany documented that industrialization increased inequality in the late 19th century (see, e.g., Frank, 1993; Gutberlet, 2014; Braun & Franke, 2022). Together with our results, these findings are consistent with the argument that regional disparities first increase, then stabilize, and finally decrease as industrialization progresses (Kuznets, 1955).

Table 5: Components of changes in regional per capita income disparity, 1957-2019

| | 1957<br>(1) | 1980<br>(2) | 2019<br>(3) | 1957-<br>1980<br>(2)-(1) | 1980-<br>2019<br>(3)-(2) | 1957-<br>2019<br>(3)-(1) |
|---|---|---|---|---|---|---|
| $\sigma_{y_t}$ | 0.234 | 0.162 | 0.191 | -0.073 | 0.029 | -0.043 |
| $\sigma_{y_t^c}$ | 0.181 | 0.162 | 0.212 | -0.019 | 0.050 | 0.032 |
| $\Delta$ | 0.054 | -0.000 | -0.021 | -0.054 | -0.021 | -0.075 |
| | (0.009) | (0.008) | (0.012) | (0.009) | (0.008) | (0.013) |

Notes: The table reports $\sigma_{y_t}$ and $\sigma_{y_t^c}$, the standard deviation of actual and counterfactual log per capita income, for 1957 (Column (1)), 1980 (Column (2)), and 2019 (Column (3)). The last row reports $\sigma_{y_t} - \sigma_{y_t^c}$, i.e., the effect of early industrialization on GDP per capita dispersion in year $t$. The last three columns report changes between 1957-1980, 1980-2019, and 1957-2019, respectively. Cells shaded in gray report the industrialization effect, $\Delta IND_{t,t-1}$, as defined in equation (4). Bootstrapped standard errors based on 200 bootstrap replications are in round brackets.

almost three-quarter—of the actual change in $\sigma$ of -0.073 over the 1957-1980 period.

The second key result from Table 5 is that the industrialization effect cannot explain the increase in regional inequality since 1980. The changes in $\sigma_{y_t}$ and $\sigma_{y_t^c}$ move largely in parallel over the 1980-2019 period. Thus, the increase in regional inequality would have occurred even if regions had not differed in their industrialization paths. If anything, regional differences in early industrialization dampened the increase by reducing current inequality. We find that differences in early industrialization reduced the dispersion of real per capita income by 0.021 log points in 2019. Because labor markets with higher counterfactual per capita income tend to have a higher 1882 industrial employment share, a moderately negative effect of early industrialization on economic development compresses the regional income distribution in our specific context.

Appendix Table A-9 shows that both of our main results also hold when we use the coefficient of variation, the Gini coefficient, and the ratio of per capita income in the labor market at the 90th percentile to that at the 10th percentile as alternative measures of regional inequality. The 90/10 ratio yields even more pronounced results than our baseline measure, with the dwindling effect of industrialization now explaining all of the decline in regional inequality between 1957 and 1980. Finally, Appendix Table A-10 shows that our results are also robust to an alternative decomposition that isolates the change in inequality due to changes in $\beta_t$, the effect of early industrialization on income. Under this alternative decomposition, we again find that all of the decline in regional inequality between 1957 and 1980 is explained by the diminishing effect of industrialization—as captured by the decline in $\beta_t$.

## 6 Conclusion

In recent decades, the spatial distribution of economic activity has changed fundamentally in many advanced countries. Germany is no exception. The former industrial powerhouses in the northwest of the country began their long and protracted decline after World War II. While regional per capita incomes converged strongly in the 1960s and 1970s, regional inequality in West Germany has increased again in recent decades. Such protracted decline and rising regional inequality are often seen as causes of populism in Europe and the US (Rodríguez-Pose, Terrero-Dávila & Lee, 2023).

This paper has shown that the profound changes in the economic geography of West Germany cannot be understood without taking into account the long-lasting legacy of regional differences in 19th-century industrialization. Early industrialization still strongly favored regional economic development in 1957. However, the positive effect diminished between 1957 and 1980, and industrialization became a drag on economic development at the turn of the 21st century. For identification, we exploit variation in access to coal resources in Europe, while controlling for connectedness to European markets. A one standard deviation increase in coal-based industrialization improved the rank of a labor market in the West German income distribution by 15.8 percentiles in 1957 but reduced the rank by 17.8 percentiles in 2019. Thus, early industrialization led to a massive decline in per capita income rank after World War II.

The initial blessing and later curse of early industrialization strongly influenced key trends in regional inequality. We show that the declining advantage of early industrializing regions significantly reduced regional inequality between 1957 and 1980. Moreover, the North-South divide that has emerged in the last forty years cannot be understood without recourse to regions' paths to industrialization 140 years ago. Our estimates suggest that differences in early industrialization can explain more than half of the current North-South gap.

Our results have important implications for the policy discourse on regional inequality and economic decline. First, they illustrate that the interpretation of contemporary changes in regional inequality, which have received much attention recently (Iammarino et al., 2018; Floerkemeier et al., 2021), requires careful consideration of the past. Not only do development processes have lasting effects, but they can also lead to future changes. Second, our results show that initial gains from industrialization may come at the expense of long-run losses (see also Matheis, 2016; Franck & Galor, 2021). This intertemporal trade-off raises the question of whether policy interventions can prevent adverse effects in the long run. Indeed, Germany has managed to avoid the shortage of university graduates in early industrialized regions observed elsewhere in Europe (Esposito & Abramson, 2021), presumably by establishing new universities in its former industrial heartland. Third, our results suggest that early industrialization favored long-term economic decline by creating a lopsided economic structure dominated by large firms in a few heavy industries. This finding aligns well with prior results for the U.S. and Great Britain that the historical presence of large firms hinders entrepreneurship and growth (Chinitz, 1961; Glaeser et al., 2015; Stuetzer et al., 2016). A more diversified economic structure—as in the Italian industrial triangle of Piedmont, Lombardy, and Liguria (Fenoaltea, 2003)—could thus avoid the negative long-term effects of industrialization in the first place. After all, Italy's old industrial triangle still generates above-average GDP per capita (e.g. Felice, 2018), in stark contrast to Germany's old industrial heartland.

## References


Abelshauser, W. (1984). *Der Ruhrkohlenbergbau seit 1945. Wiederaufbau, Krise, Anpassung.* C.H. Beck.

Achten, S. & Lessmann, C. (2020). Spatial inequality, geography and economic activity. *World Development*, 136, 105114.

Akcigit, U., Grigsby, J., & Nicholas, T. (2017). The rise of American ingenuity: innovation and inventors of the Golden Age. *NBER Working Papers* 23047, National Bureau of Economic Research.

Alder, S., Lagakos, D., & Ohanian, L. (2023). Labor market conflict and the decline of the Rust Belt. *Journal of Political Economy*, 131(10), 2780–2824.

Arbeitskreis VGR der Länder (2021). *Bruttoinlandsprodukt, Bruttowertschöpfung in den kreisfreien Städten und Landkreisen der Bundesrepublik Deutschland 1992 und 1994 bis 2019*. Stuttgart: Statistisches Landesamt Baden-Württemberg.

Ashraf, Q., Özak, Ö, & Galor, O. (2010). Isolation and development. *Journal of the European Economic Association*, 8(2-3), 401–412.

Austin, B., Glaeser, E., & Summers, L. (2018). Jobs for the heartland: place-based policies in 21st-century America. *Brookings Papers on Economic Activity*, 2018(1), 151–255.

Autor, D. H., Dorn, D., & Hanson, G. H. (2016). The China shock: learning from labor market adjustment to large changes in trade. *Annual Review of Economics*, 8(1), 205–240.

Bairoch, P. (1990). The impact of crop yields, agricultural productivity, and transport costs on urban growth between 1800 and 1910. In A. van der Woude, A. Hayami, & J. de Vries (Eds.), *Urbanization in history. A process of dynamic interactions*, Urbanization in History (pp. 134–151). Oxford University Press.

Barrios, J. M., Hochberg, Y., & Macciocchi, D. (2021). Rugged entrepreneurs: the geographic and cultural contours of new business formation. *NBER Working Papers* 28606, National Bureau of Economic Research.

Bartels, C., Jäger, S., & Obergruber, N. (2024). Long-term effects of equal sharing: evidence from inheritance rules for land. *The Economic Journal*, forthcoming.

Becker, S. O., Mergele, L., & Woessmann, L. (2020). The separation and reunification of Germany: rethinking a natural experiment interpretation of the enduring effects of Communism. *Journal of Economic Perspectives*, 34(2), 143–171.

Bergeaud, A. & Verluise, C.. (2024). A new dataset to study a century of innovation in Europe and in the US. *Research Policy*, 53(1): 104903.

Braun, S. T. & Franke, R. (2022). Railways, growth, and industrialization in a developing German economy, 1829-1910. *The Journal of Economic History*, 82(4), 1183–1221.

Braun, S. T. & Kvasnicka, M. (2014). Immigration and structural change: evidence from post-war Germany. *Journal of International Economics*, 93(2), 253–269.

Broadberry, S. & Crafts, N. (2003). Productivity performance from 1950 to 1979: a restatement of the Broadberry-Crafts view. *The Economic History Review*, 56(4), 718–735.

Cantoni, D., Mohr, C., & Weigand, M. (2020). Princes and townspeople: a collection of historical statistics on German territories and cities. 3: Town charters and first mentions.

Celli, V. (2022). Causal mediation analysis in economics: objectives, assumptions, models. *Journal of Economic Surveys*, 36(1), 214–234.

Chinitz, B. (1961). Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review: Papers and Proceedings*, 51(2), 279–289.

Ciccone, A. & Nimczik, J (2022). The long-run effects of immigration: evidence across a barrier to refugee settlement. *IZA DP No. 15185.*

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1–45.

Criscuolo, C., Martin, R. Overman, H. G., & Van Reenen, J. (1999). Some causal effects of an industrial policy. *American Economic Review*, 109(1), 48–85.

Daudin, G. (2010). Domestic trade and market size in late-eighteenth-century France. *The Journal of Economic History*, 70(3), 716–743.

Dauth, W., Findeisen, S., & Suedekum, J.(2017). Trade and manufacturing jobs in Germany. *American Economic Review: Papers and Proceedings*, 107(5), 337–342.

de Pleijt, A., Nuvolari, A., & Weisdorf, J. (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.

Dippel, C., Gold, R., Heblich, S., & Pinto, R. (2020a). Mediation analysis in IV settings with a single instrument. *Unpublished manuscript.*

Dippel, C., Ferrara, A., & Heblich, S. (2020b). Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, 20(3), 613–626.

Dippel, C., Gold, R., Heblich, S., & Pinto, R. (2021). The effect of trade on workers and voters. *The Economic Journal*, 132(641), 199–217.

Esposito, E. & Abramson, S. F. (2021). The European coal curse. *Journal of Economic Growth*, 26(1), 77–112.

Felice, E. (2018). Regional income inequality in Italy in the long run (1871–2010): patterns and determinants. In *The economic development of Europe's regions* (pp. 177–203). Routledge.

Fenoaltea, S. (2003). Peeking backward: regional aspects of industrial growth in post-unification Italy. *The Journal of Economic History*, 63(4), 1059–1102.

Fernihough, A. & O'Rourke, K. H. (2020). Coal and the European Industrial Revolution. *The Economic Journal*, 131(635), 1135–1149.

Floerkemeier, H., Spatafora, N., & Venables, A. J. (2021). Regional disparities, growth, and inclusiveness. *IMF Working Papers* 2021/038, International Monetary Fund.

Franck, R. & Galor, O. (2021). Flowers of evil? Industrialization and long run development. *Journal of Monetary Economics*, 117, 108–128.

