Railways and structural change: evidence from industrializing Britain

Eduard Alvarez*, Dan Bogart†, Max Satchell‡, Leigh Shaw-Taylor§ and Xuesheng You¶

Preliminary Draft, Oct. 2016‖

Abstract

New transport infrastructure has potentially long lasting and broad effects on an economy. This paper studies the effects of railways on structural change in employment in Britain during the industrial revolution. It uses new data on the number of males employed in 12 different occupational categories across more than 9500 parishes in England and Wales at two dates 1817 and 1881. It also uses new GIS shapefiles of transport infrastructure and endowments. The main findings show that greater access to railway stations not only increased local employment growth, it also led to structural change in employment. Specifically greater railway access decreased the share of occupations in agriculture, and increased the share in some occupations like mining, transportation, machine tools, textiles, and clothing. The effects of railways were also larger in parishes with coal or those with high population density in 1801. The results are consistent with a mechanism in which railways reinforced natural resource advantages and agglomeration economies and contributed to the spatial divergence in employment that was already underway in the early stages of the industrial revolution.

Keywords: Structural change, railways, endowments, spatial development

JEL Codes: N4

*Visiting Scholar, Department of History, University of Cambridge, edu.alvarez.palau@gmail.com

†Associate Professor, Department of Economics, UC Irvine, dbogart@uci.edu

‡Research Associate, Department of Geography, University of Cambridge max.satchell@geog.cam.ac.uk

§Professor, Department of History, University of Cambridge lmws2@cam.ac.uk

¶Research Associate, Department of History, University of Cambridge max.satchell@geog.cam.ac.uk

‖This data for this paper was created thanks to grants from the National Science Foundation and the Leverhulme Trust. We thank seminar participants at UC Irvine. All errors are our own.

1 Introduction

Discussion of transportation improvements often centers around their benefits and costs, but their long-term effects are often difficult to measure. Economic history offers insights here precisely because the long-run is observed. However, the impact of transport infrastructure on key economic changes in history, like the industrial revolution, are not well established. This paper examines how railways affected male occupational growth and structural change in nineteenth century England and Wales.

The transition from agriculture to manufacturing and services employment is one of the key features of the British industrial revolution. Recently collected data for England and Wales suggests the initial transition (in the seventeenth and eighteenth century) was largely from agriculture to manufacturing, followed by a second transition in the nineteenth century largely from agriculture to services and mining. Shaw-Taylor and Wrigley (2014) estimate that 50% of the male labor force worked in agriculture in 1710. By 1817 agricultural employment fell to 36% and by 1871 it was 19%. Over the same period male secondary (manufacturing) employment rose from 37% in 1710 to 44% in 1817 to 46% in 1871. Male tertiary (services) employment increased from 12% in 1710 to 18% in 1817 to 28% in 1871. Mining employment rose from 1% to 3% to 6% over those periods. The implications of structural change were significant. Labor productivity was typically higher in mining, manufacturing and services, and thus the overall productivity of the labor force increased due to the reallocation of occupations.

In transport, there were revolutionary changes coinciding with structural change in employment. Britain got better roads, canals, and ports during the eighteenth century, and then in the nineteenth century, steamships and railways were introduced. The first railway line was constructed in Britain in 1825. Network growth was slow until 1840 when the railway mania led to a huge expansion in stations and mileage. In 1840 Britain had 1857 railway miles and by 1850 and 1860 it had 6621 miles and 10,433 miles respectively. In subsequent decades, network growth was more steady. In 1880 Britain's railway network was 17,933 miles and in 1900 it was 21,855. By this date Britain had one of the most dense networks in the world. The system was also rare in that railways were privately financed, owned, and operated up to the mid twentieth century. Most countries had significant government ownership or subsidies during the formative phases on network expansion.[1]

It is clear that railways contributed to GDP because they dramatically lowered inland transport costs (Hawke 1970). As an illustration freight rates by rail in 1865 were one-tenth the freight rates by road in 1800 when converted into constant prices (Bogart 2014). Railways also had the potential to change Britain's occupational structure and its spatial distribution. As railways increased market access, one hypothesis is that areas close to railway stations became more attractive locations for workers and firms, increasing their total employment. Related hypotheses are that agricultural occupations declined as a share of total occupations as station access increased, while the share in secondary and tertiary sectors increased. One mechanism is that population increase near stations meant higher land rents, and thus higher production costs in land intensive sectors like agriculture. Moreover, agglomeration economies in secondary and tertiary sectors would imply more concentration of certain secondary and tertiary employment near stations if total employment rose. Occupations that made use of steam engines or other innovations of the early nineteenth century are the most likely to be affected by agglomeration economies.

