Trade and the Size Distribution of Firms: Evidence from the German Empire*

Marcus Biermann (IRES/LIDAM, UCLouvain)

January 2021

Abstract

What effect did trade have on the size distribution of firms during the first wave of globalization? Three historical datasets from the German Empire between 1875 and 1907 were collected and harmonized to answer this question. This paper combines industry census and bilateral railway trade data from the same industry and region along with industry-level tariff data. The evidence shows that increases in aggregate trade caused the share of firms to shift from smaller to larger firms. Exogenous decreases in tariffs caused an increase in the share of the largest firms. The regional distributive effects of trade on inequality between firms that are discussed in the contemporaneous literature were already present during the first wave globalization. (JEL F14, F15, N73)

*marcus.biermann@uclouvain.be. I thank Mario Larch, two anonymous referees as well as Facundo Albornoz, Swati Dhingra, Basile Grassi, Hanwei Huang, Elsa Leromain, Stephan Maurer, Gianmarco Ottaviano, Frank Pisch, Steve Pischke, Thomas Sampson, Daniel Sturm, Jens Südekum, Jacques Thisse, Catherine Thomas, Hans-Joachim Voth and participants of the CEP International Economics Workshop, Frankfurt EGIT, 15th GEP Annual Postgraduate Conference Nottingham, 16th Trans-Atlantic Doctoral Conference LBS, the Urban Economics Association Summer School at the IEC in Barcelona, Annual congress Verein für Socialpolitik in Augsburg 2016, 18th ETSG in Helsinki, Econometric Society Winter Meeting in Edinburgh 2016 for helpful comments and discussions. Financial support by the German National Academic Foundation is gratefully acknowledged.

1 Introduction

The distributional consequences of globalization are one of the most heatedly discussed topics in the public debate and in the academic literature. Import penetration from non-advanced economies - China in particular - into the US and other advanced economies has increased dramatically during the past three decades. This import penetration had effects on both, the distribution among workers and the distribution and size of firms, the latter of which this study focuses on. Import penetration caused in particular an increase in firm exit and a decrease in firm growth, which varied based on how firms were affected by import competition, thereby increasing the regional inequality between firms (for example, Bernard, Jensen, and Schott 2006; Autor, Dorn, and Hanson 2016; Bloom, Draca, and Van Reenen 2016). With regards to exports, trade theory suggests that the reduction of trade barriers induces the self-selection of productive firms into exporting, and this is a further mechanism through which globalization can increase intra-industry firm inequality (Melitz 2003). There is evidence of firms selecting into exporting in contemporaneous data (for example, Bernard et al. 2007).

Are these consequences of globalization unique to recent liberalization periods, or did they occur previously? This paper focuses on a particularly important episode of liberalization: the first wave of globalization, which took place from 1870 to 1913. Previous research on inequality between production factors in this period has examined the effects of economic integration between the Old World and the New World on factor price equalization between the two (for example, O'Rourke and Williamson 1994; O'Rourke, Taylor, and Williamson 1996). When testing the implication of factor price equalization, these studies use computable general equilibrium models and panel data at the country level. Studies employing modern econometric approaches in order to identify the distributional consequences of increased economic integration at the regional level during the first wave are scarce. Determining whether redistribution across firms took place during the first period of globalization is essential; research on this topic would shed light on the stability and generalization of the historical relationship between globalization and firm inequality. Since the regional-level data from the first wave of globalization is very limited (O'Rourke 2019), previous research has not examined whether the rise in regional inequality between firms is a recent phenomenon or if it already took place during the first significant liberalization period. This study contributes to the literature by using regional data to show that increases in aggregate trade and decreases in tariffs caused regional inequality between firms to rise between 1875 and 1907 in the German Empire. Hence, the regional distributive effects of trade were present during the first wave of globalization.

First, this study analyzes the distributional consequences of trade during the first wave of globalization by examining the effects of changes in aggregate trade flows on the size distribution of firms at the district-industry level. Second, it examines changes in industry-level tariffs on the size distribution of firms at the district-industry level in the German Empire initially introduced to respond to rising import competition. The German Empire's economic history

is highly relevant during the first wave of globalization. It experienced rapid external and internal integration during this period, and, before WWI, became the second-largest economy in international trade. The degree of trade achieved by the German Empire before WWI was reached again 60 years later. Germany's railway network became the largest in Europe and internal integration within the Empire increased significantly. There was significant variation in the changes in external and internal integration across districts and industries. On the other hand, certain industries demanded protection against international competition, which resulted in the introduction of the "iron and rye" tariffs in 1879. The tariffs were set in the national parliament and therefore exogenously determined from the perspective of individual districts. The demand for protection against foreign competition led to significant variation in protection policies across various industries during the first wave of globalization.

This study combines three datasets for the first time: industry census data, district-industry level bilateral railway trade data, and industry-level tariff data derived from foreign trade statistics. The industry census data provide rich information on the size distribution of firms at the district-industry level. The firms included in the industry census data are grouped into categories according to their size. Figure 1 shows an upsurge in the industrial concentration of larger firms with more than 50 employees between 1875 and 1907 in the German Empire. Trade theory suggests two mechanisms through which trade can alter the firm size distribution by means of firm selection into exporting as well as import competition from abroad. Firms selecting exporting expand their size by selling to foreign markets, while less productive firms are driven out of the market or lose market share as in Melitz (2003).

This study provides evidence that these mechanisms were present in the German Empire during the first wave of globalization. First, it examines the effects of increased external and internal integration through railway trade on the size distribution of district-industry pairs. The findings show that increases in aggregate railway trade caused a shift in firm share from smaller firms (with up to 50 employees) to larger firms (with more than 50 employees). The estimation results highlight that the channels are quantitatively important. The OLS estimate suggests a 1.79 percent increase in the share of large firms for the 75th percentile relative to the 25th percentile of aggregate trade growth between 1895 and 1907. The IV estimate implies a 4.15 percent increase for the 75th percentile relative to the 25th percentile of aggregate trade growth between 1895 and 1907. The alpha of a Pareto distribution is estimated and a significant decrease by 0.107 units is found for the 75th percentile relative to the 25th percentile of aggregate trade growth, which suggests that the right tail of the distribution becomes thicker. The reallocation takes place through the negative but statistically insignificant growth of the number of smaller firms and the positive, statistically significant growth of the number of larger firms. The study proposes a Bartik-type instrument for identification that can be used to overcome the endogeneity of railway trade. The robustness checks show that these patterns cannot be rationalized through alternative mechanisms such as the measured trade content by only considering railway trade, trade in intermediates, firm location choice near coal mines, and spatial

reallocation toward more urban areas.

Second, the study analyzes the effects of exogenous tariff changes on the size distribution of district-industry pairs in the German Empire. The second main finding is that tariff reductions caused an increase in the share of firms with more than 200 employees. On the other hand, tariff increases carried out in 1879 caused a decrease in the share of larger firms. The elasticity of the share of firms with more than 200 employees with respect to a one percent change in tariffs is 0.11 percent. The share of the largest firms decreased by 0.57 percent due to the median increase in tariffs across industries between 1875 and 1882. On the other hand, the decreases in tariffs between 1882 and 1895 led to an increase in the share of very large firms by 0.37 percent for the median industry. The finding also carries over, when aggregating industry level tariffs to district level tariffs and district-industry firm size distributions to district firm size distributions. The response is stronger for industries in which the import share from abroad is above the median across industries. This study investigates alternative explanations for these effects, such as agglomeration economies, migration, and lobbying through politically influential nobility. The results show that the impact of tariff changes on the share of large firms cannot be rationalized through these alternative hypotheses.

The remainder of this paper is organized as follows: First, the literature is reviewed and the contribution of this paper in relation to the literature is emphasized. The next section provides an overview of the historical setting. The following section describes the data in more detail. Then, section V lays out the estimation strategy and presents the results. The final section concludes.

2 Literature Review

This paper contributes to three strands of the literature. The allocation of resources across firms is important for aggregate productivity (Hsieh and Klenow 2009; Restuccia and Rogerson 2017). This paper relates to a body of the literature that demonstrates how trade can alter the resource allocation across firms. Melitz (2003) shows in a seminal theoretical work how trade liberalization leads to a reallocation of labor from less productive to more productive firms when firms are heterogeneous in productivity. This gives rise to aggregate productivity and welfare gains through cross-firm resource reallocation and economies of scale (Melitz and Redding 2015). Empirical studies have provided ample evidence in support of the reallocation effects of trade liberalization and found substantial productivity growth from reallocation in contemporaneous data (for example, Pavcnik 2002; Trefler 2004). Trade can also alter the allocation across firms through pro-competitive effects (De Loecker and Warzynski 2012; Edmond, Midrigan, and Xu 2015). Di Giovanni, Levchenko, and Rancière (2011) present a model to show that exporters display systematically lower power law estimates and claim that power law estimates that do not account for international trade could be misleading. The predominant theories of intra-firm trade and heterogeneous firms ("New new trade theory") have been rarely

examined in the first wave of globalization. Huberman, Meissner, and Oosterlinck (2017) undertake an important step forward in examining theories of firm heterogeneity in this historical context. They study the implications of falling fixed and variable trade costs for the extensive and intensive margins in Belgium. This study contributes by moving beyond country-level data and using district-industry specific trade data to analyze and identify the implications of increases in trade and tariffs on the firm size distribution during the first wave of globalization. It adds to this strand of the literature by providing credible evidence for reallocation effects in line with the Melitz (2003) export selection and import competition channels in the historical context of the first wave of globalization.

Secondly, this study contributes to a large literature examining the effects of railroads on economic development. In influential work, Donaldson (2018) finds that railway access facilitated price convergence across regions in India and led to an increase in real income by 16 percent. A series of papers has shown that railways spurred population growth in connected cities (for example, Hornung 2015 for Prussia) and that initial shocks can lead to persistent differences in population over time (Jedwab and Moradi 2016; Berger and Enflo 2017). Fajgelbaum and Redding (2014) and Berger (2019) show that railroads facilitated structural transformation. Atack, Haines, and Margo (2011) use historical manufacturing census data from 1850 to 1870 and find that access to the railroad had a positive causal effect on the share of firms with factory status (establishments with 16 or more workers) at the US county level. However, their data do not allow them to study the distributional consequences of aggregate trade on the firm size distribution, as this study examines. This paper contributes to this strand of the literature by showing that increased exposure to railway trade coincided with a reallocation toward larger firms along the firm size distribution. It highlights increases in railway trade as specific mechanism for the growing importance of large firms over time.

Thirdly, this study relates to a large contemporaneous literature studying the effects of trade shocks on regional outcomes. In seminal work, Autor, Dorn, and Hanson (2013) use a shift share design to instrument US exposure to Chinese import competition with the exposure of other high-income countries. They find that an increase in import competition led to higher unemployment in regions more exposed to Chinese import competition. Using a similar methodology in the context of Germany, Dauth, Findeisen, and Suedekum (2014) report that in the aggregate Germany experienced employment gains from Eastern integration due to job growth in export orientated industries. Building on their methodology this study constructs a shift share design to analyze the consequences of trade for regional firm inequality. This paper contributes to this strand of the literature by using rich data allowing to apply a similar methodology in a historical context. It also contributes by studying firm and regional inequality in the first wave of globalization, whereas the country level data used predominantly in the economic history literature on globalization in the first wave of globalization did so far not allow doing so (O'Rourke 2019).

Finally, this paper makes a data contribution by merging datasets from the German Em-

pire. It relates to previous studies that have used them separately. Gutberlet (2014) examines the effect of mechanization on spatial concentration using the same industry census data as in this paper. Wolf (2009) introduces the bilateral railway data to the literature. He tackles the question, whether the German Empire was an integrated economy or internal borders still mattered. Wolf, Schulze, and Heinemeyer (2011) use the same data to examine the effects of border changes after Versailles. This paper expands on their collection effort by harmonizing these two datasets according to the same geographic boundaries and in time, and by expanding the set of industries. It introduces an extremely rich dataset that contains bilateral trade flows and firm size distributions by district-industry that is to the best of my knowledge unique in the historical context and can be of potential use to future research.

3 Historical Context

Before 1871, German-speaking countries were organized in a loose association of 39 states, the German Confederation. The German Empire was unified politically in 1871 following the Franco-Prussian war. The period this paper is considering falls into the first wave of globalization, a key period in the economic development of the German Empire.[1] Between 1882 and 1907, real GDP per capita in the German Empire grew at an annual rate of approximately 1.9 percent.[2] According to Maddison's Historical GDP database, the German Empire became the world's second-largest economy in terms of GDP by 1908 (Bolt and van Zanden 2014).

This impressive economic development coincided with a period of rapid external and internal trade integration. Most notable was the extension of transport infrastructure. The size of the railway network expanded by 63.4 percent from 34,382 km in 1882 to 56,191 km in 1907 (Kaiserliches Statistisches Amt 1884; 1909).[3] By this time, the network was the largest in Europe. Between 1882 and 1907, transportation on railways steadily rose from 105,000 tons to 295,000 tons, which implies an annualized growth rate of 4.6 percent.[4] The rapid expansion of trade volume is even more remarkable in comparison to France, which was the fourth-largest economy in the world in 1882. According to Sympher (1913), railway trade in the German Empire grew by 4.9 percent per year, which was significantly larger than France's 2.1 percent growth rate. The unique bilateral railway data used in this paper allow to analyze the role of increased external and internal integration within the unified German Empire.

Torp (2014) presents data on international trade integration of the German Empire. The share of exports in GDP almost doubled from 8.5 percent (1874-78) to 15.8 percent (1909-13).

[1] Chronological placements of the first wave of globalization vary in the literature. The end of the period is generally characterized as the year before the outbreak of World War I (1913). Jacks, Meissner, and Novy (2010) define the period as 1870-1913, i.e., their chronological placement includes 1875 to 1907. Hereafter, this study follows their chronological definition.

[2] Calculated from Maddison Historical GDP Database on December 12, 2015: `www.worldeconomics.com/Data/MadisonHistoricalGDP/MadisonHistoricalGDPData.efp`.

[3] One mile is approximately equal to 1.61 kilometers.

[4] One metric ton is equal to 1,000 kilograms. This paper refers to a metric ton as a ton in the text.

Likewise, the share of imports in GDP rose from 15.2 percent (1874-78) to 19.2 percent (1909-13).[5] Simultaneously, the share of exports originating from the German Empire in total world exports increased steadily from 1874-78 (9.5 percent) to 1909-1913 (12.2 percent). Figure A.1 puts the sum of exports and import shares in GDP - openness - in a long run perspective: the level of openness of the German Empire, which was attained by the end of the first wave of globalization in 1913, was only reached again 60 years later in 1973.

At the same time, the German Empire underwent structural change as it transformed from an agrarian economy into an industrial economy. In the 1882 census, the population employed in agriculture was 8.24 million, whereas 5.93 million employees were working in industry.[6] The industrial sector narrowed the gap by 1895 with approximately 8 million employees compared to 8.29 million people employed in agriculture. By 1907, the industrial sector had become the leading sector, with 11.45 million of the population employed, while the employment in agriculture was 9.88 million.