Frank, H. (1993). *Regionale Entwicklungsdisparitäten im deutschen Industrialisierungsprozeß 1849–1939. Eine empirisch-analytische Untersuchung.* PhD thesis, Westfälische Wilhelms-Universität Münster.

Fremdling, R. (1977). Railroads and German economic growth: a leading sector analysis with a comparison to the United States and Great Britain. *The Journal of Economic History*, 37(3), 583–604.

Fremdling, R. (1979). Modernisierung und Wachstum der Schwerindustrie in Deutschland, 1830-1860. *Geschichte und Gesellschaft*, 5(2), 201–227.

Fritsch, M. & Wyrwich, M. (2014). The long persistence of regional levels of entrepreneurship: Germany, 1925–2005. *Regional Studies*, 48(6), 955–973.

Fritzsche, M. & Wolf, N. (2023). Fickle fossils. Economic growth, coal and the European oil invasion, 1900-2015. *CESifo Working Papers* 10805.

Frölich, M. & Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with Instrumental Variables. *Journal of the Royal Statistical Society Series B*, 79(5), 1645–1666.

Gagliardi, L., Moretti, E., & Serafinelli, M. (2023). The world's rust belts: the heterogeneous effects of deindustrialization on 1,993 cities in six countries. *NBER Working Papers* 31948, National Bureau of Economic Research.

Gaubert, C., Kline, P., Vergara, D., & Yagan, D. (2021). Trends in US spatial inequality: concentrating affluence and a democratization of poverty. *AEA Papers and Proceedings*, 111, 520–525.

Gehrsitz, M. & Ungerer, M. (2022). Jobs, crime and votes: a short-run evaluation of the refugee crisis in Germany. *Economica*, 89, 592–626.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: an empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2), 498–520.

Glaeser, E. L. & Kerr, W. R. (2009). Industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*, 18, 623–663.

Grabher, G. (1993). The weakness of strong ties: the lock-in of regional development in the Ruhr area. In *The embedded firm: on the socioeconomics of industrial networks* (pp. 255–277). Routledge.

Gutberlet, T. (2014). Mechanization and the spatial distribution of industries in the German Empire, 1875 to 1907. *The Economic History Review*, 67(2), 463–491.

Halbmeier, C. & Schröder C. (2024). The long-term implications of destruction during the Second World War on private wealth in Germany. *Journal of Economic Growth*, forthcoming.

Hamm, R. & Wienert, H. (1990). *Strukturelle Anpassung altindustrieller Regionen im internationalen Vergleich.* Duncker & Humblot.

Hassink, R. & Kiese, M. (2021). Solving the restructuring problems of (former) old industrial regions with smart specialization? Conceptual thoughts and evidence from the Ruhr. *Review of Regional Research*, 41(2), 131–155.

Herrigel, G. (1999). Governance of small and medium sized firms manufacturing in Germany: transformation of the machine tool industry in the 20th century. *Entreprises et histoire*, 211, 59–69.

Herrigel, G. (2000). *Industrial constructions: the sources of German industrial power*. Structural analysis in the social sciences 9. Cambridge University Press.

Hu, X. & Hassink, R. (2016). Explaining differences in the adaptability of old industrial areas. In *Routledge handbook of politics and technology* (pp. 162–172). Routledge.

Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2018). Regional inequality in Europe: evidence, theory and policy implications. *Journal of Economic Geography*, 19(2), 273–298.

Imai, K., Keele, L.,& Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 51–71.

Institut für Weltwirtschaft (1974). *Infrastruktur, räumliche Verdichtung und sektorale Wirtschaftsstruktur als Bestimmungsgründe des regionalen Entwicklungspotentials in den Arbeitsmarktregionen der Bundesrepublik Deutschland.* IfW Kiel.

Junkernheinrich, M. (1989). Ökonomische Erneuerung alter Industrieregionen: Das Beispiel Ruhrgebiet. *Wirtschaftsdienst*, 69(1), 28–35.

Kaelble, H. & Hohls, R. (1989). Der Wandel der regionalen Disparitäten in der Erwerbsstruktur Deutschlands 1895–1970. In *Schriften des Zentralinstituts für sozialwissenschaftliche Forschung der Freien Universität Berlin* (pp. 288–413). VS Verlag für Sozialwissenschaften.

Kaiserliches Statistisches Amt (1884). Berufsstatistik nach der allgemeinen Berufszählung vom 5. Juni 1882. 1. Berufsstatistik des Reichs und der kleineren Verwaltungsbezirke. *Statistik des Deutschen Reichs. Neue Folge*, 2.

Kaiserliches Statistisches Amt (1909a). Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VI. *Statistik des Deutschen Reichs. Neue Folge*, 218.

Kaiserliches Statistisches Amt (1909b). Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VII. *Statistik des Deutschen Reichs. Neue Folge*, 219.

Keller, W. & Shiue, C. H. (2014). Endogenous formation of free trade agreements: evidence from the Zollverein's impact on market integration. *The Journal of Economic History*, 74(4), 1168–1204.

Kiese, M. (2019). Strukturwandel 2.0: Das Ruhrgebiet auf dem Weg zur Wissensökonomie? *Standort*, 43(2), 69–75.

Kiesewetter, H. (1986). Das wirtschaftliche Gefälle zwischen Nord- und Süddeutschland in historischer Perspektive. *Neues Archiv für Niedersachsen: Zeitschrift für Stadt-, Regional- und Landesentwicklung*, 35(4), 327–347.

Kiesewetter, H. (2004). *Industrielle Revolution in Deutschland. Regionen als Wachstumsmotoren.* Franz Steiner Verlag.

Kitson, M. & Michie, J. (1996). Britain's industrial performance since 1960: underinvestment and relative decline. *The Economic Journal*, 106(434), 196–212.

Kriegesmann, B., Böttcher, M., & Lippmann, T. (2016). Die regionaloökonomische Bedeutung der Wissenschaft für das Ruhrgebiet. *Standort*, 40, 177–183.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1), 1–28.

Matheis, M. (2016). Local economic impacts of coal mining in the United States 1870 to 1970. *The Journal of Economic History*, 76(4), 1152–1181.

Nonn, C. (2001). *Die Ruhrbergbaukrise. Entindustrialisierung und Politik 1958-1969.* Vandenhoeck & Ruprecht.

Persson, J. (1997). Convergence across the Swedish counties, 1911–1993. *European Economic Review*, 41(9), 1835–1852.

Peters, M. (2022). Market size and spatial growth—evidence from Germany's post-war population expulsions. *Econometrica*, 90(5), 2357–2396.

Polish Army Topographical Service (1968). *Pergamon World Atlas.* Pergamon Press.

Redding, S. J. & Sturm, D. M. (2008). The costs of remoteness: evidence from German division and reunification. *American Economic Review*, 98(5), 1766–1797.

Rice, P. G. & Venables, A. J. (2021). The persistent consequences of adverse shocks: how the 1970s shaped UK regional inequality. *Oxford Review of Economic Policy*, 37(1), 132–151.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, 11(1), 189–209.

Rodríguez-Pose, A., Terrero-Dávila, J., & Lee, N. (2023). Left-behind versus unequal places: interpersonal inequality, economic decline and the rise of populism in the USA and Europe. *Journal of Economic Geography*, 23(5), 951–977.

Rosés, J. R. & Wolf, N. (Eds.). (2019a). *The economic development of Europe's regions.* Routledge.

Rosés, J. R. & Wolf, N. (2019b). The return of regional inequality: Europe from 1900 to today. VOX, CEPR Policy Portal.

Rosés, J. R. & Wolf, N. (2021). Regional growth and inequality in the long-run: Europe, 1900–2015. *Oxford Review of Economic Policy*, 37(1), 17–48.

Sala-i Martin, X. X. (1996). The classical approach to convergence analysis. *The Economic Journal*, 106(437), 1019–1036.

Schrader, K. & Laaser, C.-F. (2019). Unterschiede in der Wirtschaftsentwicklung im Norden und Süden Deutschlands. *Kieler Beiträge zur Wirtschaftspolitik* 20, IfW Kiel.

Simon, H. (1992). Lessons from Germany's midsize giants. *Harvard Business Review*, 70(2), 115–121.

Simon, H. (2009). *Hidden champions of the twenty-first century. The success strategies of unknown world market leaders.* Springer.

Statistisches Bundesamt (2023). *Bildungsfinanzbericht 2023.* Statistisches Bundesamt, Wiesbaden.

Streeck, W. (1991). On the institutional conditions for diversified quality production. In *Beyond Keynesianism* (pp. 21–61). Elgar.

Stuetzer, M., Obschonka, M., Audretsch, D. B., Wyrwich, M., Rentfrow, P. J., Coombes, M., Shaw-Taylor, L., & Satchell, M. (2016). Industry structure, entrepreneurship, and culture: an empirical analysis using historical coalfields. *European Economic Review*, 86, 52–72.

Tilly, R. & Kopsidis, M. (2020). *From old regime to industrial state: a history of German industrialization from the eighteenth century to World War I.* The University of Chicago Press.

Van der Ploeg, F. (2011). Natural resources: curse or blessing? *Journal of Economic Literature*, 49(2), 366–420.

Venables, A. J. (2016). Using natural resources for development: why has it proven so difficult? *Journal of Economic Perspectives*, 30(1), 161–184.

Vonyó, T. (2012). The bombing of Germany: the economic geography of war-induced dislocation in West German industry. *European Review of Economic History*, 16(1), 97–118.

Williamson, J. G. (1965). Regional inequality and the process of national development: a description of the patterns. *Economic Development and Cultural Change*, 13(4), 1–84.

Wolf, N. (2019). Regional economic growth in Germany, 1895–2010. In J. R. Rosés & N. Wolf (Eds.), *The economic development of Europe's regions* (pp. 149–176). Routledge.

Young, A. T., Higgins, M. J., & Levy, D. (2008). Sigma convergence versus beta convergence: Evidence from US county-level data. *Journal of Money, Credit and Banking*, 40(5), 1083–1093.

Ziegler, D. (2012). *Die Industrielle Revolution* (3rd ed.). Wissenschaftliche Buchgesellschaft.

# Online Appendix

## Reversing Fortunes of German Regions, 1926–2019: Boon and Bane of Early Industrialization?

Paul Berbée *  Sebastian Till Braun †  Richard Franke ‡

*ZEW–Leibniz Centre for European Economic Research, Mannheim, Germany

†Corresponding author. University of Bayreuth and IZA. Contact: University of Bayreuth, Faculty of Law, Business and Economics, 95440 Bayreuth, Germany. Email: sebastian.braun@uni-bayreuth.de. Tel.: +49 (0) 921 55 6256.

‡Trinity College Dublin and CEPH, Ireland

# A Online Appendix

## A.1 Data description and sources

This section provides a detailed description of the data set, the sources, and the construction of variables.

**Outcomes:**

*GDP per capita 1957-2019:* West Germany's federal statistical office began publishing disaggregated GDP per capita data at the county level in 1957. We digitized the data for 1957-1992 from printed sources (Statistische Landesämter, 1968, 1973, 1978, 1979, 1998). Data are available for eleven years in this period, namely for 1957, 1961, 1964, 1966, 1968, 1970, 1972, 1974, 1978, 1980 and 1992. GDP data for 1992-2019 are available at an annual frequency in electronic form (Arbeitskreis VGR der Länder, 2021). If GDP for a given year is available from two publications, we usually use the more recent GDP revision. As an exception, we take the 1992 data from Statistische Landesämter (1998) in order to have data for the last year before German unification (1980) and the first year after unification (1992) from the same source.