New data are available to examine these effects of railways. For employment we observe male occupations in more than 10,000 parishes and townships in England and Wales at two dates 1817 and 1881. The occupational data are linked to a shapefile of consistent jurisdictional units in 1817 and 1881. In most cases, the units are parishes and townships. In a minority of cases the units are aggregations of parishes because of subdivisions in territory between 1801 and 1881. For simplicity we refer to jurisdictional units as parishes. From the

[1]For an analysis of Britain's network and its ownership and regulatory structure see Casson (2009). For an international perspective on state ownership of railways see Bogart (2010).

raw data we construct parish-level variables for the growth of employment between 1817 and 1881, and the change in the shares of employment in 12 occupational categories between 1817 and 1881. For railways we use shapefiles of the network, which include the location and the date of opening for all stations.[2] The station data is linked to the shapefile of consistent parish units in the nineteenth century. Railway station access is then measured in two primary ways: (1) an indicator for whether the parish has a railway station in its boundary by 1881 and (2) the distance from the center of the parish to the nearest railway station in 1881. Other control variables are also included such as whether the parish has coal, whether it is coastal, its ruggedness, indicators for soil types, and its proximity to waterways, ports, and turnpike roads. Another set of controls include the population density of the parish in 1801 and its population growth from 1801 to 1821 both of which capture initial structure and pre-railway growth trends.

The data are analyzed using similar regression approaches applied to other studies of transport and development (see Redding and Turner 2014 for an overview). The main findings show that greater access to railway stations not only increased local employment growth, it also led to structural change in employment. Specifically greater railway access decreased the share of occupations in agriculture, and increased the share in mining, transportation, machine tools, textiles, and clothing. Most of the latter occupations are notable for being influenced by technology, and therefore more subject to agglomeration economies. In terms of magnitudes railways explain only part of the variation in employment growth and structural change. The standardized coefficients range between 0.06 and 0.19. Moreover, in some occupations having railway stations is estimated to increase employment growth by the same amount as having coal deposits. But distance to stations generally has a larger effect than distance to waterways or to turnpike roads, indicating railways were a significant

[2]See del Río, Martí-Henneberg, and Valentín (2008) for an initial description of the railways shapefile data. Additional upgrades were produced by the Cambridge group for the history of population and social structure (CamPop), see http://www.campop.geog.cam.ac.uk/research/projects/transport/data/railwaystationsandnetwork.html.

upgrade over pre-1830 transport infrastructures.

In any study of transport access there are natural questions about endogenous placement of infrastructure. In this case, the most relevant concern is that parishes experiencing employment growth and structural change for unobservable reasons were also the places getting access to railways. We address the omitted variable bias issue in several ways. First, we examine the sensitivity of the results to dropping various controls for endowments, pre-exisiting transport infrastructure, and population growth in the early nineteenth century. The coefficient for railway access is stable in the models for total employment growth and the share of agricultural occupations. There is some sensitivity of the coefficient in the models for the share of secondary and tertiary employment, suggesting some concerns over omitted variables.

As a second approach, we instrument for distance to railway stations using the inconsequential place approach (see Redding and Turner 2014). Specifically, we construct a hypothetical railway network using the major towns and boroughs of the early nineteenth century and identify the least cost paths (LCPs) that minimized elevation changes and distance. The estimated effects of railway access are still significant and precisely estimated for total employment growth and the share in agriculture, mining, transportation, textiles, and clothing. The overall conclusion is that railway access clearly contributed to employment growth and structural change in agriculture, and some sectors in secondary and tertiary. As we show in extensions, the effects of railways were larger in parishes with coal or that were already dense in 1801. Our interpretation is that railways reinforced natural resource advantages and agglomeration economies already present in the early nineteenth century. They contributed to the spatial divergence in employment already underway in the early stages of the industrial revolution.

The findings contribute to several literatures. The first addresses the causes of structural change in Britain and more generally in advanced economies during the nineteenth century.

Leading explanations center around market access and endowments like coal (Wrigley 2010, Fernihough and Hjortshøj O'Rourke 2014), while others have emphasized education and finance (Becker, Hornung, and Woessmann 2011, Heblich and Trew 2015). In the British context, most of these theories have not been adequately tested using comprehensive micro-level data on employment. Existing studies focus on English and Welsh counties or regions as the unit (Crafts and Mulatu 2006, Kelly, Mokyr, Ó Gráda 2015) but counties and regions are relatively large and cover a wide range of industries. Some studies focus on a single sector like textiles (e.g. Crafts and Wolf 2014), but cannot account for other industries.

The economic effects of railways are widely studied in the literature. Early works focused on social savings and the direct benefits of lower transport costs (Fogel 1964, Fishlow 1965, Hawke 1970). Recent works analyze their effects on population density and agricultural income.[3] Two related studies to ours, Crafts and Mulatu (2006) and Gutlberlet (2014), examine the effects of falling transport costs on regional employment structure. Crafts and Mulatu (2006) are especially notable as they argue that falling transport costs had small effects on the location of British industry from 1871 to 1911. Our study is different because it uses more dis aggregated data (parishes) and it analyzes the effects of railways from their beginnings up to 1881. Our estimates generally support the view that railways mattered but perhaps to a lesser degree than one might expect. The concentration of secondary employment in places like Manchester was largely the result of processes that predate railways. Railways were crucial in the concentration of transport employment, but not other forms of tertiary employment.

