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.

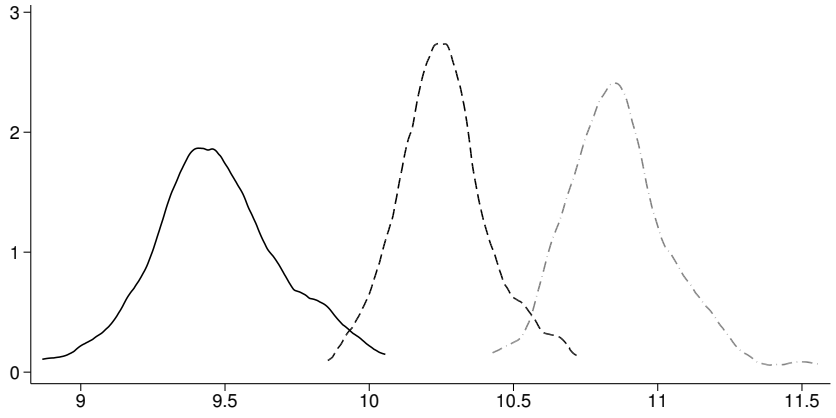
Unit of analysis:

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



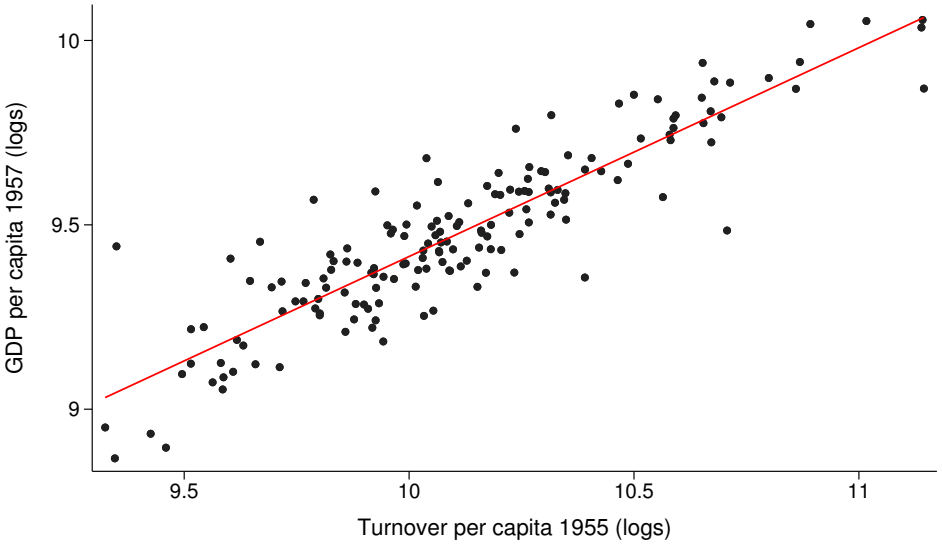
Notes: The plot shows kernel density estimates of the distribution of log GDP per capita (DM, in 1992 prices) in 1957, 1980, and 2019.

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

	1957 (1)	1980 (2)	2019 (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 β_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 β_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.

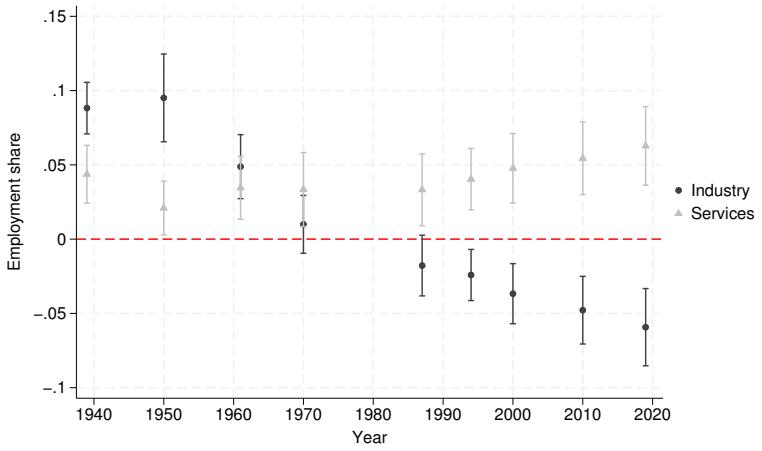
A.3 Lopsided economic structure and limited adjustment capacity

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)	