*Turnover 1926, 1935, 1950, and 1955:* We proxy regional GDP before 1957 by firm sales at the county level, as published in turnover tax statistics (Statistisches Reichsamt, 1931, 1939; Statistisches Bundesamt, 1955, 1957a).

**Explanatory variables:**

*Industrial employment share:* The share of the local workforce working in industrial occupations in 1882 based on data from the first German-wide occupation census that contains results at the county level (Kaiserliches Statistisches Amt, 1884). The industrial employment share 1882 is defined as industrial employment over total employment. The industrial employment shares in 1895 and 1907, used as robustness checks, come from Kaiserliches Statistisches Amt (1897) and Kaiserliches Statistisches Amt (1910), respectively.

*Distance to European coal fields:* We use the weighted least-cost distance to European coal fields as an instrument for the 1882 employment share in industrial occupations. To calculate the instrumental variable, we first divide Europe in a one-by-one kilometer grid, using an Equidistant Conic projection of Europe (ESRI:102031). Based on the local geography, we assign each cell a specific pre-industrial transportation cost, which we take from Daudin (2010). Daudin (2010) estimates a mode-specific transport cost vector for France in 1790. By using a pre-industrial cost vector, estimated before the first railroad was built in France, we avoid that our instrumental variable captures the fact that early industrialized regions are, for endogenous reasons, better connected to the transport network. For the same reason, we do not take into account the road network or canals, but only first nature geography, i.e., land, rivers, sea. In the absence of comparable cost vectors for Germany, we use the estimates for France, since both countries have a similar topography. We normalize the cost to one for cells that have access to the sea. Cells with access to a major river are assigned a cost value of 1.018, all other cells are assigned Daudin's value for road transport of 2.963. We take shapefiles of major European rivers (*ne_10m_rivers_lake_centerlines*) and landmass (*ne_50m_admin_0_countries*) from Fernihough & O'Rourke (2020). The shape files are *Made with Natural Earth.* The value for rivers is the average of upstream and downstream river transport. We also probe the robustness of our results to alternative costs vectors. In particular, we use squared transport costs and assign higher costs of 2.476 to river and 9.75 to road transport, respectively, following Bairoch (1990). We also restrict the set of rivers to those that are at least 20 meters wide and two meters deep in an additional robustness check. In doing so, we use the average river bankfull width and depth from the database of Konstantinos, Schumann & Pavelsky (2013). We then calculate the least-cost distance from each labor market's centroid to all European coal field centroids, using Dijkstra's algorithm and the grid as cost surface. The algorithm is implemented in the `R` package `gdistance` (van Etten, 2017) and finds the least-cost path from a labor market to a coal field, adding cell-specific costs along the way. The instrument for a given labor market $i$, $C_i$, is the (log) sum of the least-cost distances to all coal fields, using the area of coal fields as weights:

$$C_i = \log\left(\sum_{k=1}^{K} \frac{area_k}{cost_{ik}}\right), \tag{A-1}$$

where $cost_{ik}$ is the least cumulative costs from labor market $i$ to coal field $k$, and $area_k$ is the area of the coal field polygon in square kilometers. We take the location and extent of European coal fields from Fernihough & O'Rourke (2020).

**Control variables:**

*Land access:* Land access is measured as the sum of least-cost distances to all European one-by-one kilometer grid cells on land, using the same cost vector as in equation (A-1).

*Towns 1700:* Is defined as the number of towns in 1700 in a given labor market, normalized by area in square kilometers (Cantoni, Mohr & Weigand, 2020).

*Carboniferous strata:* The information on carboniferous strata are taken from Fernihough & O'Rourke (2020). The underlying geological map was compiled by Asch (2003) and reports the surface geology of Europe and adjacent areas. Therefore, coal-bearing (carboniferous) layers that are covered by other strata are not included in the map.

*Distance to coast and rivers:* Is defined as the great circle distance between a labor market's centroid and the nearest coastline and major river in kilometers, respectively. Shapefiels for coastlines and major rivers are taken from Fernihough & O'Rourke (2020) and are *Made with Natural Earth.*

*Distance to inner-German border, eastern border, Schengen border, western border*: Are defined as the great circle distance between the center of a labor market and the respective border in kilometers. The western border refers to the border with France and the Benelux countries. The eastern border refers to the CSSR and the GDR. The Schengen area refers to the year 2008.

*Lost market access:* We calculate lost firm market access (FMA) based on Redding & Sturm (2008). To do this, we use county-level population data from the 1939 census and calibrate the model so that the 1939 population distribution is in equilibrium (Statistisches

Reichsamt, 1941). For West Germany, we aggregate the data to match our 163 labor markets prior to calibration. After calibration, we calculate the lost FMA for each labor market as the difference between the FMA in 1939 equilibrium and the FMA after division, i.e., we exclude all East German counties from trade.

*Population in 1882:* We use county-level population data from the 1882 occupation census (Kaiserliches Statistisches Amt, 1884). In contrast to the 1880 population census, the occupation census does not measure the population at the place of residence (*Wohnbevölkerung*) or whereabouts during the census (*ortsanwesende Bevölkerung*), but at the place of occupation of the provider (*Berufs-Bevölkerung*).

*Population share of GDR refugees in 1961:* Is defined as the number of GDR refugees over the total population in 1961 (Schmitt, Rattinger & Oberndörfer, 1994).

*Population share of expellees in 1961:* Is defined as the number of expellees (*Heimatvertriebene*) over the total population in 1961 (Schmitt et al., 1994).

*Population share of foreigners in 1970 and 1987:* Is defined as the number of expellees (*Heimatvertriebene*) over the total population in 1970 or 1987 (Schmitt et al., 1994).

*Population share of refugees in 2016:* Is defined as the number of foreign nationals seeking protection in Germany on international, humanitarian or political grounds (*Schutzsuchende*) as a percentage of the total population in 2016. The data come from the German Federal Statistical Office.

Redevelopment regions (*Sanierungsgebiete*): This indicator variable takes the value of one if a region was classified as a redevelopment area in 1954 (Isbary, Von der Heide & Müller, 1969). This applies to all regions in the zonal border area (*Zonenrandgebiet*), a 40-kilometer wide strip along the border to the GDR and the CSSR.

*Remoteness within Germany:* Following Keller & Shiue (2014), we define remoteness within Germany as the deviation of a labor market's latitude and longitude from the average latitude and longitude in our sample.

*Rubble per capita 1946:* Untreated rubble at the end of the war over the population in 1939 (Deutscher Städtetag, 1949). The data are only available for the 199 largest West German cities. We aggregate the data to the labor market level, implicitly assuming that war destruction is zero in smaller municipalities.

*Ruhr region:* Is a binary indicator that is one for the labor markets of Duisburg, Essen, Recklinghausen, Bochum, Dortmund, and Hamm-Beckum.

*Share of damaged dwellings:* Is defined as the share of dwellings that were built before 1945 and were damaged in the war (Statistisches Bundesamt, 1956).

*Soil quality:* We calculate the average soil quality of farmland cells within each labor market based on a 250-by-250 meter raster data set from BGR (2014).[1]

[1] For the labor market of *Lindau*, we calculate the average soil quality of the neighboring labor markets because the raster data set does not indicate a farmland cell within *Lindau*.

*State fixed effects:* Indicate to which federal state (*Bundesländer* as of 1952 or state in pre-industrial Germany as of 1834) a labor market belongs. If a labor market belongs to more than one state, we choose the state with the largest area overlap. For 1834 we use shapefiles from HGIS Germany (2007).

*Sunshine hours (mean 1991-2020):* We calculate the average annual sunshine in hours within each labor market based on an one-by-one kilometer raster data set of annual average sunshine hours 1991–2020 from Deutscher Wetterdienst (DWD) (2021).

*Zollverein 1834:* Indicator variable for whether a labor market belongs to the German Customs Union (*Zollverein*) as of 1834. For 1834 we use shapefiles from HGIS Germany (2007).

**Channels:**

*Share of people with a university degree 1970, 2011:* Share of people with a university degree in the total population aged 15 and over. Data on total population aged 15 and over and population by education are from the respective population censuses. Earlier censuses did not record educational attainment. Data for 1970 and 2011 are from Schmitt et al. (1994) and `https://www.zensus2011.de`, respectively.

*Share of people with higher secondary or vocational degree in 1970, 2011:* Share of people with more than the minimum compulsory education (*Volks-* or *Hauptschulabschluss*) in the total population aged 15 and over. This includes people with a medium or higher secondary schooling degree (Mittlere Reife or Hochschulreife) and people with a vocational or university degree. Data on the total population aged 15 and over and the population by education are from the respective censuses. Earlier censuses did not record educational attainment. Data for 1970 and 2011 are from Schmitt et al. (1994) and `https://www.zensus2011.de`, respectively.

*Share of school dropouts 1970, 2011:* Share of people without a school-leaving qualification (*ohne Schulabschluss*) in the total population aged 15 and over. Data on the total population aged 15 and over and the population by education are from the respective censuses. In the 1970 census, we calculate the number of school drop-outs as the difference between the total population and the population with the different degrees recorded in the census. The 2011 census directly reports the number of people without a school-leaving qualification. Earlier censuses did not record educational attainment. Data for 1970 and 2011 are from Schmitt et al. (1994) and `https://www.zensus2011.de`, respectively.

*Population share of students 2019, by field:* Share of students enrolled in a university or university of the applied science in the total population. Data, broken down by subject, come from the German Federal Statistical Office.

*Number of universities and colleges, 1900-2010:* Number of universities, technical universities and universities of the applied sciences (*Fachhochschulen*) with and without the right to grant doctorates. We take the list of universities from Berlingieri, Gathmann & Quinckhardt (2022) (whom we thank for sharing their data with us). In total, their list contains data on 300 universities and their founding year. We define large universities as those with at least 7,500 students today.

*Industrial employment share and employment share in services 1882, 1907, 1939, 1950, 1961, 1970, 1987, 1994-2019*. Employment shares for the years 1882, 1907, and 1939 are based on the respective occupation censuses (Kaiserliches Statistisches Amt, 1884, 1910; Braun & Franke, 2021). Employment shares for 1950, 1961, 1970, and 1987 come from population censuses and are taken from Schmitt et al. (1994). Since 1992, sectoral employment shares are reported as part of official GDP estimates (Arbeitskreis VGR der Länder, 2021).

*Average firm size in industry in 1907:* Is defined as the average number of persons employed in an establishment (*Hauptbetrieb*) in industry (Kaiserliches Statistisches Amt, 1909a,b).

*Employment share in industries dominated by large firms in 1907:* Is defined as total employment in industries dominated by large firms over all employed persons (Kaiserliches Statistisches Amt, 1909a,b). We define industries dominated by large firms as those with a Germany-wide employment share of at least 50% in firms with at least 501 or 1000 employees. See Appendix Table A-2 for a list of these sectors.

*Sectoral concentration of industrial employment in 1907:* We measure employment concentration by the Hirschman-Herfindahl-Index (with $\alpha = 2$). The index is calculated as $HHI_i = \sum_{l=1}^{L}(b_{il})^2$, where $b_{il}$ is the employment share of labor market $i$ in (3-digit) industry $l$ in total industrial employment. The HHI ranges from $1/L$ (when all sectors have equal employment) to 1 (when all employment is concentrated in one sector). We use data for $L = 300$ sectors from Kaiserliches Statistisches Amt (1909a,b).