The "liberal period" of the German Empire from 1862 to 1879 (O'Rourke 2000) was characterized by free trade and low tariffs. For example, the German Empire signed a free trade agreement with France in 1871 that included a most favored nation clause. Internationally, the first wave of globalization, which began in the 1870s, was characterized by falling transportation costs (Jacks, Meissner, and Novy 2010). On the other hand, there was at the same time a backlash against globalization in the form of protectionist tariffs for certain industries. The German Empire became one of the more protectionist nations (Estevadeordal 1997).

When cheaper grain from Russia and the "New World" flooded German markets, agricultural landowners demanded protection in the second half of the 1870s (Hamerow, Duggan, et al. 2020). Gourevitch (1977) claims that Junkers, which were members of the landed nobility in Prussia, wanted to mitigate the impact of falling prices and foreign competition. Junkers that owned much of the land and assets east of the Elbe supported protection (Lehmann 2010). Lehmann (2010) argues that this was consistent with a Heckscher-Ohlin model, as capital was a scarce factor in the German Empire and its owners would have lost out from free trade. Tariffs were among the most important issues in the campaign for the election in 1878. The result of this election was political change, as Bismarck switched allies from the Liberals and collaborated with the conservative parties, who won the majority. They were in favor of protection, and on July 12, 1879, the parliament took the vote to implement the so called "iron and rye tariffs." As tariffs in the German Empire were set at the national level in Parliament, the movements can be regarded as plausibly exogenous for the 37 districts considered in the analysis later on that did not have the capability to implement tariff changes on their own. The districts took the changes in national tariff rates as given. The analysis will be based on tariff movements

[5] Torp (2014) calculates the shares of exports and imports in gross national product using constant 1913 prices.

[6] Employment in agriculture includes the subsectors market garden, livestock breeding, forestry, and fishery. The employment figures for agriculture are taken from Hohorst, Kocka, and Ritter (1978). Employment figures in industry are taken from section B of the industry census data, which comprises industry including the construction and mining industries. The employment figures include self-employment.

between 1875 and 1907, including the protectionist tariff movements in 1879.

4 Data

This section describes the three main data sources used in the empirical analysis: industry census data, railway trade data, and tariff data. This study focuses on 18 industries from the mining and industrial sector in the industry census data for which consistent and matchable data were also available for goods trade on railway and for tariffs.

4.A Industry Census Data

Data on firm size distributions was obtained from industry census data, which were published as part of the series "Statistik des deutschen Reichs". Data are available at the administrative level for 37 regions and states, which include among others Prussian provinces, kingdoms such as Bavaria, and grand-duchies such as Baden. The data stem from the industry censuses in the German Empire for which consistent data are accessible, those in 1875, 1882, 1895, and 1907. The first industry census in 1875 was collected in December. The data of the following three censuses capture employment in the first half of June across all years. Gutberlet (2014) uses data from the same census publications to examine the effects of mechanization on spatial concentration in the German Empire.

Census data capture the complete picture of employment in industry in the German Empire. The industry classification contains three levels across all census years. In 1882, there were 20 industry groups, 96 industry classes, and 248 industries at the finest level of aggregation. By 1907, the detail of aggregation changed to 23 industry groups, 129 industry classes, and 396 industries. Using tables provided in the statistics, it is possible to compare the industry classifications consistently over the years. In the following, this study uses the finest level of industry disaggregation as baseline classification in the analysis.

Industry census data were collected in conjunction with occupation census data.[7] Different establishments of one firm in different locations were counted as different units in the census, i.e., the data measure firm size at the establishment level. Each firm was classified by industry, and in case its activity could fit into more than one category, the company was assigned to the industry that corresponded to its core industry classification. If a company was active in multiple industries, it was also possible to split it into subdivisions and count these as separate establishments. For the industries under consideration, only 2.7 percent (1895) and 9.7 percent (1907) of the companies were active in multiple industries, as measured at the industry group level. Companies producing in the same industries with multiple locations accounted for only approximately 1 percent of all establishments. Hence, firms producing in multiple industries (either in the same location or in different locations) or producing in the same industry in

[7] Small firms with fewer than six employees were counted with the occupation census. All other firms, which employed more than five full-time workers, were asked to answer a separate firm survey.

multiple locations were not common, and this paper continues hereafter to refer to firms for counts in the census data instead of establishments.

Companies were located according to industry, trade, tax, and other registers. The historical data do not set any cutoff for counting, for example, a minimum number of employees to be counted, i.e., all firms irrespective of size were counted. Thus, this study has accessed industry census data, where participation was mandatory, and the population of firms was captured. The Online Appendix contains details on the counting procedure and information on the merit and quality of the German industry census data in comparison to industry census data in the UK and the US.

This paper matches industries to trade data and tariffs where the concordance between the three data sources is unambiguous. In so doing, the descriptions in each data set and industry classification are matched as narrowly as possible to the finest level of industry classification.[8] Using these criteria to merge the three data sources, the analysis considers data from 18 industries that can be mapped onto corresponding industries of the industrial sector in the census data. The industrial sector corresponds to sections B (mining and quarrying) and C (manufacturing) of the concurrent NACE 2 classification. By 1882, the 18 industries under consideration in the empirical analysis comprised 1.39 million employees and 167,765 firms, and by 1907, 2.89 million employees and 195,515 firms were counted. In 1907, the industries under consideration accounted for about 27.6 percent of total employment in the industrial sector counted in the census data. This study focuses on the count of full-time employment as opposed to secondary employment. All firms with at least one person working full-time are taken into consideration.[9] Geographically, the districts with the highest share of large firms with more than 50 employees (above 7 percent) were located in the west of the German Empire (districts 11, 12, and 17), while the districts with a low share of large firms (below 2 percent) were located in the east and north of the German Empire (districts 1, 2, 3, and 7).[10]

Overall, there are five bins of size for the number of firms in the data (1-5, 6-10, 11-50, 51-200, and more than 200 employees). As in the main analysis, the five size categories are grouped into two: firms with below and above 50 employees.

[Figure 1 here]

Figure 1 shows the evolution of the firm share in two size categories in the four census years. Table 1 displays the corresponding shares. The share of firms with more than 50 employees increased from 0.88 percent in 1875 to 4.53 percent in 1907, whereas the share of the other size category with up to 50 employees decreased within the 32-year period. Column 2 of Table

[8]The detail in the classification of industries is generally similar to three-digit or four-digit SIC industries. See Online Appendix for a list of industries considered.

[9]Self-employment from 1882 is excluded, as data were not fully collected in 1882. Self-employment is not separable in the 1875 census. The results are robust to excluding 1875 or adjusting the firm size distribution by excluding small firms.

[10]See Online Appendix for the definition of districts.

1 shows the share of firms with more than 50 employees in the industrial sector, which is the section in the industry census data to which all 18 industries belong. The share of firms with more than 50 employees increased from 0.36 percent in 1875 to 3.04 percent in 1907. The change between 1895 and 1907 in the matched sample was 1.29 percent and similar to the change in the industrial sector which was 1.06 percent. The share of firms with more than 50 employees in the selected sample was 4.53 percent in 1907, whereas the share of firms with more than 50 employees was 3.04 percent in the industrial sector. Hence, the industries under consideration tend to have a thicker right tail compared to the industries not covered in this paper. However, not all of the industries of the industrial sector that could not be matched unambiguously to trade and tariff data are tradable industries.

4.B Trade Data

In this paper, trade data from the series "Statistik der Güterbewegung auf Deutschen Eisenbahnen" were collected for 22 good categories that are among the most significant in terms of volume and that one can assign to 18 industries in the census data.[11] The trade volume of the 22 good categories accounted for 60.5 percent of total recorded trade volume on railways in 1907. Two examples of good categories are iron ore and chemistry products.[12] Bilateral trade data were collected for the transport mode, railways, by which goods were predominantly traded in the period under consideration.[13]

This study uses hand digitized 402,408 raw data points and aggregated them into 43,146 bilateral trade flows. The data comprise all inland and international railway trade.[14] Despite the fact that the railway network expanded by more than 70 percent between 1882 and 1907, there was an upswing in trade volume per mile of the network of 146 percent for the goods under consideration. Local land transport was too costly for longer distances compared to railways. Gutberlet (2012) reports estimates that suggest road transportation was nearly 40 times as expensive as waterways and more than 10 times as expensive as railways.

Data for trade on railways are available from 1883, so these data are taken to calculate predetermined trade shares in the IV calculation. All trade data are measured as quantities in tons. The railway trade data are structured according to 36 trade districts. These districts follow closely the administrative states and regions according to which the industry census data are classified. This paper merges districts, states, and regions in the industry census and railway data to construct 17 districts for both types of data, which contain the same geographic

[11] In total, 70 good categories were recorded in the railway trade data.

[12] See Online Appendix for a full list of good categories included.

[13] Bilateral data for trade on waterways is available from 1913. Statistics for land transportation were not collected, because land transportation (for example, horse-drawn vehicles) was only locally important. The automobile industry was still in its infancy between 1882 and 1907. For example, the van with an engine was invented only in 1896.

[14] Conservative estimates suggest a maximum of 1.1 percent of railway network length not covered in the statistics for 1907 and a smaller proportion for 1883 and 1895. The lines not covered were mostly short and relevant to local traffic.

boundaries, which is a unique characteristic of the data. Hence, I can construct region-specific firm size distributions and bilateral trade flows, which is to the best of my knowledge only feasible in the German Empire in the historical context.[15]

Railway data were collected for final shipments to the final destination, i.e., there is no double-counting when goods passed through other districts to their final destination. The data were collected quarterly until 1895 and annually thereafter. They were published as a separate statistical series titled "Statistik der Güterbewegung auf Deutschen Eisenbahnen." Figure A.3 contains an example of the railway trade data and illustrates that the rich data record bilateral trade flows from the origin to the final destination. Origin and final destination of trade flows were determined through rail waybills. The data from the railway statistics were introduced before in Wolf (2009) and in Wolf, Schulze, and Heinemeyer (2011).

The bilateral trade data are classified according to four modes: domestic exports and imports, and foreign exports and imports. The shares of each trade mode remain stable across all three years. Exports and imports within the German Empire constituted about 43 percent for the median industry of trade on railway, while the share of international exports accounted for 5.2 percent and international imports for 6.5 percent of railway trade in 1907 for the median industry. In Table A.4, the growth rate across industries is calculated. The median growth rate in aggregate trade between 1895 and 1907 was 129 percent. The lowest growth rate was 20 percent in the wool industry and the highest growth rate was 333 percent in chemistry products, which suggests substantial cross industry heterogeneity in growth rates. In Table A.5, the extensive margin is analyzed across the 18 industries between 1883 and 1907. The average district-industry pair added 2.90 domestic destinations (24.5 percent growth rate) and sources, 3.13 foreign destinations (62.1 percent growth rate) and 3.18 foreign sources (84.4 percent growth rate) between 1883 and 1907. The numbers are very similar for the median district-industry pair. While overall growth in trade was driven by the intensive margin, these numbers highlight a role for the extensive margin of trade.

4.C Tariff data

The data were collected from foreign trade statistics, which were published as part of the series "Statistik des deutschen Reichs", on specific tariffs levied on goods for each industry. The classification of the tariff data is similar to the one of the industry census data. Hence, tariff and industry census data were matched by industry. The tariffs are denoted in the statistics as tariff in Mark per 100 kg. The average effective rate is calculated by taking the ratio of the tariff revenue and the total value of aggregate imports for each industry.

[15] Figure A.2 shows a map of the 17 districts, and the following pages contain a definition of the constructed districts. Construction of districts is dictated by the spatial units of industry census data and trade statistics. What is more, this study constructs districts such that most industries are represented in all districts. Table A.2 contains the aggregate employment in the 18 industries for each district by year and Table A.3 shows the total population for each district by year.

[Table 2 here]

Table 2 shows that the German Empire increased protection between 1875 and 1882. There was an increase in tariffs in two-thirds of the industries, and consequently, the average tariff (median tariff) rose significantly by 5.52 percent (5.39 percent). The average tariff increased further between 1882 and 1895; however, the median tariff fell. Overall, there were tariff increases (reductions) in eight (six) industries between 1882 and 1895. Finally, between 1895 and 1907, the average tariff rate decreased, whereas the median tariff rate increased slightly. In this time period, there was a significant decrease in tariff rates in iron and steel, iron and steel products, and petroleum and other mineral oils.

5 The Effects of Aggregate Trade and Tariffs on the Firm Size Distribution

5.A Identification from Aggregate Trade

For the analysis of bilateral railway trade flow data, this paper harmonizes the data according to 17 districts.[16] Let "i" denote the sub-index for industry, "j" for district, and "t" for time. Furthermore, "m" denotes the exporting district, and "n" denotes the importing district in the bilateral trade flow data.

Bernard, Jensen, and Schott (2009) document stylized facts about importing firms in US data (1993-2000). According to their analysis, more than 50 percent of firms that import also export, and these firms account for the majority of trade. This study considers the effects of both selection through exporting and import competition on the size distribution of firms. It groups exports and imports together into aggregate trade in the analysis, because the baseline estimation implied that there was no statistically distinguishable effect between the two.

Hence, the aggregate trade of industry "i" in district "j" is defined as the following summation of all bilateral trade flows:

$$\text{aggregate trade}_{ijt} = \sum_{\substack{n=1 \\ m=j \\ n\neq j}}^{N} \sum_{\substack{m=1 \\ n=j \\ m\neq j}}^{N} \text{x}_{mnit}. \tag{1}$$

The first sum captures the exports (both intra- and international) of district j to all other 35 districts, while the second sum captures the imports (both intra- and international) of district j to all other 35 districts.[17]

The first outcome is the share of firms with more than 50 employees, which is formally defined as

$$\text{firmshare}_{ijt} = \frac{\text{Number of firms with >50 employees}_{ijt}}{\text{Number of firms}_{ijt}}. \tag{2}$$

[16] See Online Appendix for a definition of the districts.

[17] The districts are defined in the Online Appendix.

Figure 2 illustrates the construction of the outcome variable.[18] This paper examines the reallocation toward the right tail of the firm size distribution. The counts are consolidated into firms with above and below or equal to 50 employees such that the distribution becomes binary.

The following specification relates the firm share to aggregate trade:

$$\begin{aligned} \text{firmshare}_{ijt} = \mu_{ij} + \delta_{jt} + \text{product group}_{it} + \gamma \times \text{firmshare}_{ij,\text{pre-trend}} \\ + \beta \times \ln(\text{aggregate trade}_{ijt}) + \varepsilon_{ijt}. \end{aligned} \tag{3}$$

It controls for local shocks through district-time fixed effects, δ_{jt}, for time-constant heterogeneity through district-industry fixed effects, μ_{ij}, and for technology shocks through product-group-time fixed effects for five product groups, product group$_{it}$. One example of a product group includes coal mining, metallurgy, and iron and steel products.[19] Additionally, the specification includes the pre-trend in the dependent variable to ease comparison with the IV estimation.