Our paper also contributes to a more general analysis of spatial organization and development. Several theories emphasize the role of endowments and market access as determinants of economic density (e.g. Redding and Turner 2014), while others emphasize the differential responses of services and manufacturing to changes in transport costs (e.g. Desmet and

[3] For studies on the effects of railways on population density or income see Donaldson (2014) for India, Donaldson and Hornbeck (2016) for the US, and Casson (2013) and Alvarez et. al. (2013) for Britain.

Rossi-Hansberg 2015). We contribute by offering a sector-level estimate of greater market access. Our findings suggest that outside of transport, services were less responsive to changes in market access compared to the secondary sector. It is likely that services were not as susceptible to agglomeration spillovers in the nineteenth century context compared to today where information and communication technology have radically altered the location of employment.

2 Data

The paper makes use of new data analyzed in Geographic Information Software (GIS). The occupational data come from a large research project undertaken at the Cambridge Group for the History of Population and Social Structure (CamPop).[4] As of 1813 it was a legal requirement that fathers' occupations be recorded in all Anglican parish registers when their children were baptized. Current demographic evidence suggests that at this date fertility differences between major occupational groups were limited. This suggests that counts of occupations derived from baptism registers should provide a good picture of adult male occupational structure. Accordingly, CamPop collected data from virtually every parish register in England and Wales for an eight-year period (1813-1820) to create a quasi-census of male occupations. This exercise made use of 11,364 baptism registers and resulted in a data set with c.2.65 million observations (see Kitson et. al. 2012). For convenience the data set is described as referring to c.1817, the approximate mid-point of the period. Starting in 1841 the censuses provide data on a wide range of male occupations, and this paper makes use of the 1881 census in which occupational data has been digitized for every parish.

The occupational data for 1817 and 1881 are linked to GIS using a consistent set of parish boundaries produced by CamPop. Parish-level variables are created for total male employment and the share by occupational grouping in 1817 and 1881. In the analysis be-

[4]See http://www.campop.geog.cam.ac.uk/research/occupations/ for more details on the project.

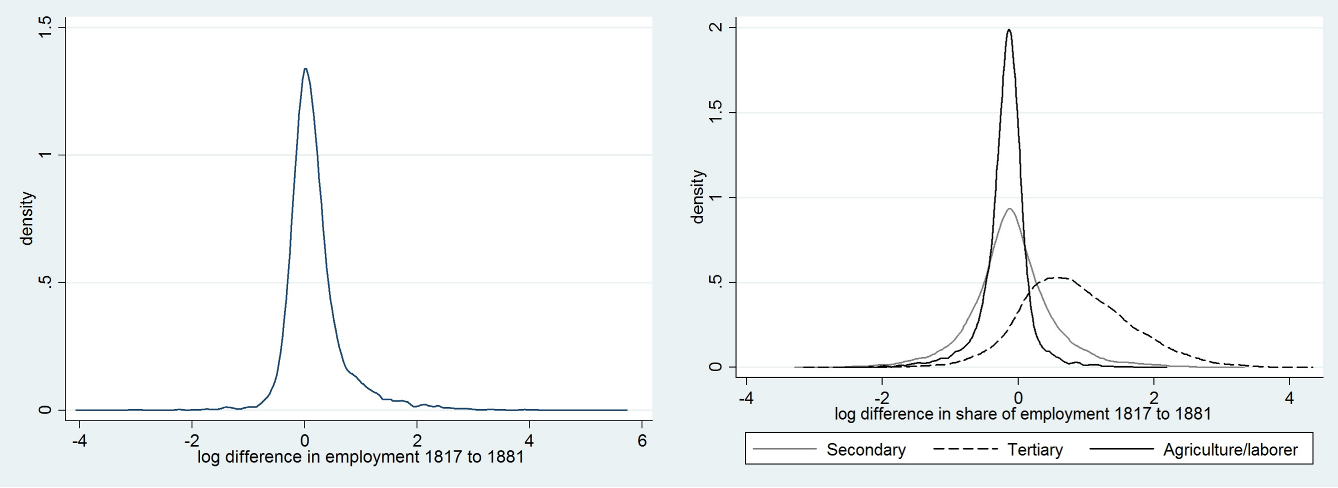
low we mainly focus on changes in the share of male employment in various groups within a parish unit between 1817 and 1881. There are two levels of employment groups. The first is the most general and includes (1) agriculture & labourer, (2) secondary, and (3) tertiary occupations. Agriculture and labourers are grouped together under the assumption that most labourers worked at least part time in agriculture. Secondary refers to the transformation of the raw materials produced by the primary sector into other commodities, whether in a craft or a manufacturing setting. Tertiary encompasses all services including transport, shop-keeping, domestic service, and professional activities. The codings are based on narrower occupational groups in the PST system detailed below.[5]

Kernel density estimates of these distributions in the main occupational groupings are shown in figure 1 along with the distribution for total employment change measured as the log difference in male employment between 1817 and 1881. The mean for log difference in employment is 0.18 indicating an average employment growth of 18 log points. Notice also that some parishes lost employment between 1817 and 1881, and thus were depopulated. The distribution for the log difference in shares is different across the three main occupational groupings. The average was highest in tertiary followed by secondary and agriculture. In other words, more parishes gained tertiary employment as a share of total employment than in other groupings. Dispersion was greatest in tertiary, followed by secondary and then agriculture.