*Sectoral concentration of industrial employment in 1950:* We again measure employment concentration by the Hirschman-Herfindahl-Index (see above). The two-digit employment data are based on the 1950 occupation census and are taken from several official publications (Bayerisches Statistisches Landesamt, 1952; Hessisches Statistisches Landesamt, 1952; Niedersächsisches Amt für Landesplanung und Statistik, 1953; Statistisches Landesamt Baden-Württemberg, 1954; Statistisches Landesamt Bremen, 1953; Statistisches Landesamt der Freien und Hansestadt Hamburg, 1953; Statistisches Landesamt Nordrhein-Westfalen, 1952a,b; Statistisches Landesamt Rheinland-Pfalz, 1952; Statistisches Landesamt Schleswig-Holstein, 1953).

*Self-employment share in 1950:* Defined as the number of self-employed over total employment in 1950. The data are based on the 1950 occupation census and are taken from Braun & Franke (2021).

*Number of years the major was member of the Social Democrats in 1950-1990 and number of years the major was member of the locally dominant party in 1950-1990:* We hand-collected names, years in office, and party affiliation of all mayors since 1945 of the main town in each regional labor market. The sources are Wikipedia entries, towns' websites and towns' archives. For each town, we define the dominant party as the party with most years in office (with independent mayors counting for no party).

*Vote share of the Socials Democrats in the national election of 1957:* Data on election outcomes by counties are available from Statistisches Bundesamt (1957b).

*Patents 1877-2013:* Geocoded utility patents are from Bergeaud & Verluise (2024). For the analysis, we consider only patents published by the German patent offices.

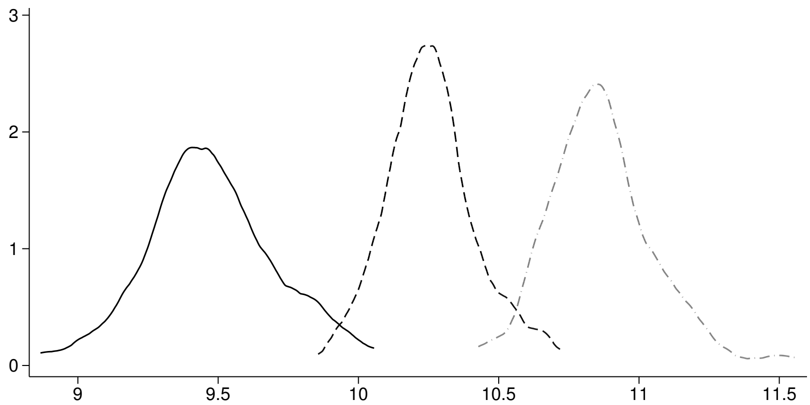
**Unit of analysis:**

*Definition:* Our unit of analysis is the 163 West German labor markets defined in Institut für Weltwirtschaft (IfW) (1974). The labor markets in IfW (1974) combine counties based on commuter flows. The classification refers to counties in their 1966 borders.

*Data aggregation:* We aggregate our source data, collected at the level of counties, to the level of labor markets using Geographical Information System (GIS) software. The definition of local labor markets is based on county boundaries in 1966. For other years, we overlay maps of historical county boundaries with the base map of local labor markets. We then use the proportion of each historical county's area that belongs to a particular local labor market to aggregate the county-level data. Shapefiles on county boundaries are taken from Max Planck Institute for Demographic Research (MPIDR) and Chair for Geodesy and Geoinformatics, University of Rostock (CGG) (2011).

**GDP per capita, 1957, 1980, 2019.** Figure A-1 shows kernel density estimates of the distribution of log GDP per capita (DM, in 1992 prices) in 1957, 1980, and 2019, and Table A-1 presents the corresponding summary statistics. As discussed in Section 5, the distribution of GDP per capita became less unequal between 1957 and 1980. This is reflected, for example, in a decreasing standard deviation or a decreasing interquartile range (i.e., a decreasing difference between the 75th and 25th percentiles). Consequently, the position of a labor market in the distribution, our main dependent variable of interest, tended to become less consequential for the labor market's relative economic power between 1957 and 1980. In contrast, the distribution of (log) GDP per capita became more unequal again between 1980 and 2019.

Figure A-1: Distribution of log GDP per capita



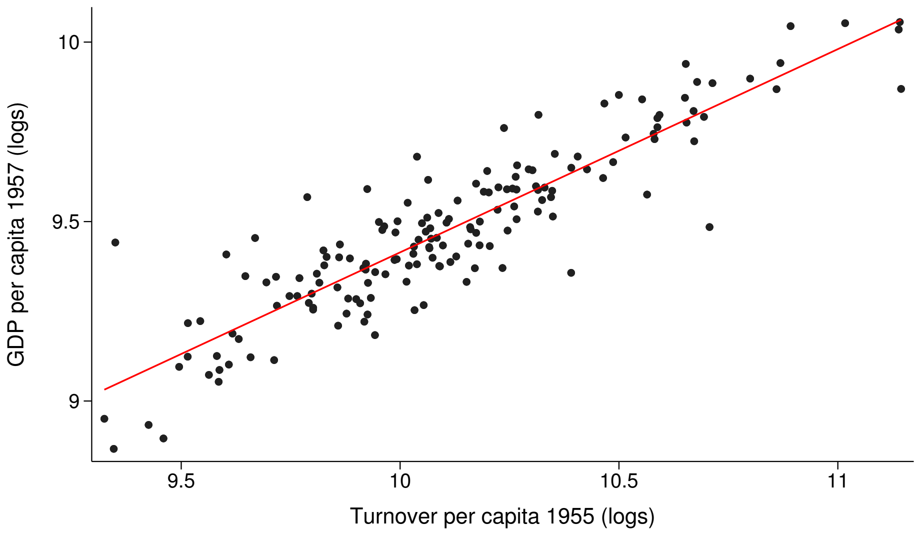

Notes: The plot shows kernel density estimates of the distribution of log GDP per capita (DM, in 1992 prices) in 1957, 1980, and 2019.

Table A-1: Summary statistics for log GDP per capita

| | 1957<br>(1) | 1980<br>(2) | 2019<br>(3) |
|---|---|---|---|
| Mean | 9.477 | 10.259 | 10.865 |
| Median | 9.455 | 10.252 | 10.852 |
| 25th Percentile | 9.332 | 10.152 | 10.729 |
| 75th Percentile | 9.606 | 10.347 | 10.964 |
| Standard deviation | 0.234 | 0.162 | 0.191 |
| Skewness | 0.150 | 0.396 | 0.665 |
| Min | 8.867 | 9.856 | 10.428 |
| Max | 10.056 | 10.734 | 11.559 |

Notes: The table shows summary statistics for log GDP per capita (DM, in 1992 prices) in 1957, 1980, and 2019.

Figure A-2: Per capita GDP 1957 vs. per capita sales 1955 (logs)




Notes: The figure plots per capita GDP in 1957 against per capita sales in 1955, along with the linear regression line in red. Each dot represents a labor market.

**Industries dominated by large firms.** Kaiserliches Statistisches Amt (1907) contains for all three-digit industries the number of persons employed in establishments with 1, 2, 3, 4-5, 6-10, 11-20, 21-50, 51-100, 101-200, 201-500, 501-1000, and more than 1000 employees in 1907 (similar statistics do not exist at the county level). Using the list, Table A-2 below identifies industries in which at least 50% of all employees work in establishment with at least 501 (Panel A) or 1000 employees (Panel B).

Table A-2: List of sectors dominated by large firms in 1907

| Industry | Employment share in large firms (1) | Total employment (2) |
|---|---|---|
| **A. At least 50% of workers employed in firms with at least 501 employees** | | |
| Shipbuilding | .502 | 49842 |
| Tramway operation | .530 | 48531 |
| Underground and submarine cables | .531 | 10186 |
| Coal tar derivatives | .559 | 6825 |
| Rubber products | .578 | 28383 |
| Pencils | .581 | 3477 |
| Jute spinning | .560 | 12868 |
| Earthenware fabrication and finishing | .601 | 19514 |
| Other electrical equipment and auxiliary products | .602 | 42001 |
| Sea and coastal shipping | .613 | 60697 |
| Iron and steel production | .639 | 170614 |
| Ore mines (excluding iron) | .681 | 43906 |
| Electrical telegraphs and telephone systems | .684 | 7830 |
| Other firearms | .689 | 10797 |
| Sewing machines | .707 | 18448 |
| Steam engines | .716 | 69513 |
| Aniline, aniline dyes | .792 | 9071 |
| Coal briquettes | .794 | 8408 |
| Electric generating machinery and engines | .806 | 27703 |
| Weapon foundries | .917 | 7551 |
| Stone coal mines | .946 | 452866 |
| **B. At least 50% of workers employed in firms with at least 1000 employees** | | |
| Coal briquettes | .523 | 8408 |
| Sea and coastal shipping | .533 | 60697 |
| Other firearms | .558 | 10797 |
| Electrical telegraphs and telephone systems | .615 | 7830 |
| Aniline, aniline dyes | .719 | 9071 |
| Stone coal mines | .801 | 452866 |
| Electric generating machinery and engines | .806 | 27703 |
| Weapon foundries | .829 | 7551 |

Notes: The table list the sectors with a Germany-wide share of workers of at least 50% in very large firms in 1907. Large firms are defined as having at least 501 (Panel A) or 1000 (Panel B) employees. Column (1) gives the corresponding employment share in large firms and Column (2) the Germany-wide employment.