Naturally, there are concerns of reverse causality - the firm size distribution affecting trade and not the other way around, as implicitly assumed in the presentation of the OLS regression analysis. For example, district-specific demand shocks can influence the shape of the firm size distribution, which in turn can increase firms' productivity and subsequently trade flows of a district. Alternatively, simultaneity may be a concern, i.e., trade flows and the size distribution are jointly influenced by a third *unobserved* factor, such as productivity. The Second Industrial Revolution broad significant technological innovations that could have led to industry wide productivity shocks. Exports and firm size increase with productivity. This would suggest that the estimates reflect both endogenous export supply due to higher productivity and exogenous demand stemming from imports, biasing the estimates upward. In the case of imports, simultaneity bias arises, as endogenous import demand and exogenous export supply coincide in equilibrium.

To alleviate such concerns, this study proposes an instrumental variable approach, the shift-share approach, which was popularized in the context of migration (Card 2001).[20] Variations of this instrument have become popular when investigating the impact of the China shock (Autor, Dorn, and Hanson 2013) or more generally to instrument for imports and exports to isolate demand and supply shocks, respectively, for example, as in Hummels et al. (2014) and Berthou et al. (2019).

To create exogenous variation in export supply in the case of imports and in import demand in the case of exports for a given district j, the predicted trade flow uses the initial share of bilateral imports (exports). This share is multiplied with the level imports (exports) from (to) all other districts of the respective trading partner while excluding district j. The underlying idea

[18] Throughout the analysis, this paper refers to 1 percent as 0.01 units in the share of firms, as for example defined in equation (2).

[19] The remaining product groups are defined in the Online Appendix.

[20] In the literature, this approach is also known as the "Bartik" instrument following Bartik (1991).

is that a district "j" that initially imports a large fraction from trading partners that expand their export supply to all other districts excluding district "j" will face higher exposure to exogenous increases in export supply. The same idea applies to increases in exogenous import demand.

Autor, Dorn, and Hanson (2013) instrument for Chinese exports to the US with the exports to other advanced countries. Adapting Autor, Dorn, and Hanson (2013) to this setting, one would need the industry-level exports and imports of other countries, such as France and Italy, to other high-income countries. To the best of my knowledge, these data are not readily available. Given that the bilateral data are only available for regions within the German Empire, spatial dependencies are more of a concern relative to the setup in Autor, Dorn, and Hanson (2013). A robustness check demonstrates that the standard errors are actually lower when using standard errors to account for spatial dependencies following Conley (1999).

The IV strategy assumes that the initial shares are exogenous from future shocks in supply and demand (Goldsmith-Pinkham, Sorkin, and Swift 2020). The inclusion of product-group-time fixed effects amounts to the assumption that within a product group unobservable industry level shocks are orthogonal to the instrument. The initial shares vary by district within a product group, and hence, different districts are differently exposed to exogenous changes in aggregate trade. A recent literature has raised concerns using Bartik instruments (Adão, Kolesár, and Morales 2019; Borusyak, Hull, and Jaravel 2018; Goldsmith-Pinkham, Sorkin, and Swift 2020). In this context, the concerns relate to the exogeneity of the initial trade shares that could correlate with subsequent changes in industrialization and the size distribution of firms. Larger trade shares with certain trading partners could be due to unobserved endogenous characteristics, which would invalidate the Bartik approach. To address these concerns, the IV estimation controls for pre-trend of the dependent variable, i.e. the change in the firm size distribution between 1875 and 1882.

Formally, the instrument for observed bilateral trade is given by the following predicted bilateral trade flows:

$$\widehat{\mathrm{x}_{mnit}} = \frac{\mathrm{x}_{mni,0}}{\sum_{\substack{n=1\\m=j\\n\neq j}}^{N} \sum_{\substack{m=1\\n=j\\m\neq j}}^{N} \mathrm{x}_{mni,0}} \times x_{mni-j,t}\,. \tag{4}$$

The leave one out trade flows are defined as:

$$x_{mni-j,t} = \sum_{\substack{n=1\\m\neq j\\n\neq j}}^{N} x_{mnit}, \quad \text{if m=j and j is an exporter,}$$

$$x_{mni-j,t} = \sum_{\substack{m=1\\n\neq j\\m\neq j}}^{N} x_{mnit}, \quad \text{if n=j and j is an importer,}$$

where $t \in \{1895, 1907\}$ and the base period 0 is 1883. Note that exports (imports) exclude district j and hence they are not affected by endogenous export supply (import demand). The instrument of aggregate trade is then defined as the summation of predicted bilateral trade

flows:

$$\widehat{\text{aggregate trade}}_{ijt} = \sum_{\substack{n=1 \\ m=j \\ n\neq j}}^{N} \sum_{\substack{m=1 \\ n=j \\ m\neq j}}^{N} \widehat{x_{mnit}}. \tag{5}$$

The first stage regression is given by

$$\begin{aligned} \ln(\text{aggregate trade}_{ijt}) = \mu_{ij} + \delta_{jt} + \text{product group}_{it} + \gamma \times \text{firmshare}_{ij,\text{pre-trend}} \\ + \beta^{IV} \times \ln(\widehat{\text{aggregate trade}}_{ijt}) + u_{ijt}. \end{aligned} \tag{6}$$

[Table 3 here]

All regressions are estimated in the matched industry census and railway data set for 1895 and 1907. Due to the two-period setting, the fixed effect estimate coincides with the first difference estimate. Standard errors are clustered at the district-industry pair level to account for serial correlation over time.

Panel A of Table 3 shows that the instrument passes conventional levels of the Kleibergen-Paap F-statistic and LM-statistic and has predictive power. A 1 percent increase in predicted trade leads to a 0.51 percent increase in actual trade. Panel B of Table 3 presents the second stage results. The point estimate for the IV estimation is positive and significant at the 10 percent level. This is consistent with heterogeneous firm models in which smaller firms lose market share in favor of larger firms due to increased aggregate trade. The IV estimate is more than double in size compared to the OLS estimates presented in Panel C, suggesting a downward bias of the OLS estimates.

The OLS estimate (IV estimate) suggests that the elasticity of the share of large firms with respect to a one percent change in aggregate trade is 0.000378 (0.000875). The interquartile range in aggregate trade growth between 1895 and 1907 across all district-industry pairs included in the estimation was 47 log points. Multiplying the elasticity estimate and the interquartile range implies, for the OLS point estimate, an increase in the share of large firms by 1.79 percent for district-industry pairs in the 75th percentile relative to district-industry pairs in the 25th percentile of the growth distribution. A likewise calculation for the IV point estimate suggests an even larger increase in the share of large firms by 4.15 percent for district pairs in the 75th percentile relative to the 25th percentile of the growth distribution. These two ceteris paribus calculations highlight that the reallocation forces due to trade along the firm size distributions were quantitatively important.

As an alternative outcome, this study considers the alpha of a Pareto distribution. The empirical CDF of the Pareto distribution with minimum value equal to one is given by:

$$F_X(x) = 1 - (\frac{1}{x})^{\alpha}. \tag{7}$$

Hence, the estimate for alpha evaluating the CDF at x equal to 5, 10, 50, and 200 is determined

by:

$$\widehat{\alpha} = \frac{-\ln(1 - F_X(x))}{\ln(x)}. \tag{8}$$

The median estimate for alpha across the four points of the CDF is specified as outcome to mitigate the impact of extreme estimates.

$$\begin{aligned} \widehat{\alpha_{ijt}} = \mu_{ij} + \delta_{jt} + \text{product group}_{it} + \gamma \times \alpha_{ij,pre-trend} \\ + \beta \times \ln(\text{aggregate trade}_{ijt}) + \varepsilon_{ijt}. \end{aligned} \tag{9}$$

[Table 4 here]

The OLS estimate in column (1) is negative and statistically significant at the one percent level. This finding implies that the CDF shifts downward for each x such that the right tail becomes thicker and cross-firm resource reallocation toward the right tail of the firm size distribution. The IV estimate is significant at the one percent level and double the size of the OLS estimate. The point estimate suggests that a one percent increase in aggregate trade decreases the alpha of the Pareto distribution by 0.002337 units. The interquartile range in aggregate trade between 1895 and 1907 was 45.7 log points. This implies that the increase in aggregate trade for the 75th percentile relative to the 25th percentile of the growth distribution reduced the Pareto alpha by approximately 0.107 units. The finding is in line with Di Giovanni, Levchenko, and Rancière (2011) who report a lower power law estimate for French exporters.

What is causing the shift in the firm size distribution? It could be that firms exit, large firms enter, or that more firms grow above 50 employees such that they belong to the largest size category.

[Table 5 here]

In Table 5, the IV estimation described before is run using the logarithm of the firm count by size category as outcome.[21] The point estimate in column (1) is negative, but insignificant. This suggests that some exit of smaller firms happened with up to 50 employees, but the effect is not statistically significant. The point estimate in column (2) is positive and significant at the 10 percent level. It suggests that a one percent increase in aggregate trade raises the firm count of large firms with more than 50 employees by 0.35 log points.

Overall, the evidence shows that the share of large firms increases through the extensive margin, i.e. more firms entering with more than 50 employees or firms growing above 50 employees. However, firms within this size category do not grow at the intensive margin.[22]

[21] Due to entry and exit of few district-industry-class-size observations, one firm is added to the outcome variables.

[22] I also examined the intensive margin of average firm size within firms above 50 employees at a higher industry aggregation. In unreported results, the average firm size of large firms increased only slightly from 155 employees to 158 employees between 1895 and 1907 and this was unrelated to changes in aggregate trade.

5.B Identification from Tariffs

Tariff changes that were set at the national level are used as an alternative source of identification. As tariffs were determined in the national parliament, the movements can be regarded as exogenous for the 37 individual districts considered in the analysis. No individual district had the capability to enforce tariffs on its own. For this exercise, 37 districts instead of 17 districts are defined, as there is no need to harmonize districts as in the case of the matched industry census and railway trade data.[23] The following specification relates changes in the share of firms to changes in tariffs.

$$\Delta \text{firmshare}_{ijt} = \delta_{jt} + \beta \times \Delta \text{tariff}_{it} + \varepsilon_{ijt}. \tag{10}$$

The structure of equation (10) is similar to equation (3).[24] However, equation (10) does not contain product group fixed effects, as the variation of tariffs is at the industry level. Furthermore, the pre-trend of the dependent variable is not included as control variable, because tariff changes at the national level are exogenous to pre-determined changes in firm size at the district level. The outcome is the share of firms with more than 200 employees. Since imports from abroad only affect a small portion of the firm population, only the largest firms are affected when it comes to reallocation to the right tail of the firm size distribution. For this exercise, the specification uses data from all four census years in 1875, 1882, 1895, and 1907. In particular, the 1875 census captures the firm size distribution before Bismarck's protectionist tariffs were implemented in 1879. The census in 1875 was counted in December in contrast to other census years. Hence, the results could be affected by seasonality. The specification controls for district-time fixed effects, δ_{jt}, to account for this.[25] Standard errors are clustered at the district-industry pair level to account for serial correlation over time. A positive point estimate indicates a movement with tariff changes, and a negative point estimate indicates a movement against tariff changes. How do average tariffs and the share of firms comove in the aggregate at the district level? In other words, do districts that have a large share of employment in industries with tariff cuts experience a relative increase in the share of a large firms with more than 200 employees? To answer this question, a weighted tariff is defined as change at the district level:

$$\Delta \text{tariff}_{jt} = \sum_i \text{industry employment share}_{ijt} \times \Delta \text{tariff}_{it}, \tag{11}$$

where the industry employment share is defined as employment in industry i in district j at time t over employment in district j at time t. The corresponding regression at the district level

[23] See Online Appendix for a definition of the 37 districts.

[24] Since equation (3) is estimated using two periods, the estimations in first differences and in levels using fixed effects produce the same result.

[25] In unreported results, the findings are robust to restricting the sample to the last three census years.

additionally controls for a macro time trend, λ_t, and becomes:

$$\Delta \text{firmshare}_{jt} = \lambda_t + \beta \times \Delta \text{tariff}_{jt} + \varepsilon_{jt}. \tag{12}$$

[Table 6 here]

The results in column (1) in Table 6 at the industry level are consistent with reallocation along the firm size distribution. The negative point estimate suggests that there is an increase of the very largest firms in response to tariff reductions. The point estimate is significant at the 1 percent level. A 1 percent reduction in tariffs coincides with a 0.11 percent increase in the share of the very largest firms. Table 2 shows an increase in the median tariff across industries of 5.29 percent between 1875 and 1882. In the following, the consequences of the changes in tariffs are compared to a scenario of zero changes in tariffs. Holding all else equal, this implies a 0.57 percent decrease in the share of very large firms due to protectionist trade policy, which corresponds to approximately 19 percent in terms of the share of very large firms in 1875. The average share of very large firms in the iron and steel industry was 11.43 percent in 1875. The point estimate suggests that the increase in tariffs from 0 to 16.12 percent resulted in a 15.1 percent decrease of the share of very large firms relative to the 1875 level, or 1.72 percent in absolute terms. Conversely, the decrease in tariffs between 1882 and 1895 caused an increase in the share of very large firms. The median tariff decreased by 3.52 percent during this time period. This implies a 0.37 percent increase in the share of very large firms, which corresponds to a 10 percent increase relative to the 1882 level of very large firms. The chemistry industry experienced a significant liberalization with a reduction of tariffs by 6.06 percent. This suggests that an increase in the share of very large firms by 0.65 percent can be attributed to the decrease in import tariffs. The results at the district level are shown in column (2). The coefficient is negative and significant at the one percent level. The coefficient estimate suggests that a 10 percent decrease in import tariffs, raised the share of firms with more than 200 employees by 1.55 percent. The magnitude is similar to the industry level estimate in column (1) suggesting small cross-sector reallocation forces. The standard deviation of changes in tariffs at the district level is 3.15 percent. This suggests that a one-standard-deviation decrease in tariffs would correspond to an increase of 0.49 percent or, put differently, a 34.47 percent increase in terms of the average share of very large firms in 1875 at the district level, which was 1.41 percent.

5.C Robustness Aggregate Trade

This section presents a number of robustness checks.

[Table A.6 here]

One concern is that industry-level shocks are driving the results. The IV strategy does not allow for the inclusion of industry-time specific fixed-effects as the cross-sectional sample size

of seventeen districts is too small. In Table A.6, sector-time trends are included instead of product-group-time fixed effects. In column (1), the outcome is the share of firms with more than 50 employees. Compared to Table 3, the coefficient remains positive but becomes smaller and insignificant. However, in column (2) the outcome is the share of firms with more than 200 employees and the coefficient remains significant at the 10 percent level after the inclusion of industry-time fixed effects. The results imply that reallocation effects toward the very largest firms continue to hold after the inclusion of industry-time fixed effects.

[Table A.7 here]

In Table A.7, equation (3) is estimated using a beta regression to take into account the boundedness of the dependent variable. Following Smithson and Verkuilen (2006), the dependent variable is transformed such that it lies between 0 and 1.[26] The point estimate is positive and significant at five percent, which suggests that the results are not driven by the linear probability model that can potentially under-predict values below zero or over-predict above one.