The second level of occupational groupings is more narrow and includes the share of males employed in 12 sub-categories. The sub-categories include (1) Agriculture, (2) Labourers, (3) Mining, (4) Building and construction, (5) Machine tools, (6) Iron & Steel, (7) Textiles (8), Footware, (9) Clothing, (10) Transport, (11) Professions and services, and (12) Dealers and sellers. The specific occupations in each sub-category are shown in table 1, parts 1 and 2. Together these categories comprise most occupations in the secondary and tertiary

[5]For more details see the PST system documentation, http://www.campop.geog.cam.ac.uk/research/projects/occupatio

Figure 1: Distribution of employment growth and structural change from 1817 to 1881



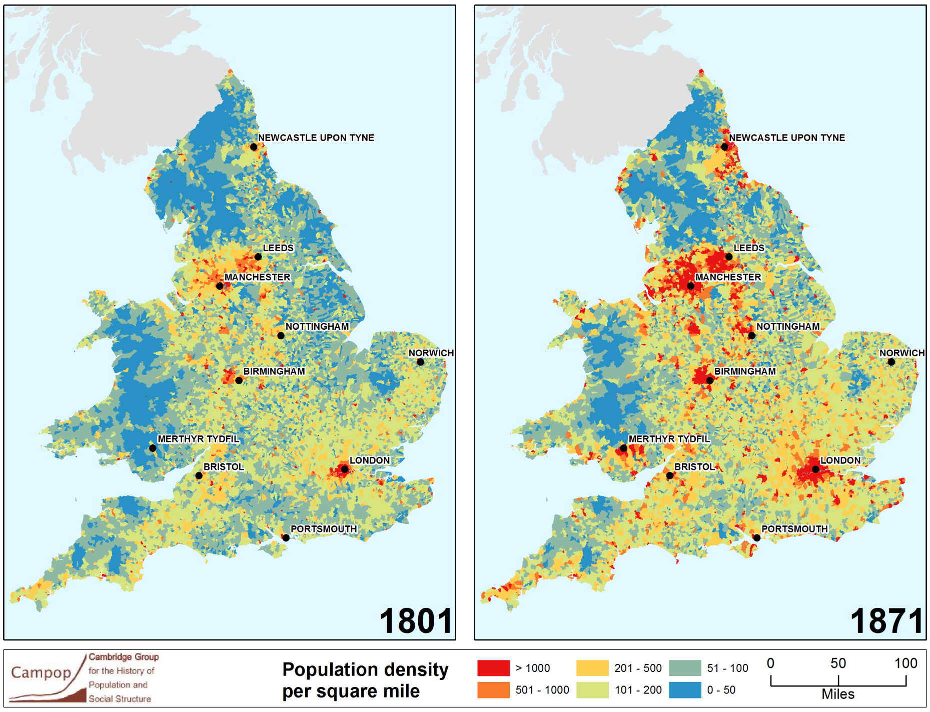
Sources: see text.

categories. The narrower classification also allow the primary sector to be dis aggregated into agriculture and mining. Also they do not assume labourers worked in agriculture.

Data on the distribution of population is also crucial to understanding the effects of railways. Population densities per square mile are displayed geographically in figure 2 for 1801 and 1871, two dates close to those in our analysis. The largest cities of 1801 are also shown. Population density was higher around the major cities in both periods as one would expect. What is most remarkable is that the geographic distribution of population density was fairly similar in 1801 and 1871. The population centers were in the northwest around Manchester and in the southeast around London. There were changes of note in the nineteenth century. Population density increased more near major cities between 1801 and 1871. The growth was especially higher in the hinterland of the major towns.

Secondary employment shares are shown geographically in figure 3. There was a high degree of specialization between Manchester and Leeds in the northwest and around Birmingham in 1817. The same was still true in 1881, although secondary employment became even more spatially concentrated around these cities. This map highlights a key finding noted by many previous authors including Shaw-Taylor and Wrigley (2014). The spatial