## A.2 Early industrialization and economic development, 1926-2019

Table A-3: Robustness checks of 2SLS estimates: Additional controls

| | 1926-2019 (1) | 1957-2019 (2) | 1926 (3) | 1957 (4) | 2019 (5) |
|---|---|---|---|---|---|
| **A. Baseline specification** | | | | | |
| Baseline specification | -30.04*** | -33.62*** | 12.22*** | 15.80*** | -17.82*** |
| | (7.52) | (7.24) | (3.25) | (3.14) | (6.89) |
| | [5.75] | [6.18] | [2.96] | [3.14] | [6.03] |
| **B. Contemporaneous shocks** | | | | | |
| B1. German division and re-unification | | | | | |
| Adding distance to inner-German border (logs) | -30.86*** | -33.78*** | 12.21*** | 15.12*** | -18.66*** |
| | (7.38) | (7.22) | (3.38) | (2.95) | (6.55) |
| | [5.72] | [6.28] | [3.07] | [2.91] | [5.76] |
| Adding market access lost in 1945 | -30.63*** | -33.54*** | 12.07*** | 14.98*** | -18.56*** |
| | (7.53) | (7.23) | (3.31) | (2.97) | (6.77) |
| | [5.84] | [6.27] | [2.97] | [2.91] | [5.98] |
| Adding distances to eastern border (logs) | -30.72*** | -33.20*** | 11.34*** | 13.83*** | -19.37*** |
| | (7.68) | (7.28) | (3.15) | (2.71) | (6.91) |
| | [6.06] | [6.44] | [2.71] | [2.55] | [6.19] |
| Adding dummy for redevelopment areas | -32.52*** | -34.88*** | 11.82*** | 14.18*** | -20.70*** |
| | (7.48) | (7.22) | (3.61) | (3.04) | (6.39) |
| | [5.91] | [6.42] | [3.28] | [2.82] | [5.84] |
| B2. Immigration | | | | | |
| Adding expellee share in 1950 | -34.39*** | -35.03*** | 12.03*** | 12.66*** | -22.37** |
| | (9.77) | (8.34) | (3.73) | (3.44) | (8.75) |
| | [8.16] | [7.83] | [3.15] | [3.28] | [8.03] |
| Adding share of GDR refugees in 1961 | -29.50*** | -32.91*** | 8.40** | 11.82*** | -21.09*** |
| | (8.08) | (7.74) | (3.33) | (3.72) | (7.72) |
| | [6.59] | [7.00] | [3.42] | [3.99] | [6.59] |
| Adding share of foreigners in 1970 | -28.36*** | -32.96*** | 14.19*** | 18.79*** | -14.17*** |
| | (6.47) | (6.65) | (2.88) | (2.88) | (4.59) |
| | [4.51] | [5.39] | [2.48] | [3.04] | [3.20] |
| Adding share of foreigners in 1987 | -31.65*** | -34.47*** | 10.36*** | 13.18*** | -21.29*** |
| | (7.20) | (7.21) | (3.34) | (2.91) | (5.25) |
| | [5.33] | [6.20] | [2.94] | [2.77] | [4.23] |
| Adding share of refugees in 2016 | -29.96*** | -34.61*** | 4.53 | 9.18* | -25.42** |
| | (9.54) | (8.72) | (5.55) | (5.57) | (10.33) |
| | [6.87] | [7.05] | [5.90] | [5.42] | [8.90] |
| B3. Bombing damage | | | | | |
| Adding rubble per capita in 1946 | -34.44*** | -37.15*** | 6.68 | 9.40** | -27.75*** |
| | (9.91) | (8.67) | (4.23) | (4.70) | (9.94) |
| | [8.37] | [8.20] | [4.53] | [4.53] | [8.92] |
| Adding share of housing stock damaged in WWII | -28.81*** | -30.25*** | 4.94 | 6.38 | -23.87** |
| | (9.78) | (8.22) | (5.51) | (6.69) | (11.90) |
| | [9.49] | [10.01] | [5.38] | [6.17] | [12.56] |
| B4. European integration | | | | | |
| Adding distance to western border (logs) | -23.44*** | -25.40*** | 9.19*** | 11.15*** | -14.26** |
| | (7.10) | (5.19) | (2.93) | (3.51) | (6.92) |
| | [6.06] | [5.51] | [2.78] | [3.36] | [6.94] |
| Adding distance to Schengen area border 2008 (logs) | -30.05*** | -33.44*** | 12.05*** | 15.43*** | -18.01** |
| | (7.68) | (7.31) | (3.31) | (3.01) | (7.00) |
| | [5.99] | [6.39] | [3.07] | [3.08] | [6.15] |
| **C. Geographic controls** | | | | | |
| Adding location at coast (0/1) | -30.32*** | -33.79*** | 12.28*** | 15.75*** | -18.04*** |
| | (7.34) | (7.13) | (3.20) | (3.17) | (6.79) |
| | [5.82] | [6.18] | [2.96] | [3.15] | [6.12] |
| Adding soil quality | -31.29*** | -34.83*** | 9.99*** | 13.53*** | -21.30*** |
| | (8.12) | (7.70) | (3.44) | (3.28) | (7.76) |
| | [6.43] | [6.80] | [3.44] | [3.39] | [6.72] |
| Adding distance to coast and rivers | -20.67*** | -27.34*** | 8.77*** | 15.44*** | -11.91*** |
| | (4.21) | (5.05) | (2.78) | (2.62) | (4.19) |
| | [3.41] | [4.79] | [3.05] | [2.77] | [4.14] |
| Adding remoteness within Germany | -29.75*** | -33.37*** | 12.13*** | 15.75*** | -17.62*** |
| | (7.13) | (6.98) | (3.05) | (3.03) | (6.60) |
| | [5.27] | [5.79] | [2.77] | [2.94] | [5.73] |
| Adding mean sunshine hours | -23.83*** | -30.26*** | 14.08*** | 20.51*** | -9.75** |
| | (6.18) | (6.44) | (3.55) | (3.41) | (4.12) |
| | [3.79] | [4.85] | [2.37] | [2.55] | [3.61] |
| Adding all of the above | -23.87*** | -31.00*** | 10.61*** | 17.75*** | -13.26*** |
| | (6.73) | (6.64) | (3.91) | (3.18) | (5.08) |
| **D. State fixed effects** | | | | | |
| Adding federal state FE | -15.46** | -22.40*** | 6.36 | 13.30*** | -9.10* |
| | (6.46) | (5.33) | (6.52) | (4.47) | (4.92) |
| | [5.59] | [6.72] | [6.38] | [2.51] | [6.51] |
| Adding geography controls & federal state FE | -19.53*** | -25.52*** | 8.71 | 14.70*** | -10.82* |
| | (6.22) | (6.05) | (5.74) | (4.34) | (5.62) |
| | [6.33] | [8.50] | [5.55] | [2.93] | [7.09] |
| Adding 1834 state FE | -19.03*** | -23.83*** | 11.07* | 15.87*** | -7.96 |
| | (6.41) | (5.51) | (5.69) | (4.66) | (5.09) |
| | [5.62] | [6.34] | [4.26] | [3.40] | [6.30] |
| Adding geography controls & 1834 state FE | -23.44*** | -27.56*** | 10.40* | 14.52*** | -13.04*** |
| | (6.68) | (6.20) | (5.40) | (4.29) | (4.98) |
| | [6.71] | [8.18] | [4.21] | [3.21] | [6.92] |
| Adding indicator for 1834 Zollverein members | -30.91*** | -33.90*** | 12.79*** | 15.78*** | -18.12*** |
| | (6.70) | (6.50) | (3.21) | (2.96) | (6.04) |
| | [5.30] | [5.84] | [2.89] | [2.96] | [5.55] |

Notes: The table reports 2SLS estimates of the $\beta_t$ coefficient in equation (2). The dependent variable is the (change in the) percentile rank in the income per capita distribution. All regressions in Panels A to D include land accessibility and the number of towns per area in 1700 as control variables. Regressions in Panels B-D add additional variables to our set of controls. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

Table A-4: Robustness checks of 2SLS estimates: Other checks and specifications

| | 1926-2019 (1) | 1957-2019 (2) | 1926 (3) | 1957 (4) | 2019 (5) |
|---|---|---|---|---|---|
| *A. Baseline specification* | | | | | |
| Baseline specification | -30.04*** | -33.62*** | 12.22*** | 15.80*** | -17.82*** |
| | (7.52) | (7.24) | (3.25) | (3.14) | (6.89) |
| | [5.75] | [6.18] | [2.96] | [3.14] | [6.03] |
| *B. Explanatory variable* | | | | | |
| Employment share industry 1895 | -29.57*** | -33.09*** | 12.03*** | 15.55*** | -17.54*** |
| | (6.02) | (5.47) | (3.19) | (3.06) | (5.92) |
| | [4.04] | [3.61] | [3.32] | [3.12] | [4.72] |
| Employment share industry 1907 | -29.90*** | -33.46*** | 12.17*** | 15.73*** | -17.74*** |
| | (5.69) | (4.97) | (3.17) | (2.97) | (5.80) |
| | [4.16] | [3.30] | [3.62] | [3.30] | [4.53] |
| Adding tertiary employment share in 1882 | -29.06*** | -33.51*** | 9.48*** | 13.93*** | -19.58** |
| | (7.47) | (7.41) | (3.63) | (3.63) | (7.68) |
| | [6.20] | [6.56] | [3.69] | [3.83] | [6.45] |
| *C. Alternative specifications of the instrument* | | | | | |
| Based on Bairoch (1990) | -17.71*** | -26.66*** | 4.53 | 13.49*** | -13.17*** |
| | (3.41) | (4.64) | (3.48) | (2.55) | (4.81) |
| | [3.30] | [5.56] | [3.58] | [2.60] | [5.10] |
| Squared transport costs | -27.59*** | -32.14*** | 10.04*** | 14.58*** | -17.56*** |
| | (6.50) | (6.52) | (3.24) | (3.07) | (6.58) |
| | [5.66] | [6.06] | [3.47] | [3.08] | [6.09] |
| Only larger river | -30.19*** | -33.56*** | 13.04*** | 16.41*** | -17.16** |
| | (7.60) | (7.29) | (3.28) | (2.99) | (6.71) |
| | [5.68] | [5.93] | [3.04] | [2.97] | [5.67] |
| Log access to carboniferous strata, Daudin | -17.53*** | -24.44*** | -0.47 | 6.44 | -18.00** |
| | (5.99) | (5.43) | (8.58) | (6.28) | (8.37) |
| | [6.52] | [5.53] | [8.46] | [6.21] | [9.10] |
| Log access to carboniferous strata, Bairoch | -21.13*** | -31.71*** | -1.75 | 8.83* | -22.88*** |
| | (5.56) | (6.85) | (7.65) | (5.01) | (8.82) |
| | [5.86] | [7.12] | [7.29] | [4.74] | [9.04] |
| Accounting for coal fields in 1967 | -28.77*** | -32.64*** | 11.63*** | 15.50*** | -17.14** |
| | (6.78) | (6.97) | (3.11) | (2.93) | (6.89) |
| | [5.08] | [5.73] | [3.06] | [2.95] | [5.91] |
| *D. Miscellaneous* | | | | | |
| Without Ruhr area | -43.52*** | -43.21*** | 17.61** | 17.30*** | -25.91** |
| | (14.44) | (12.70) | (6.91) | (5.96) | (13.20) |
| | [10.50] | [7.89] | [6.64] | [5.67] | [9.65] |
| Weighted by 1882 population | -22.67*** | -24.54*** | 8.75*** | 10.61*** | -13.93*** |
| | (3.89) | (3.66) | (2.89) | (2.81) | (4.51) |
| | [2.91] | [2.98] | [2.28] | [2.61] | [3.72] |
| Log income as dependent variable | -0.27*** | -0.24*** | 0.16*** | 0.13*** | -0.11** |
| | (0.06) | (0.04) | (0.05) | (0.02) | (0.04) |
| | [0.05] | [0.04] | [0.04] | [0.02] | [0.04] |
| *E. District level (Regierungsbezirke, N=36)* | | | | | |
| Percentile rank as dependent variable | -20.46*** | -28.18*** | 3.59 | 11.31** | -16.87* |
| | (5.64) | (9.02) | (7.27) | (5.58) | (9.49) |
| | [5.49] | [9.64] | [7.24] | [5.53] | [9.18] |
| Log income as dependent variable | -0.19*** | -0.20*** | 0.09 | 0.10*** | -0.11* |
| | (0.05) | (0.04) | (0.07) | (0.04) | (0.06) |
| | [0.05] | [0.05] | [0.07] | [0.04] | [0.05] |

Notes: The table reports 2SLS estimates of the $\beta_t$ coefficient in equation (2). If not noted otherwise, the dependent variable is the (change in the) percentile rank in the income per capita distribution. All regressions in Panels A to E include land accessibility and the number of towns per area in 1700 as control variables. Regressions in Panel B take the explanatory variable of interest, the industrial employment share, from later occupation censuses or add a control for the tertiary employment share in 1882. Regressions in Panel C vary the cost vectors for sea, river and overland transport or the coal deposits used for constructing the instrumental variable. Regressions in Panel D present miscellaneous robustness checks. Panel E re-estimates the regression at the level of districts (*Regierungsbezirke*), of which there are 36 in West Germany. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. $^{***}$, $^{**}$, and $^{*}$ denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

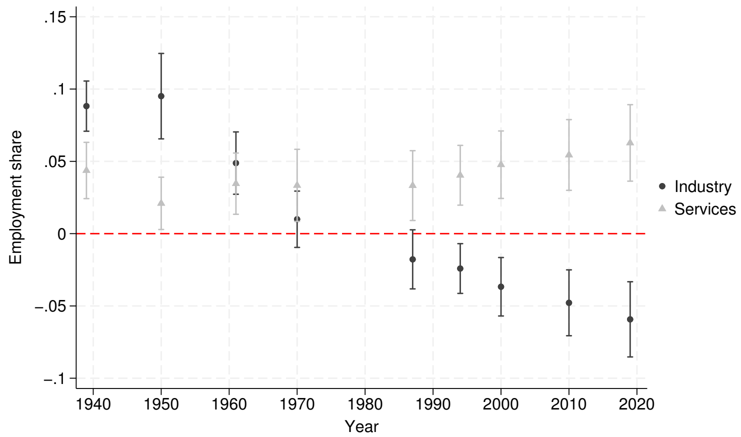
### A.3 Lopsided economic structure and limited adjustment capacity