[Table A.8 here]

In Table A.8, heterogeneity across product groups is examined by estimating separate slope coefficients for each product group. Product group 1 includes coal mining, metallurgy, and iron and steel products. Firms in this product group are by nature the largest and hence the share of large firms is quite unresponsive to changes in aggregate trade. Product group 2 comprises stone and earth products. The point estimate is large and significant at the five percent level. The third product group includes chemistry products, fat and oils, petroleum and other mineral oils. The point estimate is positive but marginally insignificant. The fourth product group comprises cotton and wool. The point estimate is positive but small and insignificant. Lastly, the fifth product group contains paper, leather, and timber. The point estimate is positive and statistically highly significant at the one percent level.

[Table A.9 here]

The results in Table A.9 rule out a number of alternative explanations for the positive relationship between railway trade and the increase the share of large firms. The specification in column (1) addresses the concern that bilateral waterway trade data are not available and hence cannot be included in the estimation. Construction of railway or waterway may have diverted trade from one transport mode to another. Hence, growth in railway trade may be due to a change in transport mode rather than an increase in actual trade. Waterway trade is available in the aggregate at the district-industry level but not bilaterally. The specification in column (1) adds the share of railway trade of trade as portion of railway and waterway trade to the baseline controls. Rail was the predominant means of transport during the sample period. On

[26] Formally, the new variable is given by: $\widehat{\text{firmshare}}_{ijt} = \text{firmshare}_{ijt} \times \frac{(612-1)+0.5}{612}$.

average, 89 percent of goods were shipped via railways in 1883 in the sample. The 25th and 50th percentiles of the railway's share are 85.5 percent and 98.4 percent, respectively.

The coefficient in column (1) becomes somewhat smaller but remains significant at the 10 percent level. This suggests that the results are not driven by districts where there is a change in the relative importance of railway, and hence, the conclusions appear to hold by and large irrespective of the importance of railway trade for a specific district-industry pair. The railway share correlates negatively with the share of large firms.

Column (2) tests for heterogeneity by districts with sea access, for which trade via railway may be less significant than for landlocked districts. For seaports, the railway trade data also capture trade of goods transferred to most seaports in the German Empire via railway. For the set of landlocked districts, the data therefore capture most trade apart from local land transportation and waterway transport. The baseline coefficient remains significant at 10 percent. The interaction term is positive, but small and insignificant. This finding corroborates that the results are not driven by using railway trade only.

Next, the coefficients in column (3) distinguish industries by whether they produce intermediate products or not based on the goods' description in the trade statistics. My classification of intermediate producing industries comprises four industries: brown coal, hard coal, iron ore, and petroleum and other mineral oils. Intermediate producers generally have low values on the diagonal of the input-output matrix. The idea underlying this robustness check is to examine heterogeneity by industries with a low value on the diagonal of the input-output matrix in the estimation. The baseline coefficient remains significant at 10 percent and increases marginally. The interaction term is positive and large with 0.05, but insignificant. This implies that intermediate producing industries do not drive the results.

Column (4) tests for heterogeneity by the three largest coal producing districts, which include the Province of Rhineland, the Province of Schlesia, and the Province of Westphalia, that were responsible for 73.4 percent of the total production in 1895. Kopsidis and Bromley (2016) argue that coal was an important factor in Germany in explaining regional development after the introduction of the railway. Gutberlet (2014) finds that the adoption of coal-fired steam powered machinery led to the concentration of manufacturers in industry clusters near coal mines between 1875 and 1895, but not between 1895 and 1907. However, increased aggregate trade could interact of firms' co-location with coal mines in two ways: it may lead to a concentration of firms to regions with coal, or it may lower the transport costs of coal enough to enable coal-intensive firms to be located elsewhere. The baseline coefficient remains positive and significant, whereas the interaction term is small, negative and insignificant. The findings imply that clustering around coal mines cannot explain the result.

In column (5), an alternative explanation the spatial reallocation of firms into large cities as further explanation. Gaubert (2018) shows that larger and more productive firms sort into cities. Increases in aggregate trade could interact with the reallocation of firms into cities. In order to examine this channel, heterogeneity by the four districts with the largest share of the population

in cities in 1895 is tested.[27] The baseline coefficient is positive and significant at the 10 percent level. The interaction term is small, negative and insignificant. Hence, it is unlikely that the results are driven by reallocation of firms into more productive locations within districts.

Since the analysis makes use of bilateral trade flows of regions within the German Empire, spatial dependencies between regions may be a reason for concern. The estimation in column (6) uses HAC consistent standard errors that allow for spatial dependencies following Conley (1999). Compared to the baseline estimation with robust standard errors, the standard errors decrease, and the significance level increases to one percent. This implies that the analysis is robust to considering spatial dependencies in the calculation of standard errors.

5.D Robustness Tariffs

In this subsection, several additional analyses are carried out to investigate alternative explanations (agglomeration, migration, lobbying) and heterogeneity with respect to the import share is examined.

[Table A.10 here]

Combes et al. (2012) study the relative importance of agglomeration and selection toward higher productivity in a larger city. Their structural approach allows them to empirically distinguish between the two mechanisms. While selection left-truncates firms' productivity distribution, agglomeration right-shifts and dilates it. In the case of agglomeration, all firms enjoy the benefits of being located in a large city. Combes et al. (2012) find stronger evidence for agglomeration economies than for firm selection. To explore heterogeneity across districts to investigate whether there are effects similar to those of Combes et al. (2012), a measure of agglomeration is defined. In so doing, the measure of agglomeration is defined as the population density $\frac{\text{Population}}{\text{Area in km}^2}$ as it is common in the literature. The median density by census year is calculated. Then, the heterogeneity category is defined as all districts that were above median in at least two of the four census years as agglomerated. The results in column (1) show that there is no significant interaction coefficient between agglomeration and tariffs. The baseline coefficient remains significant at the 10 percent level. The interaction term is positive, but insignificant.

Alternatively, migration may play a role. There were significant internal migration flows as a consequence of the unification and economic differences across districts. Additionally, there was a surge in emigration to the United States. There was also immigration from elsewhere in Europe. Districts experiencing an increase in competition from lower tariffs may also be districts that have an outflow of migrants. As the migration rate is difficult to construct and requires detailed data, population growth between two census years is used as a proxy for changes in the labor supply. Heterogeneity is investigated with respect to population growth

[27] The districts are 4, 6, 9, and 12 as defined in the Online Appendix.

above median in this time span. The results are presented in column (2) of Table A.10. The overall effect for the share of firms with more than 200 employees remains significant at the 5 percent level irrespective of whether the district is characterized by high population growth or not.

A further concern is that political economy considerations may render the tariffs endogenous. Tariffs were supported by the Junker elite (Lehmann 2010), which chancellor Bismarck was part of. The Junker elite owned much of the land and industries located east of the Elbe. Hence, the heterogeneity category is a dummy equal to one for districts east of the Elbe for which the Junker elite had a monopoly to capture political connectedness. The results are presented in column (3). The interaction term is positive and the total effect for districts dominated by the Junker elite is not statistically significant. This suggests that the reallocation effect of tariffs was actually weaker for districts dominated by the politically connected Junker elite.

Finally, heterogeneity by the propensity to import across industries is tested. If an industry has a low share of imports from abroad, then changes in tariffs may have less pronounced effects compared to an industry that imports a large fraction of its total trade from abroad. This hypothesis is investigated by defining a dummy equal to one if the import share from abroad via railway between 1883 and 1907 was below the median across industries. The median share of imports via railways was 6.5 percent, and the mean share, 8.3 percent. Indeed, the results in column (4) point out that the reallocation effects are driven by industries with an import share above median. The baseline effect for industries is significant at the 1 percent level (p-value 0.002) in industries with an import share above median. The interaction term is positive and marginally insignificant. This suggests that changes in tariffs mattered only in industries with a high propensity to import from abroad. Tariffs did not affect sectors that were less exposed to import competition.

6 Conclusion

To the best of my knowledge this is the first study that uses district-industry specific data to illustrate the importance of aggregate trade in explaining observed changes in firm distributions across different size categories. For this purpose, this study used census data from a period of external and internal integration in the German Empire, a unique setting characterized by closer trade integration through the expansion of the transportation infrastructure and the reduction of transportation costs. At the same time, there was significant variation in trade costs across industries due to the introduction of protectionist tariffs. This paper emphasizes the role of economic integration in shaping the industrial employment structure during the industrialization of the German Empire.

The first finding of this paper is to show that the firm size distribution shifts to larger firms in response to increases in aggregate trade. This study leverages the regional dimension of the data by proposing a Bartik-type instrument for identification. Furthermore, this paper takes

advantage of the regional dimension of the data by arguing that tariff movements, which were determined at the national level, were exogenous to individual districts. The second finding is that exogenous decreases in tariffs caused an increase in the share of the largest firms. On the other hand, increases in tariffs between 1875 and 1882 led to a decrease in the share of the largest firms.

This paper assesses key implications of the literature on heterogeneous firm models using historical data. The analysis confirms, in an unparalleled historical setting, that the effects theoretically described by Melitz (2003) have also been present during the first wave of globalization - a century before the emergence of this pioneering literature. The past three decades have seen an unprecedented increase of imports from non-advanced countries into advanced economies, which has led to a rise in regional inequality between firms. This study is the first to use regional data to show that regional distributive effects of trade between firms already took place during the first wave of globalization. It would be interesting to investigate whether the findings in this study generalize to other settings during the first wave of globalization. Considering the effects of the first wave of globalization on worker inequality is another promising avenue for future research.

References

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics*, 134(4): 1949–2010.

Atack, Jeremy, Michael R. Haines, and Robert A. Margo. 2011. "Railroads and the Rise of the Factory: Evidence for the United States, 1850-1870." *Economic Evolution and Revolutions in Historical Time.* P. Rhode, J. Rosenbloom and D. Weiman, (eds). Palo Alto, CA: Stanford University Press.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–68.

Autor, David H., David Dorn, and Gordon H. Hanson. 2016. "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade." *Annual Review of Economics*, 8(1): 205–240.

Bartik, Timothy J. 1991. "Who Benefits from State and Local Economic Development Policies?" *Kalamazoo, MI: W.E. Upjohn Institute for Employment Research*.

Berger, Thor. 2019. "Railroads and Rural Industrialization: Evidence from a Historical Policy Experiment." *Explorations in Economic History*, 74: 101277.

Berger, Thor and Kerstin Enflo. 2017. "Locomotives of Local Growth: The Short- and Long-term Impact of Railroads in Sweden." *Journal of Urban Economics*, 98: 124–138. Urbanization in Developing Countries: Past and Present.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2006. "Survival of the Best Fit: Exposure to Low-wage Countries and the (Uneven) Growth of U.S. Manufacturing Plants." *Journal of International Economics*, 68(1): 219–237.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2009. "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." *Producer Dynamics: New Evidence from Micro Data*. T. Dunne, J. B. Jensen, and M. J. Roberts (eds). Chicago:University of Chicago Press.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives*, 21(3): 105–130.

Berthou, Antoine, Jong Hyun Chung, Kalina Manova, and Charlotte Sandoz Dit Bragard. 2019. "Trade, Productivity and (Mis)allocation." *CEPR Discussion Paper No. 14203*.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2016. "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." *Review of Economic Studies*, 83(1): 87–117.

Bolt, Jutta and Jan Luiten van Zanden. 2014. "The Maddison Project: Collaborative Research on Historical National Accounts." *Economic History Review*, 67(3): 627–651.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2018. "Quasi-Experimental Shift-Share Research Designs." *NBER Working Paper No. 24997*.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics*, 19(1): 22–64.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection." *Econometrica*, 80(6): 2543–2594.

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

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum. 2014. "The Rise of the East and the Far East: German Labor Markets and Trade Integration." *Journal of the European Economic Association*, 12(6): 1643–1675.

De Loecker, Jan and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–71.

di Giovanni, Julian, Andrei A. Levchenko, and Romain Rancière. 2011. "Power Laws in Firm Size and Openness to Trade: Measurement and Implications." *Journal of International Economics*, 85(1): 42–52.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, 108(4-5): 899–934.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, Markups, and the Gains from International Trade." *American Economic Review*, 105(10): 3183–3221.

Estevadeordal, Antoni. 1997. "Measuring Protection in the Early Twentieth Century." *European Review of Economic History*, 1(1): 89–125.

Fajgelbaum, Pablo P. and Stephen J. Redding. 2014. "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914." *NBER Working Paper No. 20217*.

Gaubert, Cecile. 2018. "Firm Sorting and Agglomeration." *American Economic Review*, 108(11): 3117–53.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review*, 110(8): 2586–2624.

Gourevitch, Peter Alexis. 1977. "International Trade, Domestic Coalitions, and Liberty: Comparative Responses to the Crisis of 1873-1896." *Journal of Interdisciplinary History*, 8(2): 281–313.

Gutberlet, Theresa. 2012. "Cheap Coal, Market Access, and Industry Location in Germany." *Working Paper*.

Gutberlet, Theresa. 2014. "Mechanization and the Spatial Distribution of Industries in the German Empire, 1875 to 1907." *Economic History Review*, 67(2): 463–491.

Hamerow, Theodore S., Lawrence G. Duggan, et al. 2020. "Germany from 1871 to 1918. The German Empire, 1871–1914." *Encyclopædia Britannica*.

Hohorst, Gerhard, Jürgen Kocka, and Gerhard A. Ritter. 1978. Sozialgeschichtliches Arbeitsbuch II. Materialien zur Statistik des Kaiserreichs 1870-1914. Munich: C.H. Beck Verlag, Second edition.

Hornung, Erik. 2015. "Railroads and Growth in Prussia." *Journal of the European Economic Association*, 13(4): 699–736.

Hsieh, Chang-Tai and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

Huberman, Michael, Christopher M. Meissner, and Kim Oosterlinck. 2017. "Technology and Geography in the Second Industrial Revolution: New Evidence from the Margins of Trade." *Journal of Economic History*, 77(1): 39–89.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. 2014. "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data." *American Economic Review*, 104(6): 1597–1629.

Jacks, David S., Christopher M. Meissner, and Dennis Novy. 2010. "Trade Costs in the First Wave of Globalization." *Explorations in Economic History*, 47(2): 127–141.

Jedwab, Remi and Alexander Moradi. 2016. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics*, 98(2): 268–284.

Kaiserliches Statistisches Amt. Statistisches Jahrbuch für das Deutsch Reich. Dreißigster Jahrgang. 1909.

Kaiserliches Statistisches Amt. Statistisches Jahrbuch für das Deutsch Reich. Fünfter Jahrgang. 1884.

Kopsidis, Michael and Daniel Bromley. 2016. "The French Revolution and German Industrialization: Dubious Models and Doubtful Causality." *Journal of Institutional Economics*, 12(1): 161–190.

Lehmann, Sibylle H. 2010. "The German Elections in the 1870s: Why Germany Turned from Liberalism to Protectionism." *Journal of Economic History*, 70(1): 146–178.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Melitz, Marc J. and Stephen J. Redding. 2015. "New Trade Models, New Welfare Implications." *American Economic Review*, 105(3): 1105–1146.