Figure 2: Population Density in England and Wales



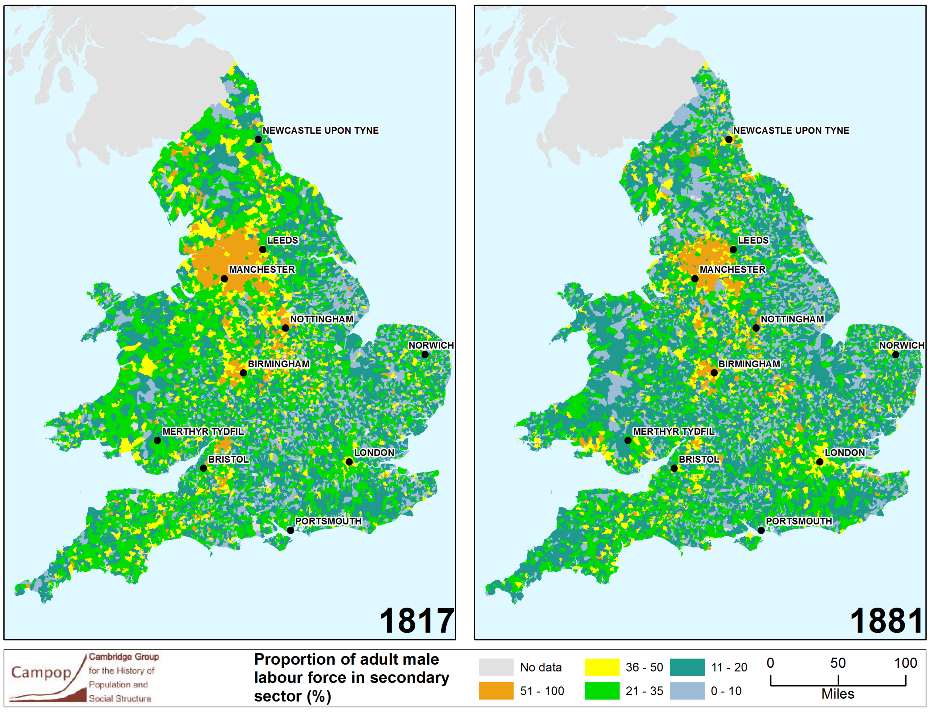
concentration of secondary employment was high already in the early nineteenth century, and during the railway age secondary concentration accelerated. Some authors have gone further in suggesting that the industrial revolution has its origins in some earlier period, and was not caused by railways.

There is a different story for the tertiary sector. The geographic distribution is shown in figure 4. Tertiary shares were generally low throughout England and Wales in 1817 and concentrated near London. By 1881 tertiary is more common everywhere, but especially in the north, and near the large manufacturing towns of Leeds, Manchester, and Birmingham. It is also remarkable that by 1881 tertiary employment became concentrated in similar areas as the secondary sector. Thus services and manufacturing employment tended to co-locate.

The paper also uses new GIS data on British transport networks.[6] Most importantly the data include the spatial location of railway lines and rail stations with their opening dates

[6]See http://www.campop.geog.cam.ac.uk/research/projects/transport/ for more details on the project.

Figure 3: Shares of males in secondary in England and Wales



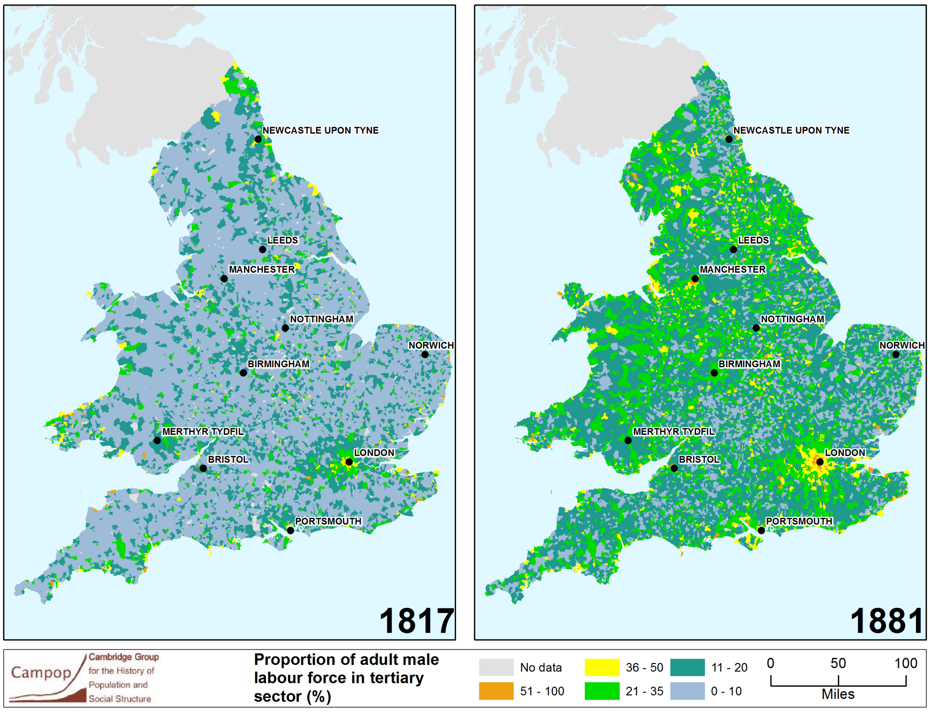
Sources: see text.

and closing dates by 1881. Figure 5 shows railway lines in 1851 and 1881. The network of 1851 was largely conceived in the railway mania of 1844 when hundreds of lines were proposed. The railway mania has been described as an irrational event (Odlyzko 2010), but it did bring railway connections to all major cities, even if many other medium and small towns were still not connected.

By 1881 the network had grown much larger. Now all major towns had connections not only with each other, but also with smaller towns in their region. The British rail network had essentially matured.