Table A-5: 2SLS mediation analysis

| | $I_{1882} \rightarrow y_{1957-2019}$ (1) | $I_{1882} \rightarrow M_{1907}$ (2) | $M_{1907} \rightarrow y_{2019}$ (3) |
|---|---|---|---|
| **A. Firm size 1907** | | | |
| Employment share industry 1882 ($I_{1882}$) | -33.62*** | 2.875*** | 3.364 |
| | (6.180) | (0.542) | (9.531) |
| Firm size 1907 ($M_{1907}$) | | | -12.865** |
| | | | (4.864) |
| | Direct effect: 3.36 (9.53) | | |
| | Indirect effect: -36.98** (15.62) | | |
| **B. Employment share in sectors dominated by firms w/ $\geq$ 501 workers in 1907** | | | |
| Employment share industry 1882 ($I_{1882}$) | -33.62*** | 0.071*** | -8.873 |
| | (6.180) | (0.019) | (5.984) |
| Employment share in large-firm sectors 1907 ($M_{1907}$) | | | -349.7*** |
| | | | (123.5) |
| | Direct effect: -8.87 (5.98) | | |
| | Indirect effect: -24.75** (10.97) | | |
| **C. Employment share in sectors dominated by firms w/ $\geq$ 1000 workers in 1907** | | | |
| Employment share industry 1882 ($I_{1882}$) | -33.62*** | 0.060*** | -11.76** |
| | (6.180) | (0.021) | (5.289) |
| Employment share in large-firm sectors 1907 ($M_{1907}$) | | | -366.9** |
| | | | (150.1) |
| | Direct effect: -11.76** (5.29) | | |
| | Indirect effect: -21.86* (11.73) | | |
| **D. Industry concentration in 1907** | | | |
| Employment share industry 1882 ($I_{1882}$) | -33.62*** | 0.049* | -13.97** |
| | (6.180) | (0.027) | (5.634) |
| HHI index of industry concentration size 1907 ($M_{1907}$) | | | -404.4 |
| | | | (250.9) |
| | Direct effect: -13.97** (5.63) | | |
| | Indirect effect: -19.65 (16.34) | | |
| **E. First principal component** | | | |
| Employment share industry 1882 ($I_{1882}$) | -33.62*** | 2.249*** | -8.491 |
| | (6.180) | (0.691) | (6.034) |
| First principal component ($M_{1907}$) | | | -11.171** |
| | | | (4.545) |
| | Direct effect: -8.49 (6.03) | | |
| | Indirect effect: -25.13*** (12.81) | | |

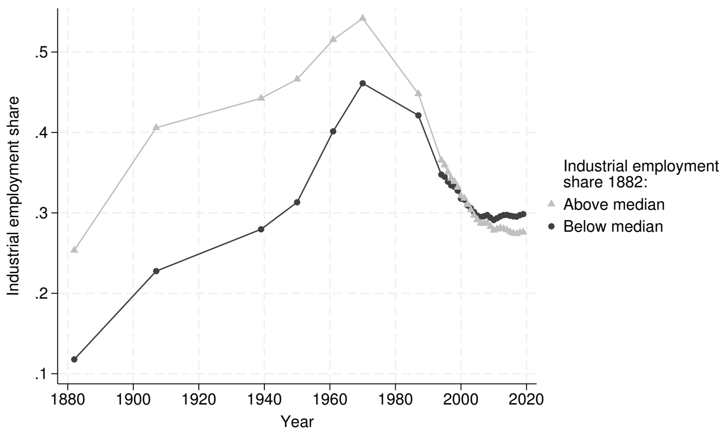
Notes: The table presents second stage results of the causal mediation framework for linear IV models introduced in Dippel, Gold, Heblich & Pinto (2020a). The outcome variable is the change in the GDP per capita rank between 1957 and 2019. The mediator variables are firm size in 1907 (Panel A), the employment share in sectors in which at least 50% of employees work in firms with at least 501 workers (Panel B) or 1000 workers (Panel C), the HHI-Index of industry concentration with $\alpha = 2$ (Panel D), and the first principal component of the four aforementioned indicators (Panel E). Column (1) reproduces, from Column (2) of Table 1, 2SLS regression results of the effect of early industrialization on the outcome. Column (2) reproduces, from Table 2, 2SLS regression results of the effect of early industrialization on the mediator variable. Column (3) presents estimates of the effect of the mediator variable on the outcome (controlling for early industrialization). The indirect effect is the product of the coefficients on early industrialization in Column (2) and on the mediator variable in Column (3). The instrument is the weighted average distance to European coal fields where coal field sizes serve as weights (see Section 2). All regressions include land accessibility and the number of towns per area in 1700 as control variables. Standard errors clustered at the level of administrative districts (*Regierungsbezirke*) are in round brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively.

Figure A-3: The effect of early industrialization on the employment shares in industry and services, 1939-2019




Notes: The figure plots the coefficient estimates from separate 2SLS regressions of the employment share in industry (blue) and services (yellow) in different years (1939, 1950, 1961, 1970, 1987, 1994, 2000, 2010, 2019) on the 1882 industrial employment share. Point estimates are marked by a dot. The vertical bands indicate the 90% confidence interval of each estimate. The 1882 employment share in industry is standardized with a mean of zero and a standard deviation of one. All regressions include land accessibility and the number of towns per area in 1700 as control variables.

Figure A-4: Mean industrial employment share by 1882 industrial employment, 1882-2019

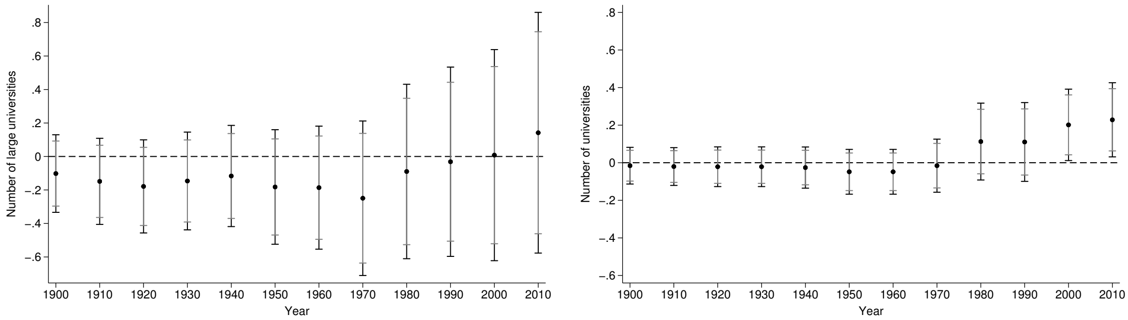



Notes: The figure plots the average employment share of the labor force in industry (in %), separately for labor markets with above and below median industrial employment share in 1882.

Figure A-5: Impact of early industrialization on the number of universities and colleges, 1900-2010

(a) All universities and colleges

(b) Large universities and colleges




Notes: The figure plots 2SLS estimates of the impact of early industrialization on the number of universities and colleges (including technical universities and universities of the applied sciences). Panel (a) counts all universities and colleges, while panel (b) focuses on large universities and colleges with at least 7,500 students today. The vertical bands in gray and black indicate 90% and 95% confidence intervals, respectively. The 1882 employment share in industry, our explanatory variable of interest, is standardized with a mean of zero and a standard deviation of one.

Table A-6: Effects of early industrialization on the population share of enrolled students, 2019

| | | By subject | | | | | |
|---|---|---|---|---|---|---|---|
| | Total | Humanities | Law, econ, social sciences | Maths, natural sciences | Medicine | Engineering | Others |
| Employment share industry 1882 | 0.762** | 0.130** | 0.252* | 0.122** | 0.031 | 0.231** | -0.004 |
| | (0.311) | (0.062) | (0.137) | (0.052) | (0.023) | (0.104) | (0.014) |
| | [0.367] | [0.069] | [0.130] | [0.054] | [0.024] | [0.142] | [0.015] |
| **Outcome statistics** | | | | | | | |
| Mean | 2.417 | 0.327 | 0.923 | 0.264 | 0.119 | 0.701 | 0.082 |
| Standard deviation | 2.417 | 0.517 | 0.979 | 0.395 | 0.223 | 0.776 | 0.175 |

Notes: The table reports the results of 2SLS regressions of the effect of early industrialization on the number of students enrolled in 2019 relative to the local population. Column (1) focuses on the total number of students, while Columns (2)-(7) distinguish between the subjects that students study. Humanities include arts; medicine include health sciences; other subjects include agricultural, forestry and nutritional sciences, sorts, and unclassified subjects. All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

Table A-7: Effects of early industrialization on educational attainment

| | Population share with | | | | | |
|---|---|---|---|---|---|---|
| | university degree | | at least higher secondary or vocational degree | | no school-leaving qualification | |
| | 1970 (1) | 2011 (2) | 1970 (3) | 2011 (4) | 1970 (5) | 2011 (6) |
| Employment share industry 1882 | -0.175 | -0.059 | -1.246* | -1.351** | 0.755** | 1.146*** |
| | (0.109) | (0.494) | (0.729) | (0.615) | (0.302) | (0.127) |
| | [0.096] | [0.681] | [0.466] | [0.565] | [0.292] | [0.143] |
| **Outcome statistics** | | | | | | |
| Mean | 2.294 | 11.701 | 19.060 | 71.164 | 3.275 | 4.494 |
| Standard deviation | 0.614 | 3.187 | 3.973 | 2.044 | 0.944 | 1.200 |

Notes: The table reports results from 2SLS regressions of the effect of early industrialization on regional educational outcomes. Columns (1)-(2) focus on the population share with a university degree (in %) as dependent variable, Columns (3)-(4) on the population share with at least a higher secondary degree (*Mittlere Reife*) or a vocational degree (including those with tertiary education), and Columns (5)-(6) on the population share without a school-leaving qualification. The data come from the population censuses of 1970 and 2011. The 1882 employment share in industry is standardized with a mean of zero and a standard deviation of one. All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