O'Rourke, Kevin and Jeffrey G. Williamson. 1994. "Late Nineteenth-Century Anglo-American Factor-Price Convergence: Were Heckscher and Ohlin Right?" *The Journal of Economic History*, 54(4): 892–916.

O'Rourke, Kevin H. 2000. "Tariffs and Growth in the Late 19th Century." *Economic Journal*, 110(463): 456–483.

O'Rourke, Kevin H., Alan M. Taylor, and Jeffrey G. Williamson. 1996. "Factor Price Convergence in the Late Nineteenth Century." *International Economic Review*, 37(3): 499–530.

O'Rourke, Kevin H. 2019. "Economic History and Contemporary Challenges to Globalization." *Journal of Economic History*, 79(2): 356–382.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies*, 69(1): 245–276.

Restuccia, Diego and Richard Rogerson. 2017. "The Causes and Costs of Misallocation." *Journal of Economic Perspectives*, 31(3): 151–74.

Smithson, Michael and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-likelihood Regression with Beta-distributed Dependent Variables." *Psychological methods*, 11(1): 54.

Sympher, Leo. 1913. "Das Anwachsen der deutschen Binnenschifffahrt von 1875 bis 1910." *Zeitschrift für Binnenschifffahrt*, 1: 3–12.

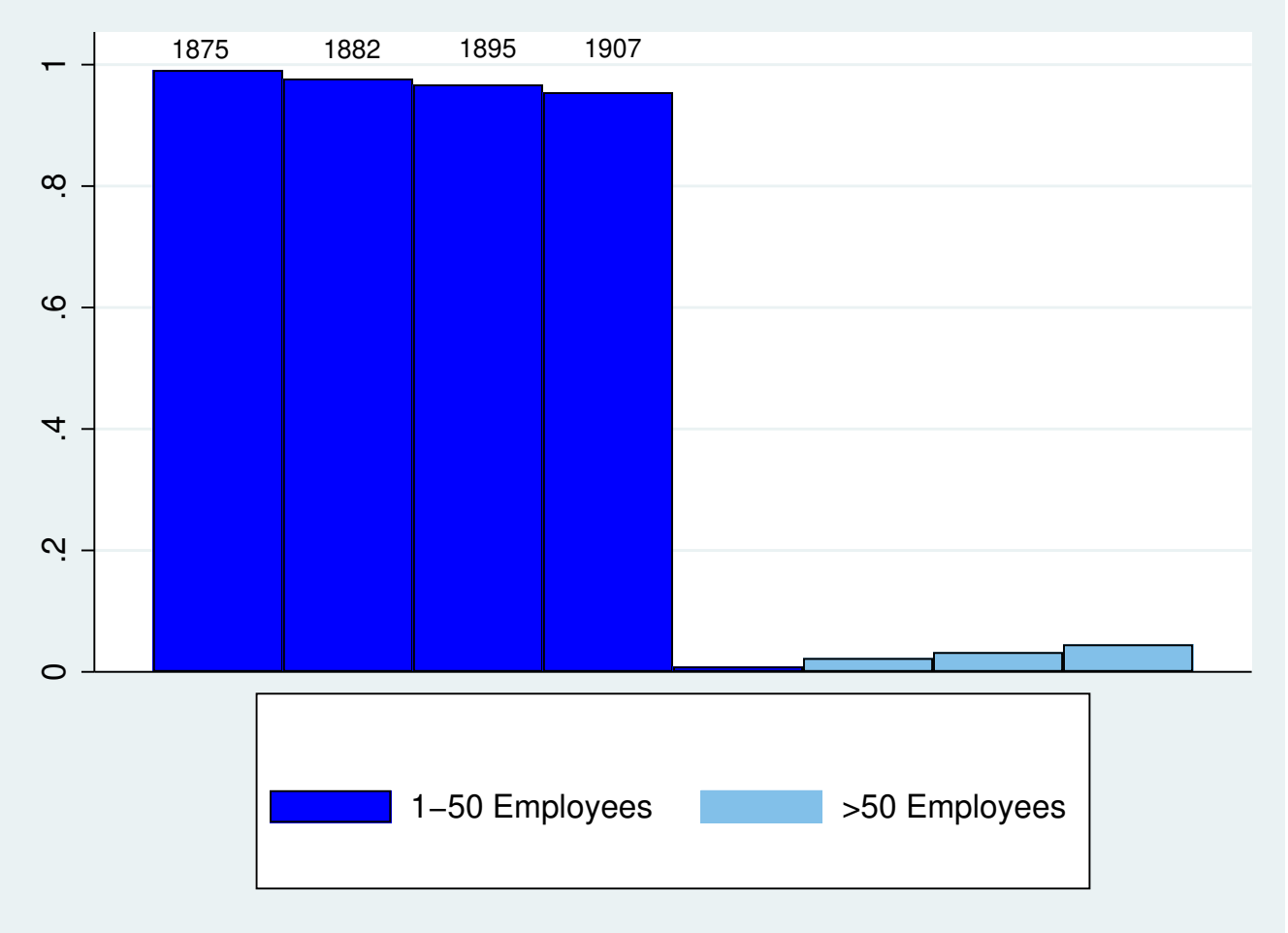
Torp, Cornelius. 2014. "The Challenges of Globalization: Economy and Politics in Germany, 1860-1914." New York and Oxford: Berghahn Books.

Trefler, Daniel. 2004. "The Long and Short of the Canada-U. S. Free Trade Agreement." *American Economic Review*, 94(4): 870–895.

Wolf, Nikolaus. 2009. "Was Germany Ever United? Evidence from Intra- and International Trade, 1885–1933." *Journal of Economic History*, 69(3): 846–881.

Wolf, Nikolaus, Max-Stephan Schulze, and Hans-Christian Heinemeyer. 2011. "On the Economic Consequences of the Peace: Trade and Borders After Versailles." *Journal of Economic History*, 71(4): 915–949.

Figure 1: Firm shares across size categories



Notes: The figure shows the share of firms with up to 50 employees and with more than 50 employees aggregated across 18 industries by census date. Author's calculations based on various volumes of Statistik des deutschen Reichs.

Figure 2: Sample calculation of the firm share from industry census data

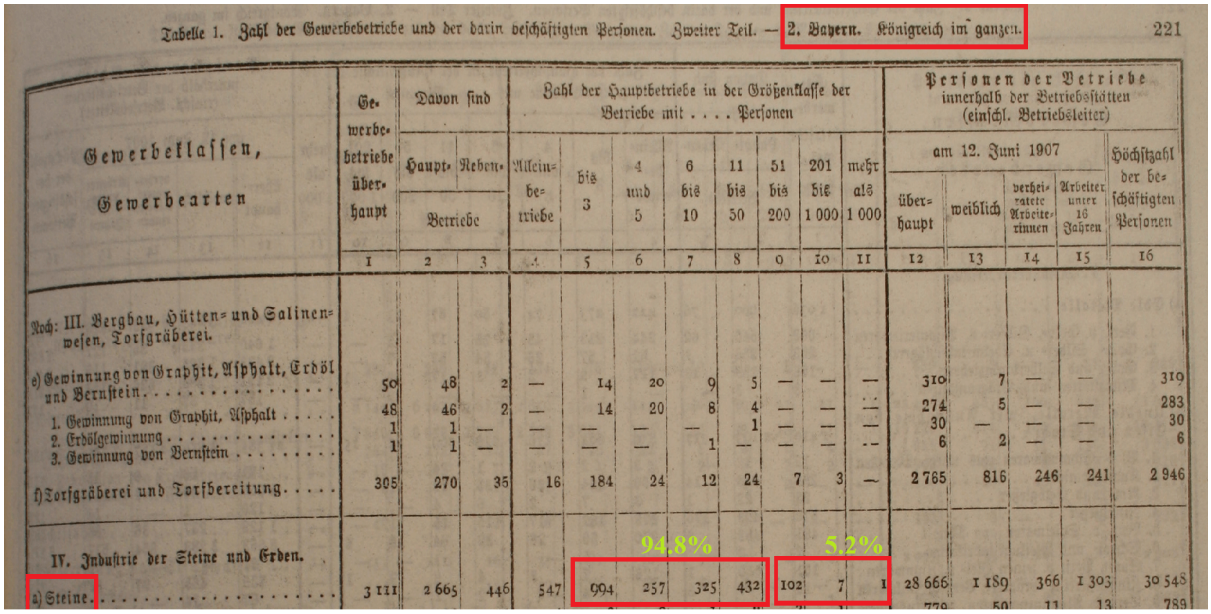


Tabelle 1. Zahl der Gewerbebetriebe und der darin beschäftigten Personen. Zweiter Teil. — 2. Bayern. Königreich im ganzen. 221

Gewerbeklassen, Gewerbearten	Gewerbebetriebe überhaupt	Davon sind: Haupt-Betriebe	Davon sind: Neben-Betriebe	Zahl der Hauptbetriebe in der Größenklasse der Betriebe mit ... Personen: Alleinbetriebe	bis 3	4 und 5	6 bis 10	11 bis 50	51 bis 200	201 bis 1 000	mehr als 1 000	Personen der Betriebe innerhalb der Betriebsstätten (einschl. Betriebsleiter) am 12. Juni 1907: überhaupt	weiblich	verheiratete Arbeiterinnen	Arbeiter unter 16 Jahren	Höchstzahl der beschäftigten Personen
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Aus: III. Bergbau, Hütten- und Salinenwesen, Torfgräberei.																
e) Gewinnung von Graphit, Asphalt, Erdöl und Bernstein	50	48	2	—	14	20	9	5	—	—	—	310	7	—	—	319
1. Gewinnung von Graphit, Asphalt	48	46	2	—	14	20	8	4	—	—	—	274	5	—	—	283
2. Erdölgewinnung	1	1	—	—	—	—	—	1	—	—	—	30	—	—	—	30
3. Gewinnung von Bernstein	1	1	—	—	—	—	1	—	—	—	—	6	2	—	—	6
f) Torfgräberei und Torfbereitung	305	270	35	16	184	24	12	24	7	3	—	2 765	816	246	241	2 946
IV. Industrie der Steine und Erden.																
a) Steine	3 111	2 665	446	547	904	257	325	432	102	7	1	28 666	1 189	366	1 303	30 545

94,8%

5,2%

Notes: The figure shows, how the variable firm share is constructed. The data point is the stone industry in district 15 (Kingdom of Bavaria and Bavarian Palatine) in the census year 1907. Overall, there were 2,118 firms counted for this district-industry pair. The firm share of firms with more than 50 employees is then given by $\frac{102+7+1}{2.118} \approx 5.19\%$. Self-employment is excluded from 1882, because the information was not collected in the 1882 census. The share of firms with more than 200 employees is calculated correspondingly. Source: Statistik des deutschen Reichs, Volume 215.

Table 1: Firm share across size categories

Number of employees	(1) > 50	(2) > 50
Sample	Matched	Aggregate
1875	0.88	0.36
1882	2.28	1.12
1895	3.24	1.97
1907	4.53	3.04

Notes: The table shows the distribution of firms by size category aggregated across districts and industries. Column (1) shows the distribution for the industries that are matched to trade and tariff data. Column (2) shows the distribution of all industries in section B (industrial sector including mining and construction industry) from the industry census data. Source: various volumes of Statistik des deutschen Reichs.

Table 2: Tariffs by industry

Industry	(1) 1875	(2) 1882	(3) 1895	(4) 1907
Iron ore	0	0	0	0
Iron and steel	0	16.12	19.73	11.30
Hard coal	0	0	0	0
Brown coal	0	0	0	0
Stone	0.31	7.31	1.26	2.70
Earth, gravel, and sand	0	0	0	0
Lime and cement	0	0	0	0
Glass	4.00	10.66	14.12	10.91
Iron and steel products	4.14	11.95	14.18	9.07
Chemistry products	0.60	6.44	0.38	1.28
Fertilizer	0	0	0	0
Fat and oils	2.09	6.27	9.01	9.73
Petroleum and other mineral oils	0	33.16	76.79	56.47
Wool	2.58	3.61	1.08	2.04
Cotton	1.04	8.75	2.58	5.92
Paper	3.27	6.38	7.07	6.79
Leather	0.68	4.61	3.63	3.07
Timber	0	2.73	6.12	6.65
Average Tariff	1.04	6.56	8.66	7.00
Median Tariff	0.15	5.44	1.92	2.88

Notes: Average effective tariff rates in percent calculated from foreign trade statistics published as part of the "Statistik des deutschen Reichs." Average tariff is the sample average across industries by year, whereas median tariff is the median tariff across industries by year.

Table 3: Firm share and aggregate trade

Panel A: First-stage	ln(aggregate trade$_{ijt}$)
ln($\widehat{\text{aggregate trade}}_{ijt}$)	0.5058***
	(0.0924)
Estimation method	OLS
Kleibergen-Paap LM-statistic	22.13
Kleibergen-Paap F-statistic	29.96
R^2	0.866
Observations	554
Panel B: IV	**firmshare$_{ijt}$**
Number of employees	> 50
Mean of dep var	0.1629
ln(aggregate trade$_{ijt}$)	0.0875*
	(0.0458)
Estimation method	Fixed eff. IV
R^2	0.172
Number of districts	17
Number of industries	18
Observations	554
Panel C: OLS	**firmshare$_{ijt}$**
Number of employees	> 50
Mean of dep var	0.1629
ln(aggregate trade$_{ijt}$)	0.0378*
	(0.0211)
Estimation method	Fixed eff.
R^2	0.190
Number of districts	17
Number of industries	18
Observations	554

Notes: Panel A shows the first stage regression, which is defined in equation (6). Panel B presents instrumental variable estimates of equation (3). Panel C displays the OLS estimate of equation (3). The unit of observation is at the district-industry-class-size level at time t. The dependent variable is the share of firms with more than 50 employees in industry i, district j at time t. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The sample consists of all district-industry pairs in 1895 and 1907. All regressions include district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend in the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Pareto alpha and aggregate trade

	(1)	(2)
	$\widehat{\alpha}_{ijt}$	
Mean of dep var	0.6178	
ln(aggregate trade$_{ijt}$)	−0.1248***	−0.2337***
	(0.0259)	(0.0798)
Estimation method	Fixed eff.	Fixed eff. IV
R^2	0.580	0.525
Number of districts	14	14
Number of industries	13	13
Observations	150	150

Notes: The OLS estimate of equation (9) is presented in column (1). The IV estimation of equation (9) is shown in column (2). The unit of observation is a district-industry pair at time t. The dependent variable is the median estimate for the alpha of a Pareto distribution estimated from the empirical CDF. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The sample consists of all district-industry pairs with positive probability mass in all size categories between 1875 and 1907. All regressions include district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend in the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Firm count and aggregate trade

	(1)	(2)
	ln(firmnumber$_{ijt}$)	
Number of employees	≤ 50	> 50
ln(aggregate trade$_{ijt}$)	–0.0508	0.3457*
	(0.1806)	(0.1966)
Estimation method	Fixed eff. IV	Fixed eff. IV
R^2	0.392	0.545
Number of districts	17	17
Number of industries	18	18
Observations	554	554