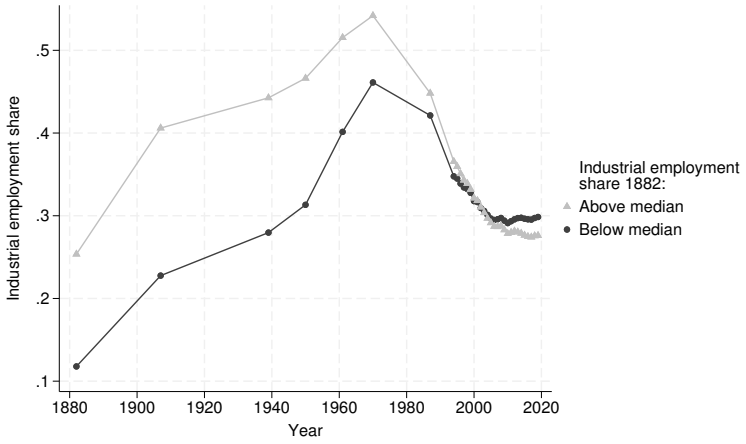
Notes: The table presents second stage results of the causal mediation framework for linear IV models introduced in Dippel, Gold, Heblich & Pinto (2020a). The outcome variable is the change in the GDP per capita rank between 1957 and 2019. The mediator variables are firm size in 1907 (Panel A), the employment 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.

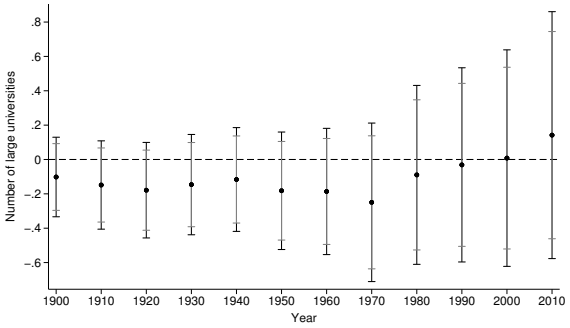
Figure A-4: Mean industrial employment share by 1882 industrial employment, 1882-2019



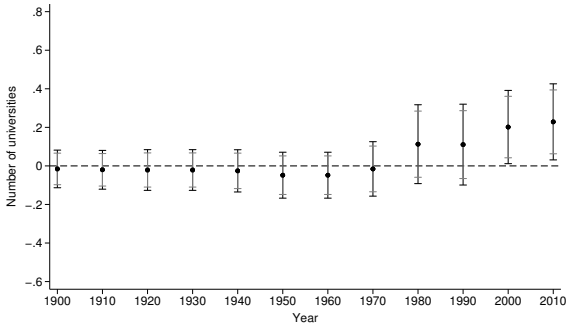
Notes: The figure plots the average employment share of the labor force in industry (in %), separately for labor markets with above and below median industrial employment share in 1882.

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	2011	1970	2011	1970	2011
	(1)	(2)	(3)	(4)	(5)	(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.

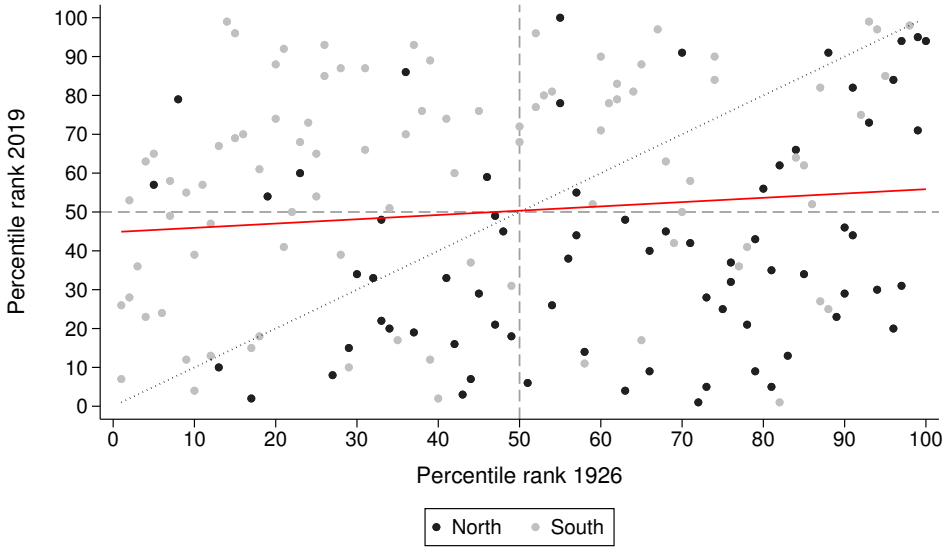
Table A-8: Early industrialization, self-employment, and political outcomes

	Economic structure			Political outcomes		
	Industrial apprentices share 1970 (1)	HHI index of industry concentration 1950 (2)	Self-employment share 1950 (3)	Years w/ major from dominant party 1950-90 (4)	Years w/ Social Dem. major 1950-90 (5)	Vote share Social Democrats 1957 (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.