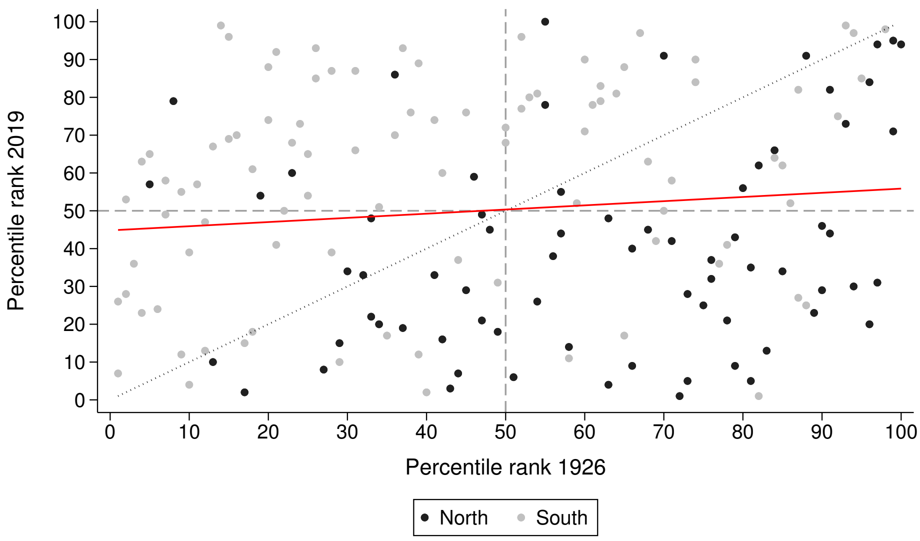
Table A-8: Early industrialization, self-employment, and political outcomes

| | Economic structure | | | Political outcomes | | |
|---|---|---|---|---|---|---|
| | Industrial apprentices share 1970 | HHI index of industry concentration 1950 | Self-employment share 1950 | Years w/ major from dominant party 1950-90 | Years w/ Social Dem. major 1950-90 | Vote share Social Democrats 1957 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Employment share industry 1882 | -0.019*** | 0.055*** | -0.039*** | 6.067*** | 5.243*** | 0.044*** |
| | (0.006) | (0.021) | (0.008) | (1.679) | (1.895) | (0.014) |
| | [0.007] | [0.027] | [0.009] | [1.857] | [1.584] | [0.013] |
| **Outcome statistics** | | | | | | |
| Mean | 0.054 | 0.110 | 0.168 | 27.008 | 16.683 | 0.269 |
| Standard deviation | 0.045 | 0.071 | 0.038 | 10.009 | 13.730 | 0.092 |

Notes: The table shows the results of 2SLS regressions of the effect of early industrialization on the employment share of industrial apprentices in industry in 1970 (Column (1)), the sectoral concentration of industrial employment in 1950 (Column (2)), the self-employment share in 1950 (Column (3)), the number of years the major was member of the locally dominant party in 1950-1990 (Column (4)), the number of years the major was member of the Social Democrats in 1950-1990 (Column (5)), and the vote share of the Socials Democrats in the national election of 1957 (Column (6)). We measure employment concentration by the HHI-Index (with $\alpha = 2$). The 1882 employment share in industry is standardized with a mean of zero and a standard deviation of one. All regressions include land accessibility and the number of towns per area in 1700 as control variables. Conley standard errors (Bartlett kernel, 100 km cut-off) are reported in round brackets; standard errors clustered at the level of administrative districts (*Regierungsbezirke*) in square brackets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively, based on Conley standard errors.

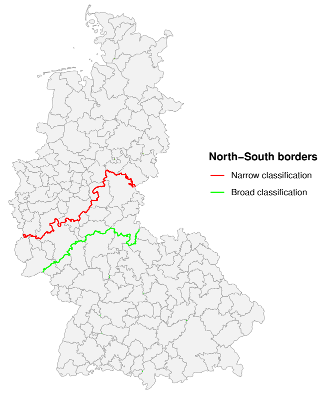
## A.4 North-south reversal and changing inequality

Figure A-6: Per capita income rank of northern versus southern German labor markets in 1926 and 2019




Notes: The figure plots the percentile rank in the 2019 income per capita distribution against the rank in 1926, along with the linear regression line in red. Each dot represents a labor market. Labor markets in northern Germany are indicated by dark gray dots, labor markets in southern Germany by light gray dots. Dashed horizontal and vertical lines indicate median percentile ranks. The dotted line indicates identical percentile ranks in 2019 and 1926.

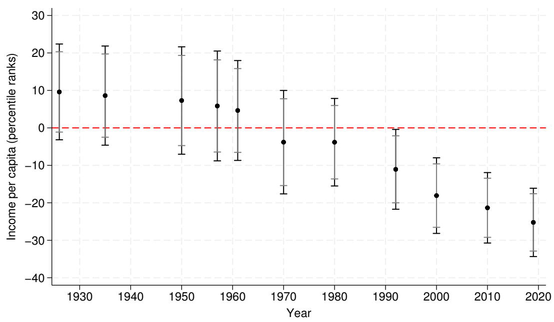
Figure A-7: Definitions of North and South Germany



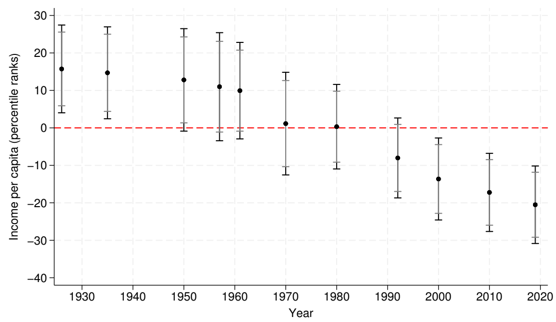

Notes: The narrower baseline definition adheres to federal states borders. It classifies labor markets located in Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, and Schleswig-Holstein as northern. Southern labor markets are those in Bavaria, Baden-Württemberg, Hesse, and Rhineland-Palatinate. The second, broader classification also assigns the northern parts of Hesse and Rhineland-Palatinate to the north.

Figure A-8: Percentile rank differences between northern and southern labor markets, alternative classifications, 1926-2019

(a) Broad definition




(b) Split at mean latitude




Notes: The figure plots the average difference in percentile ranks between northern and southern regions for alternative classifications of northern and southern regions. Point estimates are marked by a dot. The vertical bands in gray and black indicate 90% and 95% confidence intervals, respectively. Panel (a) classifies labor markets located in Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, Schleswig-Holstein, and the northern parts of Hesse and Rhineland-Palatinate as North Germany. Southern labor markets are those in Bavaria, Baden-Württemberg, and the southern parts of Hesse and Rhineland-Palatinate (see Figure A-7). Panel (b) uses latitude to assign regions. Those with above-mean latitude are classified as North Germany.

Table A-9: Components of changes in regional income inequality, 1957-2019, alternative measures

| | 1957 (1) | 1980 (2) | 2019 (3) | 1957-1980 (2)-(1) | 1980-2019 (3)-(2) | 1957-2019 (3)-(1) |
|---|---|---|---|---|---|---|
| **Panel A.** Standard deviation of log GDP per capita | | | | | | |
| $\sigma_{y_t}$ | 0.234 | 0.162 | 0.191 | -0.073 | 0.029 | -0.043 |
| $\sigma_{y_t^c}$ | 0.181 | 0.162 | 0.212 | -0.019 | 0.050 | 0.032 |
| $\Delta$ | 0.054 | -0.000 | -0.021 | -0.054 | -0.021 | -0.075 |
| | (0.009) | (0.008) | (0.012) | (0.009) | (0.008) | (0.013) |
| **Panel B.** Coefficient of variation | | | | | | |
| $CV_{y_t}$ | 0.242 | 0.168 | 0.208 | -0.073 | 0.039 | -0.034 |
| $CV_{y_t^c}$ | 0.186 | 0.169 | 0.224 | -0.017 | 0.055 | 0.037 |
| $\Delta$ | 0.056 | -0.00 | -0.016 | -0.056 | -0.016 | -0.072 |
| | (0.010) | (0.008) | (0.014) | (0.010) | (0.011) | (0.016) |
| **Panel C.** Gini coefficient | | | | | | |
| $Gini_{y_t}$ | 0.132 | 0.091 | 0.107 | -0.041 | 0.016 | -0.024 |
| $Gini_{y_t^c}$ | 0.100 | 0.091 | 0.120 | -0.009 | 0.029 | 0.020 |
| $\Delta$ | 0.031 | -0.000 | -0.013 | -0.032 | -0.013 | -0.044 |
| | (0.005) | (0.005) | (0.007) | (0.006) | (0.005) | (0.008) |
| **Panel D.** 90/10 ratio | | | | | | |
| $90/10_{y_t}$ | 1.799 | 1.514 | 1.578 | -0.286 | 0.064 | -0.222 |
| $90/10_{y_t^c}$ | 1.515 | 1.517 | 1.695 | 0.002 | 0.178 | 0.180 |
| $\Delta$ | 0.284 | -0.004 | -0.118 | -0.288 | -0.114 | -0.402 |
| | (0.082) | (0.040) | (0.077) | (0.083) | (0.063) | (0.106) |

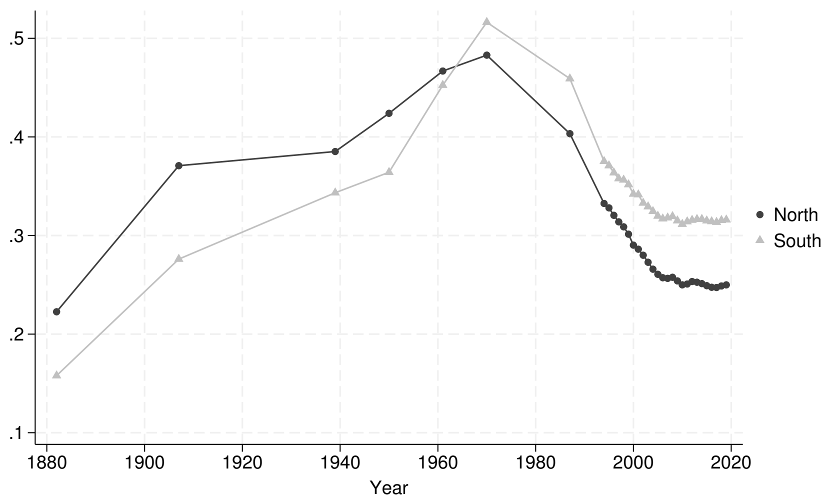
Notes: The table reports and decomposes the change in regional inequality between 1957 and 2019 using different measures of regional inequality. Panel A. uses the standard deviation of log GDP per capita, i.e., our baseline measure from Table 5. Panel B. uses the coefficient of variation of GDP per capita, Panel C. the Gini coefficient of GDP per capita, and Panel D. the ratio of GDP per capita in the labor market at the 90th percentile to that at the 10th percentile. Each panel reports the inequality measure as applied to actual and counterfactual per capita GDP for 1957 (Column (1)), 1980 (Column (2)), and 2019 (Column (3)). The last row of each panel reports the difference between inequality in actual and counterfactual income, i.e., the effect of early industrialization on income dispersion in year $t$. The last three columns report changes between 1957-1980, 1980-2019, and 1957-2019, respectively. Gray shaded cells report the industrialization effect, $\Delta IND_{t,t-1}$, as defined in equation (4) for our baseline measure. Bootstrapped standard errors based on 200 bootstrap replications are in round brackets.

Table A-10: Decomposition of changes in $\sigma_{y_t}$, 1957-2019

| | 1957-1980 (1) | 1980-2019 (2) | 1957-2019 (3) |
|---|---|---|---|
| Total change | -0.073 | 0.029 | -0.043 |
| Of which due to: | | | |
| change in $\hat{\beta}_t$ | -0.084 | -0.021 | -0.106 |
| remainder | 0.011 | 0.050 | 0.063 |

Notes: The table decomposes the change in $\sigma_{y_t}$ into two components, the effect of changes in $\hat{\beta}_t$ and a remainder. The decomposition is $\overbrace{[\sigma_{y_t} - \sigma_{y_{t-1}}]}^{\text{Actual change}} = \overbrace{[\sigma_{y_t} - \sigma_{y^*}]}^{\text{Coefficient effect}} + \overbrace{[\sigma_{y^*} - \sigma_{y_{t-1}}]}^{\text{Remainder}}$ where $y^* = y_t - (\hat{\beta}_t - \hat{\beta}_{t-1})I_{i,1882}$. To obtain $y^*$, we thus replace the coefficient $\hat{\beta}_t$ in equation (2) by $\hat{\beta}_{t-1}$, while holding the effect of other observables and the distribution of residuals fixed. The decomposition is similar in spirit to that proposed by Juhn, Murphy & Pierce (1993). In our context, however, observable characteristics, including $I_{i,1882}$, do not vary over time. See Fortin, Lemieux & Firpo (2011) for an overview of the scope and limitations of different methods for decomposing distributional statistics.