Notes: Instrumental variable estimates of equation (3) using the firm count as outcome. The unit of observation is at the district-industry-class-size level at time t. The dependent variable in column (1) is the logarithm of the count of firms with up to 50 employees in industry i, district j at time t. The dependent variable in column (2) is the logarithm of the count of firms with more than 50 employees in industry i, district j at time t. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The first-stage regression is defined in equation (6). The sample consists of all district-industry pairs in 1895 and 1907. All regressions include district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend of the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Firm share and tariffs

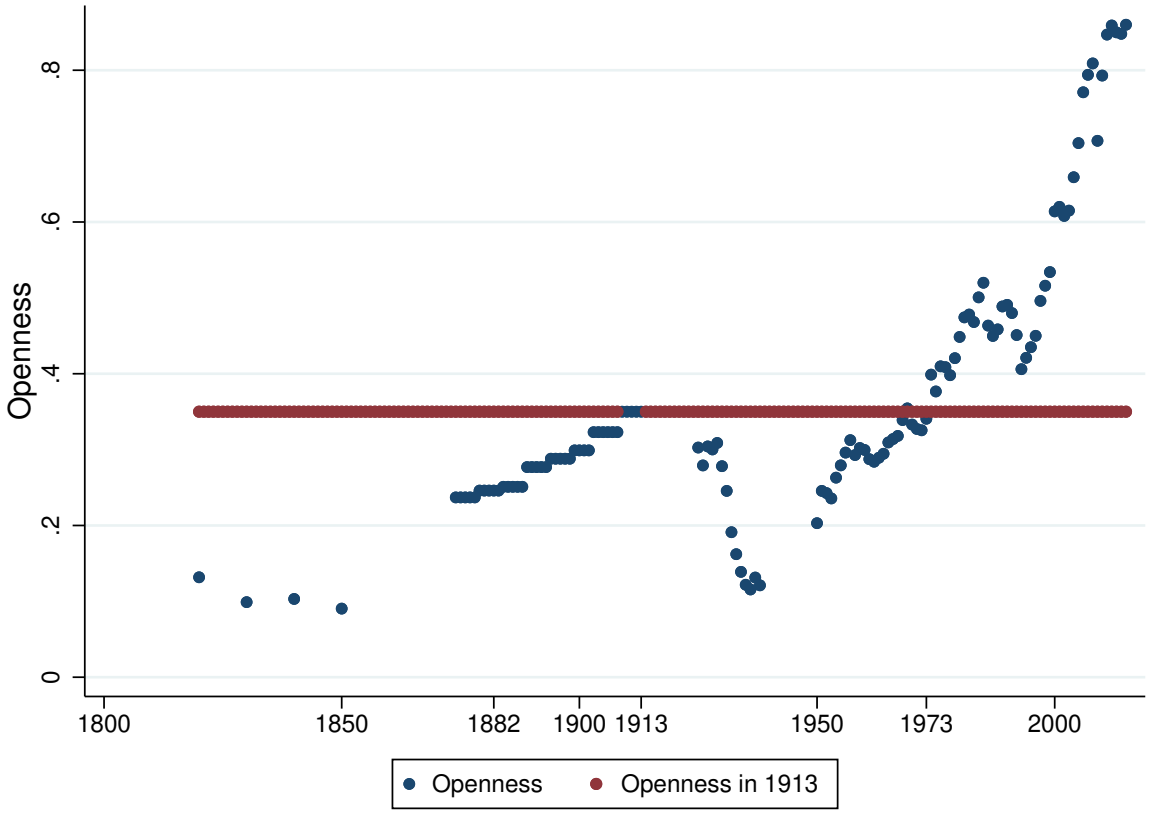
	(1) Δfirmshare$_{ijt}$	(2) Δfirmshare$_{jt}$
Number of employees	> 200	> 200
Mean of dep var	0.0095	0.0026
Δtariff$_{it}$	−0.1071***	
	(0.0348)	
Δtariff$_{jt}$		−0.1545***
		(0.0556)
Estimation method	First diff.	First diff.
R^2	0.074	0.216
Number of districts	37	37
Number of industries	18	
Observations	1,559	111

Notes: The regression in column (1) presents estimates of equation (10). The unit of observation in column (1) is at the district-industry-class-size level at time t. The dependent variable in column (1) is the first difference of the share of firms with more than 200 employees in industry i, district j at time t. The independent variable in column (1) is the first difference in tariffs measured in percent for the respective industry. The regression in column (1) includes district-time fixed effects as control. The regression in column (2) shows estimates of equation (12). The unit of observation in column (2) is at the district-class-size level at time t. The dependent variable in column (2) is the share of firms with more than 200 employees in district j at time t. The independent variable in column (2) is the first difference in tariffs measured in percent at the district level. The regression in column (2) includes time fixed effects as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

7 Online Appendix

Figures

Figure A.1: Openness in Germany from 1820 to 2015



Notes: The variable openness is defined as share of exports plus imports over GDP. The red line illustrates the level of openness reached in 1913 and the graph shows that this level was passed sustainable only sixty years later. The data steam from different sources with possibly different price indices. GDP data for 1820, 1830, 1840, and 1850 interpolated from Fremdling (1995) and nominal trade data from Bondi (1958). The data for 1874-1913 are five-year averages taken from Torp (2014). Torp (2014) measures openness normalized by GNP. The data for 1925-1938 taken from Ritschl (2002). Observations from 1950-2015 calculated from statistics published by Federal Statistical Office of Germany (2016).
Federal Statistical Office of Germany (2016) online sources retrieved on 20 September, 2016:
`https://www.destatis.de/DE/ZahlenFakten/Indikatoren/Globalisierungsindikatoren/Tabellen/01_02_03_AH.html`
`https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Inlandsprodukt/Tabellen/Volkseinkommen1925_pdf.pdf?__blob=publicationFile`
`https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Aussenhandel/lrahl01.html`

Figure A.2: Map of districts

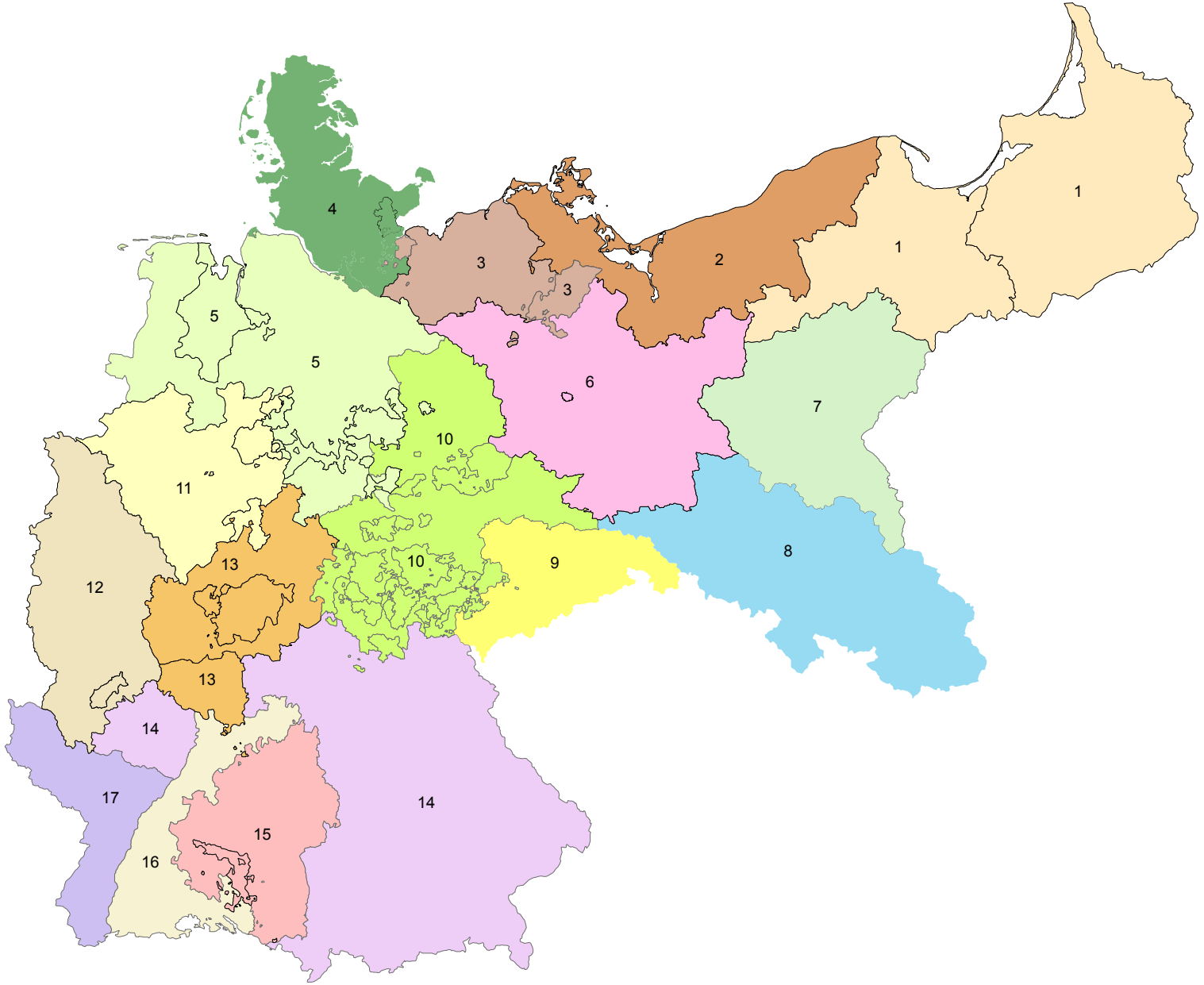
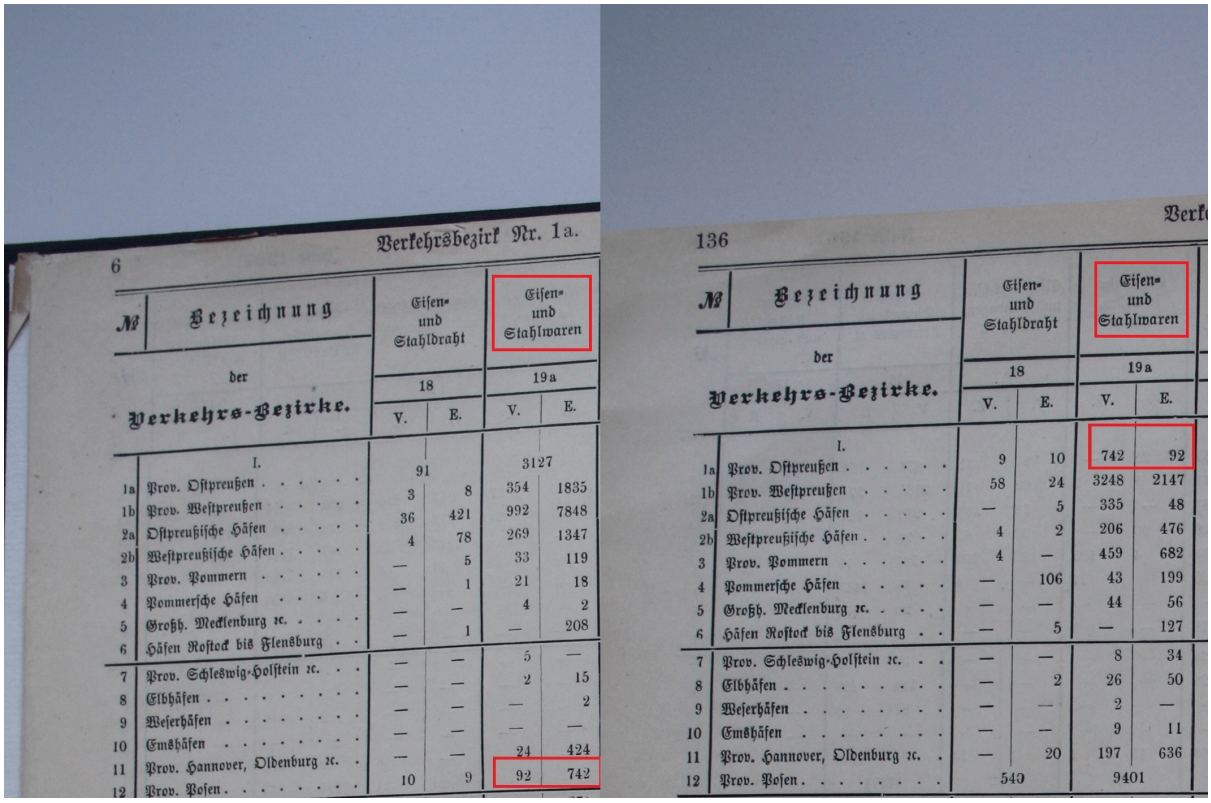


Figure A.3: Sample bilateral trade flow



6 Verkehrsbezirk Nr. 1a.

№	Bezeichnung der Verkehrs-Bezirke.	Eisen- und Stahldraht 18 V.	Eisen- und Stahldraht 18 E.	Eisen- und Stahlwaren 19a V.	Eisen- und Stahlwaren 19a E.
	I.	91		3127	
1a	Prov. Ostpreußen	3	8	354	1835
1b	Prov. Westpreußen	36	421	992	7848
2a	Ostpreußische Häfen	4	78	269	1347
2b	Westpreußische Häfen	—	5	33	119
3	Prov. Pommern	—	1	21	18
4	Pommersche Häfen	—	—	4	2
5	Großh. Mecklenburg ꝛc.	—	1	—	208
6	Häfen Rostock bis Flensburg	—	—	5	—
7	Prov. Schleswig-Holstein ꝛc.	—	—	2	15
8	Elbhäfen	—	—	—	2
9	Weserhäfen	—	—	—	—
10	Emshäfen	—	—	24	424
11	Prov. Hannover, Oldenburg ꝛc.	10	9	92	742
12	Prov. Posen				

136 Verk

№	Bezeichnung der Verkehrs-Bezirke.	Eisen- und Stahldraht 18 V.	Eisen- und Stahldraht 18 E.	Eisen- und Stahlwaren 19a V.	Eisen- und Stahlwaren 19a E.
	I.			742	92
1a	Prov. Ostpreußen	9	10	3248	2147
1b	Prov. Westpreußen	58	24	335	48
2a	Ostpreußische Häfen	—	5	206	476
2b	Westpreußische Häfen	4	2	459	682
3	Prov. Pommern	4	—	43	199
4	Pommersche Häfen	—	106	44	56
5	Großh. Mecklenburg ꝛc.	—	—	—	127
6	Häfen Rostock bis Flensburg	—	5	8	34
7	Prov. Schleswig-Holstein ꝛc.	—	—	26	50
8	Elbhäfen	—	2	2	—
9	Weserhäfen	—	—	9	11
10	Emshäfen	—	—	197	636
11	Prov. Hannover, Oldenburg ꝛc.	—	20	9401	
12	Prov. Posen	549			

Notes: The figure shows a bilateral trade flow between one trade district of district 1 and district 7. The "V" at the top of the column indicates exports and the "E" represents imports. It shows for the industry "Iron and steel products" (framed in red at the top of both pages) in 1907. The red framed boxes illustrate exports of trade district one correspond to the imports of trade district twelve and vice versa. Source: Statistik der Güterbewegung auf Deutschen Eisenbahnen im Jahre 1907.