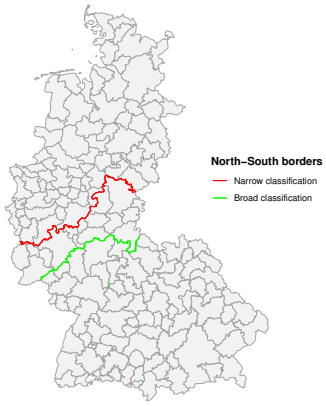
A.4 North-south reversal and changing inequality

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

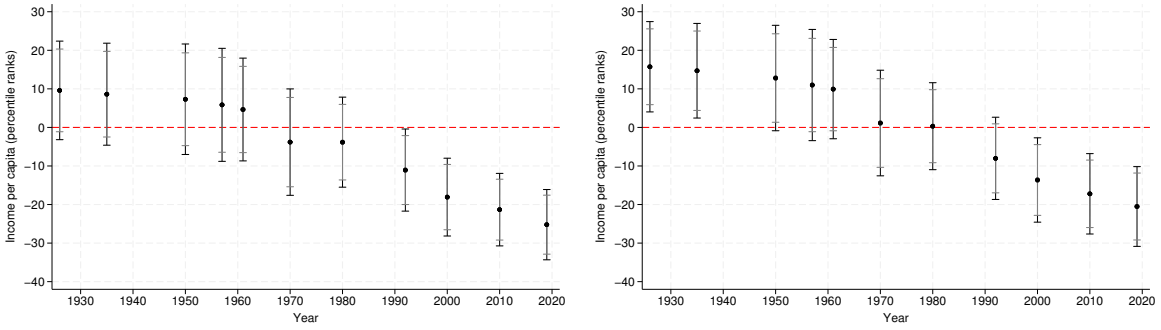


Notes: The narrower baseline definition adheres to federal states borders. It classifies labor markets located in Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, and Schleswig-Holstein as northern. Southern labor markets are those in Bavaria, Baden-Württemberg, Hesse, and Rhineland-Palatinate. 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						
σ_{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
Δ	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
Δ	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
Δ	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
Δ	0.284	-0.004	-0.118	-0.288	-0.114	-0.402
	(0.082)	(0.040)	(0.077)	(0.083)	(0.063)	(0.106)

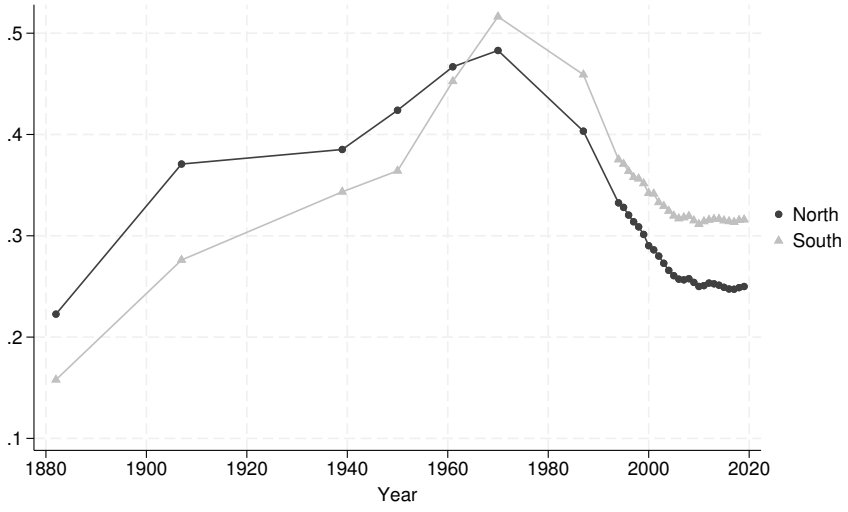
Notes: The table reports and decomposes the change in regional inequality between 1957 and 2019 using different measures of regional inequality. Panel A. uses the standard deviation of log GDP per capita, i.e., our baseline measure from Table 5. Panel B. uses the coefficient of variation of GDP per capita, 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 σ_{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 σ_{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.