We combine the data on parish units with the rail GIS to identify whether parishes had stations and to calculate the distance between the center of the parish and railway stations. The center of the town is defined in two steps. If the parish had a market town at some

Figure 4: Shares of males in tertiary in England and Wales

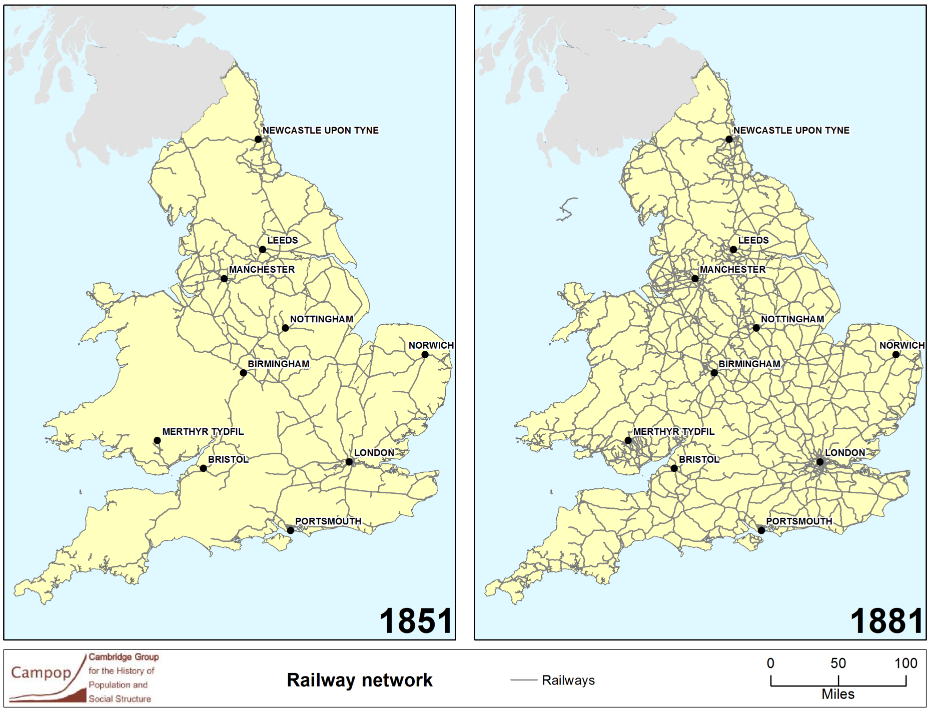


Sources: see text.

point between 1600 and 1850 then the market town is taken as the center. This applies to 1425 parishes out of 9500. If the parish had no market town, then it was likely to be rural and the centroid is taken as the parish center. The area around Birmingham in figure 6 provides an example. The red line shows the town footprint of Birmingham in 1891, the purple circles are the parish centroids, the yellow circles are the railway stations, and the green lines represent the distances to the nearest station. As another variable we identify the distance between the parish center and the second nearest station. We then average over the first and second distance to incorporate density of nearby stations.

The analysis below focuses on the effects of railways on occupational structure, but clearly railways were not the only factor. We create several control variables to capture some of these factors, especially as some were related to railway access. First, we create a set of controls for pre-railway population density, population growth, and occupational structure. The population density variable is equal to the log of the parish's 1801 population divided

Figure 5: Railway lines and stations in 1851 and 1881

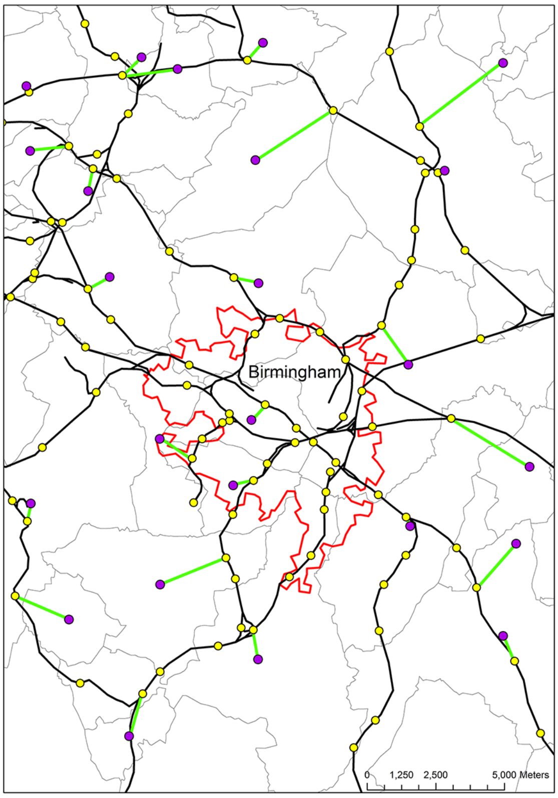


Sources: see text.

by its land area. The population growth variable is equal to the log difference in parish population between 1821 and 1801. The pre-railway occupational structure is simply the share of secondary and tertiary occupations in 1817.