Definitions

Product group definition

- Product group 1:
  - Iron ore
  - Iron and steel
  - Hard coal
  - Brown coal
  - Iron and steel products
- Product group 2:
  - Stone
  - Earth, gravel, and sand
  - Lime and cement
  - Glass
- Product group 3:
  - Chemistry products
  - Fertilizer
  - Fat and oils
  - Petroleum and other mineral oils
- Product group 4:
  - Wool
  - Cotton
- Product group 5:
  - Paper
  - Leather
  - Timber

List of districts (Tables 3-4 and Tables A.6-A.9)

- 1: Provinces of East and West Prussia and seaports Memel, Pillau, Königsberg, Elbing, and Neufahrwasser
- 2: Province of Pomerania and seaports Stolpmünde, Rügenwalde, Colberg, Stettin, Swinemünde, Wolgast, and Stralsund
- 3: Grand Duchies of Mecklenburg-Strelitz and Mecklenburg-Schwerin and seaports Rostock, Warnemünde, and Wismar[28]
- 4: Province of Schleswig-Holstein, City of Lübeck, City of Hamburg, Principality of Lübeck and seaports Flensburg, Kiel, Lübeck, Hamburg, Altona, and Glückstadt[28][29]
- 5: City of Bremen, Province of Hanover, Duchy of Oldenburg, Duchy of Braunschweig, Principality of Schamburg-Lippe, counties Pyrmont and Rinteln, and seaports Harburg, Stade, Cuxhafen, Bremen, Vegesack, Geestemünde, Bremerhafen, Nordenham, Brake, Elsfleth, Emden, Leer, and Papenburg[28][29]
- 6: Urban district of Berlin and Province of Brandenburg
- 7: Province of Posen
- 8: Province of Schlesien
- 9: Kingdom Saxony
- 10: Province of Saxony, Grand Duchy Sachsen-Weimar, Duchies of Sachsen-Meiningen, Sachsen-Altenburg, Sachsen-Coburg-Gotha, Anhalt, Principalities of Schwarzburg-Sonderhausen, Schwarzburg-Rudolfstadt, Reuss-Greiz, and Reuss-Gera, and county Schmalkalden
- 11: Province of Westphalia, Principality of Lippe, Principality of Walbeck without county Pyrmont
- 12: Province of Rhineland and Principality of Birkenfeld without county Wetzlar[29]

[28]Two trade districts defined in the railway statistics include seaports from two of the seventeen districts: railway district "Seaports Rostock, Wismar, Flensburg, Kiel and Lübeck" contains seaports from districts three and four. Railway district "Seaports Hamburg, Altona, Glückstadt, Harburg, Stade and Cuxhafen" contains seaports from districts four and five. As allocation rule to proxy trade in each district, I assign the share of the trade flow to each district according to their share of the number of ships trading goods in each of these ports.

[29]The Grand Duchy Oldenburg consisted of separate territories Duchy of Oldenburg, Principality of Birkenfeld, and Principality of Lübeck, which by construction belong to three different districts. For 1882 it is feasible to correctly allocate the number of firms to each spatial unit. For years 1895 and 1907 this study makes use of industry census data that contain information about the number of firms of each spatial unit. Firms are allocated according to the corresponding industry shares taking into account the differences in total employment within each industry.

- 13: Province of Hessia-Nassau and Grand Duchy of Hessia with county Wetzlar but without counties Rinteln and Schmalkalden
- 14: Kingdom of Bavaria and Bavarian Palatine[30]
- 15: Kingdom of Württemberg and Province of Hohenzollern
- 16: Grand Duchy of Baden[30]
- 17: Alsace Lorraine

Districts in railway statistics assign the following counties different from census data which follow administrative boundaries:

- County Pyrmont in Waldeck located in district eleven is assigned to district five
- County Rinteln located in district thirteen is assigned to district five
- County Schmalkalden located in district thirteen is assigned to district ten
- County Wetzlar located in district twelve is assigned to district thirteen

The administrative structure divided the German Empire into 1,049 counties in 1900. I cannot correct for this assignment, because firm size distribution data are not available at the county level. However, this departure from administrative boundaries does not induce any systematic measurement error given the size of these counties.

[30]Mannheim belonged to Grand Duchy Baden and Ludwigshafen to Kingdom of Bavaria. Hence, the railway district "Mannheim and Ludwigshafen" contains parts of two districts. Trade flows of the railway district are apportioned according to the share of employment of Mannheim (Grand Duchy Baden) and Ludwigshafen (Kingdom of Bavaria) of total employment of Mannheim and Ludwigshafen in the corresponding industry group.

List of 17 countries and regions outside of German Empire included as trading partners in international exports and imports on railway

- Russia
- Poland
- Galicia and Bukovina
- Romania
- Hungary, Slavonia, Croatia, Transylvania, Bosnia and Herzegovina
- Serbia, Bulgaria, Turkey, and Greece
- Bohemia
- Austria (without Bohemia and Galicia)
- Switzerland
- Italy
- France
- Luxembourg
- Belgium
- Netherlands
- Great Britain
- Sweden and Norway
- Denmark

List of districts (Table 6 and Table A.10)

- Provinces of East and West Prussia; Province of Pomerania;
- Grand Duchy of Mecklenburg-Strelitz; Grand Duchy of Mecklenburg-Schwerin;
- Province of Schleswig-Holstein; City of Lübeck; City of Hamburg[31]; City of Bremen[31];
- Province of Hanover; Grand Duchy Oldenburg; Duchy of Braunschweig; Principality of Schamburg-Lippe;
- Berlin and Province of Brandenburg; Province of Posen;
- Province of Schlesien; Kingdom Saxony;
- Province of Saxony; Grand Duchy Sachsen-Weimar; Duchy of Sachsen-Meiningen; Duchy of Sachsen-Altenburg; Duchy of Sachsen-Coburg-Gotha; Duchy of Anhalt;
- Principality of Schwarzburg-Sonderhausen; Principality of Schwarzburg-Rudolfstadt;
- Principality of Reuss-Greiz; Principality of Reuss-Gera;
- Province of Westphalia; Principality of Lippe; Principality of Walbeck;
- Province of Rhineland; Province of Hessia-Nassau; Grand Duchy of Hessia;
- Kingdom of Bavaria; Kingdom of Württemberg;
- Province of Hohenzollern; Grand Duchy of Baden; Alsace Lorraine

[31]The cities of Bremen and Hamburg joined the Zollverein in 1888. Observations for these districts are included in the sample from 1895.

Details on Industry Census Data

Industry surveys were conducted at the municipality level. Municipalities were divided into counting districts and one assistant was responsible for each counting district. Assistants distributed and collected the surveys and checked their consistency. Assistants were asked to survey not more than 50 households and companies in a counting district such that distribution and collection of surveys was feasible within a time frame of one week.[32] Information about the counting was announced early and no public events should take place two days prior, on the day of the counting, and one day after the counting. Assistants filled out control lists passed to the head of the municipality. Thereafter, the head of the municipality had to confirm the consistency of the counting by signature before the results were transferred to the county and then to the regional statistical office. Not only summaries were transferred, but also the questionnaires themselves were shipped to the responsible statistical office of the state in which a county was located. Companies were forced by law to participate in the survey and the director of each company had to confirm the truthfulness of their information by signature. Misreporting was punished by 30 Mark in all census years which corresponds to a nominal value in 2015 of 213 Euro (1882) and 174 Euro (1907). Hence, measurement error stemming from untruthful reporting is unlikely to be an issue.[33] On the other hand, municipalities were offered an incentive to provide full coverage of industry and occupation census. For each inhabitant, they received one Pfennig before the counting, one Pfennig three and a half months after the counting day, and one Pfennig on January 1st in 1896 in the case of the 1895 census.[34]

To the best of my knowledge, the availability of data in the German Empire provides an unparalleled opportunity to combine trade and industry census data at the regional level in this historical context, particularly in contrast to the other two largest economies in terms of GDP in this period - the United Kingdom and the United States. In the United Kingdom, no industry census was conducted for the period under consideration. In the United States, information on companies was collected early on incidentally to the collection of population data with a frequency of decennial census. However, only companies with a value of production of more than 500 dollars were counted. According to Hesse (1914), a significant fraction of factories and small companies had not been counted arbitrarily. In addition, Hesse cites Francis Walker, who in 1869 was chief of the statistical office in Washington and census superintendent for the 1870 and 1880 censuses. According to Walker, respondents had little incentives to report truthfully due to fear of the use of census data by tax authorities or simply to escape counting.

[32] Information on the counting procedure is exemplary presented for 1895 but it was similar in other census years.

[33] Information retrieved from `https://www.bundesbank.de/Redaktion/DE/Downloads/Statistiken/Unternehmen_Und_Private_Haushalte/Preise/kaufkraftaequivalente_historischer_betrage_in_deutschen_waehrungen.pdf?_blob=publicationFile.` (accessed on 31 March, 2016), values are purchasing power equivalent of the average value of one Euro in 2015.

[34] 100 Pfennig corresponded to one Mark, the currency of the German Empire. Payment before the counting was based on the recent population census data. For 1895 the reference year was the population census in 1890.

A comparison across years and regions before 1900 is hardly reasonable. Only from 1900 did the quality of data improve and did industry censuses comprise full coverage of all firms.[35]

Appendix References

Atack, Jeremy, and Fred Bateman. 1999. "Nineteenth-Century U.S. Industrial Development Through the Eyes of the Census of Manufactures a New Resource for Historical Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*32(4): 177–188.

Atack, Jeremy, Michael R. Haines, and Robert A. Margo. 2011. "Railroads and the Rise of the Factory: Evidence for the United States, 1850–1870." In *Economic Evolution and Revolutions in Historical Time.* Edited by P. Rhode, J. Rosenbloom, and D. Weiman. Palo Alto: Stanford University Press.

Bondi, Gerhard. 1958. Deutschlands Aussenhandel 1815–1870. Berlin: Akademie Verlag.

Fremdling, Rainer. 1995. "German National Accounts for the 19th and Early 20th Century." *Scandinavian Economic History Review* 43(1): 77–100.

Hesse, Albert. 1914. "Gewerbestatistik." In *Grundriß zum Studium der Politischen Ökonomie* Von Prof. Dr. J. Conrad, 2nd ed. Jena: Verlag Von Gustav Fischer.

Ritschl, Albrecht. 2002. Deutschlands Krise und Konjunktur. Binnenkonjunktur, Auslandsverschuldung und Reparationsproblem Zwischen Dawes-Plan und Transfersperre 1924–34. Berlin: Akademie Verlag.

Smithson, Michael, and Jay Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-likelihood Regression with Beta-distributed Dependent Variables." *Psychological Methods* 11(1): 54.

Torp, Cornelius. 2014. The Challenges of Globalization: Economy and Politics in Germany, 1860–1914. New York and Oxford: Berghahn Books.

[35] Walkers' concerns are also discussed in Atack and Bateman (1999). The quality of the US census data at the aggregate level is doubtful. Atack, Haines, and Margo (2011) use digitized original firm surveys. However, for the same industries as considered in this paper their sample contains 2,567 establishments in 1880 compared to 167,767 establishments in the 1882 census data deployed here.

Tables

Table A.1: Firm number across size categories

Number of employees	(1) 1-50	(2) >50
1875	356,493	3,187
1882	163,942	3,825
1895	176,537	5,941
1907	186,653	8,862

Notes: The table shows the distribution of firms by size category aggregated across districts and industries. The category 1-5 employees includes self employment in 1875. Author's calculations based on various volumes of Statistik des deutschen Reichs.

Table A.2: Employment across districts

District Number	(1) 1882	(2) 1895	(3) 1907
1	22,457	30,426	48,970
2	18,321	23,283	30,783
3	7,356	9,790	12,112
4	26,154	39,178	64,155
5	61,925	86,074	121,060
6	98,294	145,924	202,948
7	10,089	14,652	20,050
8	135,555	201,906	284,813
9	149,207	218,240	300,201
10	120,570	154,649	211,147
11	148,009	225,042	372,090
12	255,638	346,811	546,836
13	58,036	93,204	122,288
14	108,678	185,385	238,849
15	43,736	66,083	85,581
16	35,424	55,505	80,467
17	90,647	99,973	146,698
Sum	1,390,096	1,996,125	2,889,048

Notes: Total employment across industries by district. Author's calculations based on various volumes of Statistik des deutschen Reichs.

Table A.3: Population across districts

District Number	(1) 1882	(2) 1895	(3) 1907
1	3,302,528	3,450,746	3,633,579
2	3,434,972	4,409,244	5,706,576
3	1,517,712	1,575,052	1,702,286
4	1,665,617	1,774,046	1,964,806
5	3,998,782	4,355,477	4,993,098
6	3,754,116	4,328,073	4,889,295
7	1,689,621	2,080,890	2,599,051
8	2,922,288	3,364,889	3,986,105
9	2,234,514	2,850,951	3,980,652
10	4,147,917	5,090,825	6,697,844
11	5,268,761	5,779,176	6,598,168
12	3,014,822	3,753,262	4,585,500
13	2,023,843	2,136,572	2,406,659
14	1,558,598	1,719,238	2,057,561
15	2,474,327	2,768,928	3,351,508
16	674,160	709,836	747,592
17	1,539,580	1,623,079	1,820,249
Total	45,222,158	51,770,284	61,720,529

Notes: Population across districts and time. Author's calculation based on Statistik des deutschen Reichs Volume 111 and Statistik des deutschen Reichs Volume 213.

Table A.4: Growth in aggregate trade flows by industry

Industry	(1) 1895	(2) 1907	(3) Growth 1907-1895
Iron ore	6593530.5	17571503	166.50
Iron and steel	8388174.5	18715335.5	123.12
Hard coal	62952158	119445599.5	89.74
Brown coal	8897812.5	20767049	133.40
Stone	7775992	19756658.5	154.07
Earth, gravel, and sand	2973685	9270452.5	211.75
Lime and cement	3117064	7340835.5	135.50
Glass	571520	1150965.5	101.39
Iron and steel products	877290.5	2604348	196.86
Chemistry products	328992.5	1424734.5	333.06
Fertilizer	3288050	9176433.5	179.08
Fat and oil	415939	824169	98.15
Petroleum and other mineral oil	706633.5	1586558	124.52
Wool	374385.5	447877.5	19.63
Cotton	613178.5	1089256.5	77.64
Paper	799640.5	2005809	150.84
Leather	317944.5	556044	74.89
Timber	7976467	14792960.5	85.46

Notes: The table displays aggregate trade flows by industry in 1895 and 1907. Column (3) shows the growth rate in this period. Author's calculation based on bilateral trade flows digitized from various volumes of Statistik der Güterbewegung auf deutschen Eisenbahnen.