Figure A-9: Average industrial employment shares in northern and southern labor markets, 1882-2019




Notes: The figure plots the average employment share of the labor force in industry (in %), separately for northern and southern German labor markets.

## References


Arbeitskreis VGR der Länder (2021). *Bruttoinlandsprodukt, Bruttowertschöpfung in den kreisfreien Städten und Landkreisen der Bundesrepublik Deutschland 1992 und 1994 bis 2019.* Stuttgart: Statistisches Landesamt Baden-Württemberg.

Asch, K. (2003). *The 1:5 Million International Geological Map of Europe and Adjacent Areas.* Schweizerbart Science Publishers.

Bairoch, P. (1990). The impact of crop yields, agricultural productivity, and transport costs on urban growth between 1800 and 1910. In A. van der Woude, A. Hayami, & J. de Vries (Eds.), *Urbanization in history. A process of dynamic interactions*, Urbanization in History (pp. 134–151). Oxford University Press.

Bayerisches Statistisches Landesamt (1952). Volks- und Berufszählung am 13. September 1950 in Bayern. Berufszählung. Die Erwerbstätigkeit der Bevölkerung. *Beiträge zur Statistik Bayerns*, 186.

Bergeaud, A. & Verluise, C.. (2024). A new dataset to study a century of innovation in Europe and in the US. *Research Policy*, 53(1): 104903.

Berlingieri, F., Gathmann, C., & Quinckhardt, M. (2022). College openings and local economic development. *CEPR Discussion Paper*, 17374, CEPR.

BGR (2014). Ackerbauliches Ertragspotential der Böden in Deutschland (SOIL QUALITY RATING). Technical report, Bundesanstalt für Geowissenschaften und Rohstoffe, Stilleweg 2, 30655 Hannover.

Braun, S. T. & Franke, R. (2021). A county-level database on expellees in West Germany, 1939–1961. *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*, 108(4), 522–540.

Cantoni, D., Mohr, C., & Weigand, M. (2020). Princes and townspeople: a collection of historical statistics on German territories and cities. 3: Town charters and first mentions.

Daudin, G. (2010). Domestic trade and market size in late-eighteenth-century France. *The Journal of Economic History*, 70(3), 716–743.

Deutscher Städtetag (1949). *Statistisches Jahrbuch Deutscher Gemeinden 1949.* Schwäbisch Gmünd: Alfons Bürger Verlag.

Deutscher Wetterdienst (DWD) (2021). DWD Climate Data Center (CDC): Multi-annual grids of annual sunshine duration over Germany 1991-2020,version v1.0, 2021.

Dippel, C., Gold, R., Heblich, S., & Pinto, R. (2020a). Mediation analysis in IV settings with a single instrument. *Unpublished manuscript.*

Fernihough, A. & O'Rourke, K. H. (2020). Coal and the European Industrial Revolution. *The Economic Journal*, 131(635), 1135–1149.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In *Handbook of Labor Economics* (pp. 1–102). Elsevier.

Hessisches Statistisches Landesamt (1952). Wirtschaftliche und soziale Gliederung der Bevölkerung in den hessischen Kreisen. Ergebnisse der Berufszählung vom 13. September 1950. *Beiträge zur Statistik Hessens. Sonderreihe: Berufszählung 1950*, 2.

HGIS Germany (Andreas Kunz, Leonhard Dietze). IEG-Maps / Berlin Data Pool (Andreas Kunz, Robert Moeschl). Institut fur Europaische Geschichte (Mainz, Rhineland-Palatinate, Germany). Fachhochschule Mainz (Mainz, Rhineland-Palatinate, Germany). (2007). Germany State Boundaries, 1834, German Historical GIS. [Shapefile]. HGIS Germany, retrieved from https://hgl.harvard.edu/catalog/harvard-ghgis1834core.

Institut für Weltwirtschaft (1974). *Infrastruktur, räumliche Verdichtung und sektorale Wirtschaftsstruktur als Bestimmungsgründe des regionalen Entwicklungspotentials in den Arbeitsmarktregionen der Bundesrepublik Deutschland.* IfW Kiel.

Isbary, G., Von der Heide, H.-J., & Müller, G. (1969). Gebiete mit gesunden Strukturen und Lebensbedingungen. Merkmale und Abgrenzung. Gutachten im Auftrag des Bundesministers des Innern. *Veröffentlichungen der ARL: Abhandlungen 57.* Hannover.

Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3), 410–442.

Kaiserliches Statistisches Amt (1884). Berufsstatistik nach der allgemeinen Berufszählung vom 5. Juni 1882. 1. Berufsstatistik des Reichs und der kleineren Verwaltungsbezirke. *Statistik des Deutschen Reichs. Neue Folge*, 2.

Kaiserliches Statistisches Amt (1897). Berufs- und Gewerbezählung vom 14. Juni 1895. Berufsstatistik der kleineren Verwaltungsbezirke. *Statistik des Deutschen Reichs. Neue Folge*, 109.

Kaiserliches Statistisches Amt (1907). Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung I. *Statistik des Deutschen Reichs. Neue Folge*, 213.

Kaiserliches Statistisches Amt (1909a). Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VI. *Statistik des Deutschen Reichs. Neue Folge*, 218.

Kaiserliches Statistisches Amt (1909b). Berufs- und Betriebsählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VII. *Statistik des Deutschen Reichs. Neue Folge*, 219.

Kaiserliches Statistisches Amt (1910). Berufs- und Betriebsählung vom 12. Juni 1907. Berufsstatistik. Abteilung VIII, kleinere Verwaltungsbezirke. *Statistik des Deutschen Reichs. Neue Folge*, 209.

Keller, W. & Shiue, C. H. (2014). Endogenous formation of free trade agreements: evidence from the Zollverein's impact on market integration. *The Journal of Economic History*, 74(4), 1168–1204.

Konstantinos, M. A., Schumann, G. J.-P., & Pavelsky, T. (2013). A simple global river bankfull width and depth database. *Water Resources Research*, 49(10), 7164–7168.

Max Planck Institute for Demographic Research (MPIDR) and Chair for Geodesy and Geoinformatics, University of Rostock (CGG) (2011). MPIDR Population History GIS Collection (partly based on Bundesamt für Kartographie und Geodäsie). Mosaic Project.

Niedersächsisches Amt für Landesplanung und Statistik (1953). Die Bevölkerung Niedersachsens nach den Ergebnissen der Volkszählung am 13. September 1950. B. Tabellenteil. *Zählung der Bevölkerung, Gebäude, Wohnungen und nichtlandwirtschaftlichen Arbeitsstätten 1950*, 1.

Redding, S. J. & Sturm, D. M. (2008). The costs of remoteness: evidence from German division and reunification. *American Economic Review*, 98(5), 1766–1797.

Schmitt, K., Rattinger, H., & Oberndörfer, D. (1994). Kreisdaten (Volkszählungen 1950-1987). *GESIS Datenarchiv, Köln. ZA2472*, (Datenfile Version 1.0.0).

Statistische Landesämter (1968). Das Bruttoinlandsprodukt der kreisfreien Städte und Landkreise 1957 bis 1966. *Volkswirtschaftliche Gesamtrechung der Länder*, 3.

Statistische Landesämter (1973). Das Bruttoinlandsprodukt der kreisfreien Städte und Landkreise 1961, 1968 und 1970. *Volkswirtschaftliche Gesamtrechung der Länder*, 4.

Statistische Landesämter (1978). Bruttoinlandsprodukt und Bruttowertschöpfung der kreisfreien Städte und Landkreise in der Bundesrepublik Deutschland 1970, 1972 und 1974. *Volkswirtschaftliche Gesamtrechung der Länder*, 8.

Statistische Landesämter (1979). Bruttoinlandsprodukt und Bruttowertschöpfung der kreisfreien Städte und Landkreise in der Bundesrepublik Deutschland 1976. *Volkswirtschaftliche Gesamtrechung der Länder*, 10.

Statistische Landesämter (1998). Bruttowertschöpfung der kreisfreien Städte und Landkreise in der Bundesrepublik Deutschland 1980, 1992, 1994 und 1996. *Volkswirtschaftliche Gesamtrechung der Länder*, 33.

Statistisches Bundesamt (1955). Die Umsätze der Umsatzsteuerpflichtigen und deren Besteuerung (Ergebnisse der Statistik über die Umsatzsteuerveranlagung für 1950). *Statistik der Bundesrepublik Deutschland*, 112.

Statistisches Bundesamt (1956). Gebäude- und Wohnungszählung in der Bundesrepublik Deutschland vom 13. September 1950. *Statistik der Bundesrepublik Deutschland*, 38.

Statistisches Bundesamt (1957a). Die Umsätze der Umsatzsteuerpflichtigen und deren Besteuerung (Ergebnisse der Statistik über die Umsatzsteuerveranlagung für 1955). *Statistik der Bundesrepublik Deutschland*, 184.

Statistisches Bundesamt (1957b). Die Wahl zum 3. Deutschen Bundestag am 15. September 1957. *Statistik der Bundesrepublik Deutschland*, 200, 4.

Statistisches Landesamt Baden-Württemberg (1954). Ergebnisse der Volks- und Berufszählung vom 13. September 1950. Berufszählung. *Statistik von Baden-Württemberg*, 5.

Statistisches Landesamt Bremen (1953). Die Volks- und Berufszählung am 13.9.1950 im Land Bremen. *Statistische Mitteilungen aus Bremen*, 4. Sonderheft.

Statistisches Landesamt der Freien und Hansestadt Hamburg (1953). Die Berufszählung in Hamburg am 13. September 1950. *Statistik des Hamburgischen Staates*, 37.

Statistisches Landesamt Nordrhein-Westfalen (1952a). Die Wohnbevölkerung nach der Erwerbstätigkeit in Nordrhein-Westfalen. Landesteil Nordrhein. Ergebnisse der Berufszählung vom 13. September 1950. *Sonderreihe Volkszählung 1950*, 5a.

Statistisches Landesamt Nordrhein-Westfalen (1952b). Die Wohnbevölkerung nach der Erwerbstätigkeit in Nordrhein-Westfalen. Landesteil Westfalen. Ergebnisse der Berufszählung vom 13. September 1950. *Sonderreihe Volkszählung 1950*, 5b.

Statistisches Landesamt Rheinland-Pfalz (1952). Die Berufszählung in Rheinland-Pfalz. Volkszählung am 13. September 1950. *Statistik von Rheinland-Pfalz*, 13.

Statistisches Landesamt Schleswig-Holstein (1953). Die Erwerbstätigkeit in Schleswig-Holstein. Ergebnisse der Berufszählung vom 13. Septermber 1950. *Statistik von Schleswig-Holstein*, 9.

Statistisches Reichsamt (1931). Umsatz und Umsatzsteuer in Deutschland nach den Umsatzsteuerveranlagerungen 1926 bis 1928. *Statistik des Deutschen Reichs*, 361.

Statistisches Reichsamt (1939). Umsatzsteuerstatistik 1935. *Statistik des Deutschen Reichs*, 511(IV).

Statistisches Reichsamt (1941). Die Bevölkerung des Reichs, der Reichsteile, der größeren und kleineren Verwaltungsbezirke, der Gaue der NSDAP, sowie der Großstädte nach der Religionszugehörigkeit auf Grund der Volkszählung vom 17. Mai 1939 *Wirtschaft und Statistik*, 21(9). Verlag für Sozialpolitik, Wirtschaft und Statistik.

van Etten, J. (2017). R Package gdistance: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, 76(13), 1–21.