The second set of controls includes indicators for natural endowments. Using GIS data on endowments, we identify whether parishes overlapped with exposed coalfields. Exposed coalfields are those where coal bearing strata are not concealed by rocks laid down more recently. In Britain coal is predominantly found in rocks laid down during the Carboniferous Period. The GIS does not capture a handful of tiny post carboniferous coal deposits, such as that at Cleveland (Yorkshire) which was worked in the 19th century. The omission of these is not a problem because they were of only of local importance. Other parish-level endowment data include overlap with the coast, the average elevation, the ruggedness of parishes measured by the average slope to nearby cells in the parish, and finally 27 categories

Figure 6: Birmingham area example



Sources: see text.

for soil types.[7]

The third set of controls is based on pre-railway transport networks. These include ports in 1830, waterways in 1830, and turnpike roads in 1830.[8] Similar to railway stations, variables are created for distance to ports, waterways, and turnpike roads in 1830. The fourth set of controls include county indicators. There are 54 counties in England and Wales, and each gets its own dummy variable.

Table 2 reports summary statistics for the main dependent and explanatory variables. Note the 27 indicators for soil types are omitted as are the 54 dummies for counties. The summary statistics reveal two interesting facts: (1) 25% of parish units had a railway station

[7]See http://www.campop.geog.cam.ac.uk/research/projects/transport/data/ for more details on the endowment data.

[8]See http://www.campop.geog.cam.ac.uk/research/projects/transport/ for more details on the network data.

and (2) the average distance from a parish center to the nearest station was 3.8 km. It is clear that Britain had a very dense network of stations by 1881.

3 Baseline analysis

It is instructive to begin with a difference in difference analysis where parishes are sorted into two groups depending on whether they had high or low access to railways. We should caution readers that the standard parallel trends assumption in diff-n-diff analysis does not hold here. Parishes getting greater access to railways were already experiencing population growth, and perhaps structural change, prior to railways. Therefore we view these preliminary diff-n-diff results as suggestive but not definitive. Panel A in table 3 shows that the log difference in population and employment were higher in parishes with greater railway access by 1881. For example, secondary employment in parish units with railway stations was 47 log points higher than parishes without railway stations. Similarly parishes with below median distance to their nearest station had 31 log points higher secondary employment than parishes with above the median distance. Panel B in table 3 shows that railway access has a different association with employment shares in agriculture and labour compared to secondary and tertiary. For example, on average parishes with railway stations had 14.7 log points lower share of agriculture and labourer, but they had 3.5 and 9.5 higher log point shares in secondary and tertiary.

Panel C in table 3 repeats the exercise examining the association between railway access and employment shares in the 12 sub-categories. Here the shares are not reported in log differences but rather standard differences. Also the shares are examined only if the parish unit had a positive share in either 1817 or 1881. The reason is that some parishes gained a occupation sub-category for the first time in 1881 and some entirely lost the occupational sub-category by 1881. The log difference requires a positive share, and so all these parishes would be dropped if log differences were used. For each sub-category in Panel C the mean employment share in 1817 is reported as long as it was positive. The number of parishes with a positive share in 1817 is also reported. If a sub-sector's sample size is larger in the difference in means analysis (see machine tools for example) then this implies that many parishes added that particular sub-category by 1881. The preliminary results show that employment shares in agriculture, textiles, footware, clothing, and professions decline significantly with greater railway access. The rest are mostly positively and significantly associated with greater railway access. The effects differ in size, but most represent a change between 10 to 33 percent of the mean share in 1817. For example, parishes with railway stations had a 0.003 higher share of employment in machine tools, which is 18.75% of the mean share in 1817.

Differences in means are suggestive of the effects, but it is obviously important to check whether the same patterns hold after controlling for other variables. Our first such analysis involves OLS regressions of the log difference in employment or the log difference in employment shares on variables for railway access plus controls. The following equation describes the specification:

$$\Delta logemployment_{ij} = \alpha + \beta_1 rail_{ij} + \beta_2 x_{ij} + \beta_3 countyFE_j + \varepsilon_i$$

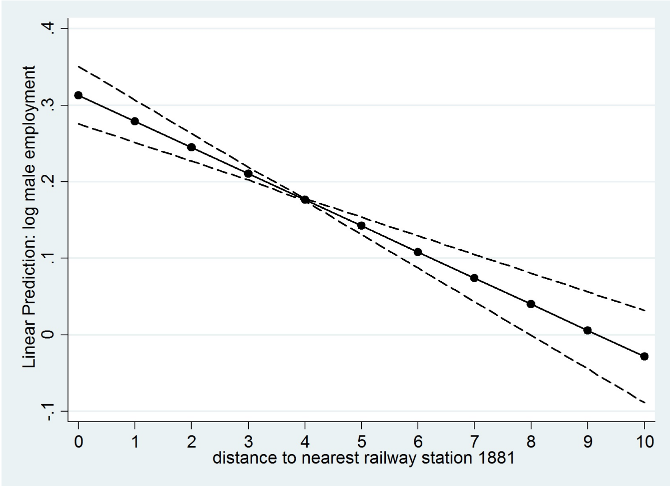
where $\Delta logemployment_{ij}$ is the log difference in total employment or the log difference in the employment share in parish i in county j, $rail_{ij}$ is one of the two measures of railway access in parish i in county j, x_{ij} includes controls for parish endowments, pre-railway economic structure, and pre-railway transport infrastructure all shown in table 2, $countyFE_j$ is a vector that includes indicators for 54 counties in England and Wales, and finally ε_{ij} is the error term. For the moment the identifying assumption is that rail access is exogenous conditional on parish controls and county FE. We will relax this assumption later. Note also that the standard errors are clustered on counties in all specifications.