Table A.5: Extensive margin for domestic and international trade

	(1) 1883	(2) 1895	(3) 1907	(4) Difference 1907-1883
Panel A: Mean across industries				
Domestic destination/source	11.81	13.40	14.71	2.90
Foreign destination	5.04	6.53	8.17	3.13
Foreign source	3.77	4.95	6.95	3.18
Panel B: Median across industries				
Domestic destination	12.39	14.06	15.50	3.11
Domestic source	12.11	13.56	14.83	2.72
Foreign destination	5.17	7.06	8.78	3.61
Foreign source	3.78	5.11	7.11	3.33

Notes: Panel A displays the mean number of destinations for seventeen districts defined in the Online Appendix across industries. Panel B shows average of the median number of destinations by industry. Domestic destination is the number of domestic regions that import from a domestic district including the district itself. Domestic source is the number of domestic regions that exports to a domestic district including the district itself. Foreign destination is the number of foreign regions that import from a domestic district. Foreign source is the number of foreign regions that export to a domestic district. The number of domestic and foreign regions is seventeen, respectively. Author's calculation based on bilateral trade flows digitized from various volumes of Statistik der Güterbewegung auf deutschen Eisenbahnen.

Table A.6: Firm share and aggregate trade
- adding industry-time fixed effects

	(1)	(2)
	firmshare_{ijt}	
Number of employees	> 50	> 200
Mean of dep var	0.1629	0.0599
$\ln(\text{aggregate trade}_{ijt})$	0.0286	0.0247*
	(0.0298)	(0.0143)
Estimation method	Fixed eff.	Fixed eff.
R^2	0.285	0.278
Number of districts	17	17
Number of industries	18	18
Observations	554	554

Notes: OLS estimates of equation (3) using industry-time fixed effects. The unit of observation is at the district-industry-class-size level at time t. The dependent variable in column (1) is the share of firms with more than 50 employees in industry i, district j at time t. The dependent variable in column (2) is the share of firms with more than 200 employees in industry i, district j at time t. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The sample consists of all district-industry pairs in 1895 and 1907. All regressions include district-industry, district-time fixed effects, industry-time fixed effects, and the pre-trend of the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Firm share and aggregate trade - beta regression

	firmshare_{ijt}
Number of employees	> 50
Mean of dep var	0.1629
$\ln(\text{aggregate trade}_{ijt})$	0.4098**
	(0.1944)
Estimation method	Maximum likelihood
log likelihood	1148.674
Number of districts	17
Number of industries	18
Observations	554

Notes: Maximum likelihood estimates of equation (3). The unit of observation is at the district-industry-class-size level at time t. The dependent variable is the share of firms with more than 50 employees in industry i, district j at time t. Following Smithson and Verkuilen (2006) the firm share is transformed such that it lies in between 0 and 1. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The sample consists of all district-industry pairs in 1895 and 1907. The regression includes district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend of the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Firm share and aggregate trade - heterogeneity across product groups

	firmshare_{ijt}
Number of employees	> 50
Mean of dep var	0.1629
$\ln(\text{aggregate trade}_{ijt}) \times (\text{product group } 1_i)$	–0.0064
	(0.0368)
$\ln(\text{aggregate trade}_{ijt}) \times (\text{product group } 2_i)$	0.2003**
	(0.0933)
$\ln(\text{aggregate trade}_{ijt}) \times (\text{product group } 3_i)$	0.0362
	(0.0236)
$\ln(\text{aggregate trade}_{ijt}) \times (\text{product group } 4_i)$	0.0105
	(0.0277)
$\ln(\text{aggregate trade}_{ijt}) \times (\text{product group } 5_i)$	0.1110***
	(0.0417)
Estimation method	Fixed eff.
R^2	0.222
Number of districts	17
Number of industries	18
Observations	554

Notes: OLS estimates of equation (3). The unit of observation is at the district-industry-class-size level at time t. The dependent variable is the share of firms with more than 50 employees in industry i, district j at time t. The independent variables are the logarithm of aggregate trade as defined in equation (1) interacted with five product group fixed-effects for the respective district-industry pair. The sample consists of all district-industry pairs in 1895 and 1907. All regressions include district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend of the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Firm share and aggregate trade - robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)
			firmshare_{ijt}			
Number of employees			> 50			
Mean of dep var			0.1629			
$\ln(\text{aggregate trade}_{ijt})$	0.0777*	0.0844*	0.1135*	0.0949*	0.0958*	0.0875***
	(0.0432)	(0.0453)	(0.0615)	(0.0561)	(0.0537)	(0.0289)
$\ln(\text{aggregate trade}_{ijt}) \times (\text{sea access}_j)$		0.0118				
		(0.0536)				
$\ln(\text{aggregate trade}_{ijt}) \times (\text{intermediate}_i)$			0.0533			
			(0.0618)			
$\ln(\text{aggregate trade}_{ijt}) \times (\text{top 3 coal}_j)$				–0.0202		
				(0.0494)		
$\ln(\text{aggregate trade}_{ijt}) \times (\text{top 4 urban share}_j)$					–0.0215	
					(0.0467)	
Estimation method	F.E. IV	F.E. IV	F.E. IV	F.E. IV	F.E. IV	F.E. IV
R^2	0.183	0.176	0.099	0.171	0.166	0.172
Number of districts	17	17	17	17	17	17
Number of industries	18	18	18	18	18	18
Observations	554	554	554	554	554	554

Notes: Instrumental variable estimates of equation (3). The unit of observation is at the district-industry-class-size level at time t. The dependent variable is the share of firms with more than 50 employees in industry i, district j at time t. The independent variable is the logarithm of aggregate trade as defined in equation (1) for the respective district-industry pair. The first-stage regression is defined in equation (6). The sample consists of all district-industry pairs in 1895 and 1907. The regression in column (1) adds the share of railway trade as of trade on railway and waterway as control. Column (2) analyzes heterogeneity by sea access. Column (3) studies heterogeneity by intermediate producing industries. Column (4) tests for heterogeneity by the three largest coal producing districts. Column (5) examines heterogeneity by the top four districts in which the share of population in cities above 100k inhabitants as of the total population in 1895. Column (6) calculates HAC consistent standard errors following Conley (1999). All regressions include district-industry, district-time fixed effects, product-group-time fixed effects, and the pre-trend of the dependent variable as control. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Firm share and tariffs - robustness checks

	(1)	(2)	(3)	(4)
		Δfirmshare$_{ijt}$		
Number of employees		> 200		
Mean of dep var		0.0095		
Δtariff$_{it}$	−0.1272*	−0.0929**	−0.1290***	−0.1256***
	(0.0666)	(0.0394)	(0.0459)	(0.0395)
Δtariff$_{it} \times$ (agglomeration$_j$)	0.0321			
	(0.0775)			
Δtariff$_{it} \times$ (population growth above median$_j$)		−0.0282		
		(0.0710)		
Δtariff$_{it} \times$ (junker district$_j$)			0.0597	
			(0.0660)	
Δtariff$_{it} \times$ (low import share$_i$)				0.0984
				(0.0663)
Estimation method	First diff.	First diff.	First diff.	First diff.
R^2	0.074	0.074	0.075	0.076
Number of districts	37	37	37	37
Number of industries	18	18	18	18
Observations	1,559	1,559	1,559	1,559

Notes: Estimates of equation (10). The unit of observation is at the district-industry-class-size level at time t. The dependent variable is the first difference of share of firms with more than 200 employees in industry i, district j at time t. The independent variable is the first difference in tariffs measured in percent for the respective industry. Column (1) analyzes heterogeneity by agglomeration with a dummy equal to one if the population density was above median in at least two census years. Column (2) investigates heterogeneity by population growth with a dummy equal to one if the population growth was above median between two censuses. Column (3) tests for lobbying with a dummy equal to one if the district was dominated by the Junker elite. The districts were Grand Duchy Mecklenburg-Schwerin, Grand Duchy Großherzogtum Mecklenburg-Strelitz, Berlin, Duchy Anhalt, Province of Saxony, Province of Pommern, Province of Posen, Province of Ostpreußen, and Province of Schlesien. Column (4) examines heterogeneity by the import share with a dummy equal to one if the industry import share from abroad via railway was below the median between 1883 and 1907. All regressions include district-time fixed effects as control. The regression in column (4) additionally controls for a dummy, which is equal to one if the industry import share from abroad via railway was below the median between 1883 and 1907. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

8 Statistical Appendix

Kaiserliches Statistisches Amt. Auswärtiger Waarenverkehr des Deutschen Zollgebiets im Jahre 1875. Berlin, 1876, *Statistik des deutschen Reichs Alte Folge, Band 22.*

Kaiserliches Statistisches Amt. Die Ergebnisse der Deutschen Gewerbezählung vom 1. Dezember 1875. Berlin, *Statistik des deutschen Reichs Alte Folge, Band 34, Theil 1.*

Kaiserliches Statistisches Amt. Die Ergebnisse der Deutschen Gewerbezählung vom 1. Dezember 1875. Berlin, *Statistik des deutschen Reichs Alte Folge, Band 34, Theil 2.*

Kaiserliches Statistisches Amt. Die Ergebnisse der Deutschen Gewerbezählung vom 1. Dezember 1875. Berlin, *Statistik des deutschen Reichs Alte Folge, Band 35, Theil 1.*

Kaiserliches Statistisches Amt. Gewerbestatistik nach der allgemeinen Berufszählung vom 5. Juni 1882: 2. Gewerbestatistik der Staaten und größeren Verwaltungsbezirke. Berlin, 1886, reedition as *Statistik des deutschen Reichs, Volume 7.* Osnabrück, 1973.

Kaiserliches Statistisches Amt. Berufs- und Gewerbezählung vom 14. Juni 1895. Die berufliche und soziale Gliederung des Deutschen Volkes. Berlin, 1899, reedition as *Statistik des deutschen Reichs, Volume 111.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Gewerbezählung vom 14. Juni 1895. Gewerbe-Statistik der Bundesstaaten. Erster Theil. Berlin, 1898, reedition as *Statistik des deutschen Reichs, Volume 114.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Gewerbezählung vom 14. Juni 1895. Gewerbe-Statistik der Bundesstaaten. Zweiter Theil. Berlin, 1898, reedition as *Statistik des deutschen Reichs, Volume 115.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Gewerbezählung vom 14. Juni 1895. Gewerbe-Statistik der Verwaltungsbezirke. Zweiter Theil. Berlin, 1898, reedition as *Statistik des deutschen Reichs, Volume 118.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Berufsstatistik. Abteilung X. Berlin, 1913, reedition as *Statistik des deutschen Reichs, Volume 211.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung I. Berlin, 1910, reedition as *Statistik des deutschen Reichs, Volume 213.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung III. Berlin, 1910, reedition as *Statistik des deutschen Reichs, Volume 215.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung IV. Berlin, 1910, reedition as *Statistik des deutschen Reichs, Volume 216.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VII. Berlin, 1909, reedition as *Statistik des deutschen Reichs, Volume 219.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Berufs- und Betriebszählung vom 12. Juni 1907. Gewerbliche Betriebsstatistik. Abteilung VIII. Berlin, 1914, reedition as *Statistik des deutschen Reichs, Volume 220.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Auswärtiger Handel des deutschen Zollgebiets im Jahre 1895. 1. Theil: Der auswärtige Handel nach Menge und Werth der Waaren-Gattungen und der Verkehr mit den einzelnen Ländern. Berlin, 1896, reedition as *Statistik des deutschen Reichs, Volume 85.* Osnabrück, 1975.

Kaiserliches Statistisches Amt. Auswärtiger Handel im Jahre 1907. Darstellung nach Warengattungen. Berlin, 1909, reedition as *Statistik des deutschen Reichs, Volume 189.* Osnabrück, 1976.

Kaiserliches Statistisches Amt. Auswärtiger Handel im Jahre 1907. Durchfuhr. Niederlage-, Veredlelungsverkehr. Zollerträge usw. Seefischerei. Berlin, 1909, reedition as *Statistik des deutschen Reichs, Volume 191.* Osnabrück, 1976.

Kaiserliches Statistisches Amt. 1. Theil: Der auswärtige Waarenverkehr des deutschen Zollgebiets im Jahre 1882 geordnet nach den einzelnen Waarengattungen sowie der Veredlungsverkehr. Berlin, 1883, reedition as *Statistik des deutschen Reichs, Alte Folge, Volume 60.* Osnabrück, 1969.

Kaiserliches Statistisches Amt. Der Verkehr auf den deutschen Wasserstraßen, insbesondere der Schiffs und Güterverkehr auf den deutschen Wasserstraßen nebst den beobachteten Wasserständen im Jahre 1874. Berlin, 1876.

Kaiserliches Statistisches Amt. Der Verkehr auf den deutschen Wasserstraßen, insbesondere der Schiffs und Güterverkehr auf den deutschen Wasserstraßen nebst den beobachteten Wasserständen im Jahre 1882. Berlin, 1883.

Kaiserliches Statistisches Amt. Die Binnen-Schifffahrt im Jahre 1895. Berlin, 1897, reedition as *Statistik des deutschen Reichs, Volume 88. Osnabrück, 1975.*

Kaiserliches Statistisches Amt. Monatshefte zur Statistik des Deutschen Reichs für das Jahr 1882. *Statistik des deutschen Reichs, Volume 59, No. 2. Berlin, 1883.*

Kaiserliches Statistisches Amt. Vierteljahreshefte zur Statistik des Deutschen Reichs für das Jahr 1895. *Vierteljahreshefte zur Statistik des deutschen Reichs, Volume 5, No. 4.* Berlin, 1896.

Kaiserliches Statistisches Amt. Vierteljahreshefte zur Statistik des Deutschen Reichs für das Jahr 1907. *Vierteljahreshefte zur Statistik des deutschen Reichs, Volume 17, No. 4.* Berlin, 1908.

Kaiserliches Statistisches Amt. Die Binnen-Schifffahrt im Jahre 1907 sowie der Bestand der deutschen Binnenschiffe am 31. Dezember 1907. Berlin, 1909, reedition as *Statistik des deutschen Reichs, Volume 192.* Osnabrück, 1976.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen III. Quartal 1883. Berlin, 1884.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen IV. Quartal 1883. Berlin, 1884.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen I. Quartal 1884. Berlin, 1884.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen II. Quartal 1884. Berlin, 1884.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen im I. Quartal 1895. Berlin, 1896.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen im II. Quartal 1895. Berlin, 1896.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen im III. Quartal 1895. Berlin, 1896.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen im Jahre 1895. Berlin, 1896.

Königlich-Preußisches Ministerium der öffentlichen Arbeiten. Statistik der Güterbewegung auf Deutschen Eisenbahnen im Jahre 1907. Berlin, 1908.

Königlich-Sächsisches Statistisches Bureau. Zeitschrift des K. Sächsischen Statistischen Bureaus. Die Ergebnisse der sächsischen Gewerbezählung vom 5. Juni 1882. Volume 32. Dresden, 1886.

Königlich Statistisches Bureau in Berlin. Zeitschrift des Königlich Preussischen Statistischen Bureaus. *Volume 20.* Berlin, 1880.

Königlich Statistisches Bureau in Berlin. Die Gewerbebetriebe im preussischen Staate nach der Aufnahme vom 5. Juni 1882. *Preussische Statistik, Volume 83*. Berlin, 1885.

Königlich Statistisches Landesamt. Statistisches Handbuch für das Königreich Württemberg. Jahrgang 1899. Stuttgart, 1900.