The results for the baseline regression model are shown in table 4. Panel A reports results where rail access is defined by whether a parish has a railway station. Employment is higher in parishes with railway stations, and again we see that employment is especially higher in secondary and tertiary occupations compared to agriculture and labourers. The effect of railways on occupational shares shows the same patterns. The share in agriculture & labourer is 16 log points lower in parishes with railway stations, the share in secondary is 16 log points higher, the share in tertiary is 15 log points higher. The estimates have the same signs as the difference in means analysis, but the magnitudes are different indicating that adding some controls matters. For example, the effect of railway stations on the share in the secondary sector is larger after controlling for endowments and pre-railway population growth.

Panel B reports results using distance to stations in km as the main rail variable. Here the signs are the opposite from panel A as greater distance implies lower access to railway stations. In all cases the coefficients are statistically significant. The magnitudes are best represented by the standardized coefficients (the share of a standard deviation in log employment associated with a one standard deviation change in distance to stations) towards the bottom of panel B. The standardized is largest for log employment (-0.19) and the log share of agriculture (0.169) and smaller for the log share of secondary (-0.067) and tertiary (-0.083). In other words railway access explains more variation in total employment and agricultural structure than shares of secondary and tertiary.

The estimates show that employment and hence population increased more in parishes whose nearest railway station was closer to zero. But did greater distance to railways stations make parishes lose employment and population? The predicted change in log male employment for 0 to 10 km distance to the nearest station is shown in figure 7. Of most interest is the predicted change in parishes distant from railway stations, say 10km. Here the predicted change is indistinguishable from zero suggesting that poor railway access did

Figure 7: The marginal effects of distance to railway stations



Sources: see text.

not systematically lead to depopulation or absolute employment loss.

Another way of evaluating the magnitudes it so compare the railway access coefficients with some control variables of interest. Panel A in table 5 compares the coefficient for the railway station indicator to the coefficients for the coal and coastal indicators. The coastal coefficient generally has the same sign as the railway coefficient but is smaller in magnitude. The coal coefficient is also similar to railway stations in sign, but its relative magnitude varies. Coal has a similarly-sized effect on total employment and agricultural shares, but coal has a smaller effect on the secondary share. Thus while the impacts differ to some degree, coal and railways were of similar importance for the occupational changes of the nineteenth century.

Panel B in table 5 compares the coefficient for distance to the nearest railway station with the coefficients for distance to the nearest waterway and nearest turnpike road. Distance to railway stations has a larger effect on total employment, agricultural shares, and tertiary shares than distance to turnpike roads and waterways. But interestingly, distance to turnpike roads has a larger effect on the secondary share than distance to railway stations. This suggests that pre-railway transport innovations played an equally if not more important role in determining the occupational structure of the secondary sector.

An analysis of the occupational shares in secondary and tertiary may mask significant heterogeneity across narrower occupational groups. Our analysis of the twelve occupational sub-categories will identify whether this is the case. The sub-categories also point to potential channels as occupations differ in the importance of transport costs and the degree of agglomeration economies. The results in table 6 show the effects of distance to nearest railway station on the change in the employment share for each of the 12 sub-categories. The standardized coefficient is also reported to assess magnitudes. A large positive effect is found for the agricultural share implying that greater distance to stations increased the share in agriculture. A large and significant negative effect is found for mining, machine tools, and transport implying that greater distance significantly reduced the share of these occupations. A statistically significant but smaller negative effect is found for labourers, iron and steel, and textiles.

The result for transport shares in table 6 is not surprising as railways were themselves large employers. The effect for mining probably reflects the importance of transport costs in shipping heavy minerals like coal. It could also reflect agglomeration economies because mining relied on steam engine technology. The large effect for machine tools also points to a role for agglomeration economies as this was the 'high-tech' industry of the nineteenth century. Agglomeration is also the most likely channel explaining the effect of railway access on textile manufacturing. Textiles had high value to weight and the sector was continually evolving with improvements in steam power during the nineteenth century.

More insights can be gained by looking at when improvements in railway access affected outcomes by 1881. As noted in the introduction, the railway network expanded significantly in the mid nineteenth century following the railway mania of the 1840s. Across parish units, the average distance to the nearest railway declined from 36.3 km in 1841 to 10.5 in 1851. There were subsequent expansions in the network in the 1860s and 1870s which are also notable. In 1861 the average distance to the nearest station was 6.11 km and in 1871 the average was 4.5 km. One might imagine that the formation of the early network had the largest effect on structural change. Alternatively the process might have been more incremental with structural change reacting to each improvement in railway access over time.

Table 7 reports a series of regressions that address the timing of effects. The first reported in row A includes the distance to the nearest railway station in 1851 along with all the controls used in previous regressions. The second, third, and fourth in rows B, C, and D include the distance to the nearest railway station in 1861, 1871, and 1881 respectively. The key comparison is between the coefficients for station distance in 1851, 1861, 1871, and 1881, and how large the former are to the latter. The results suggest that the incremental model applies to most occupations groupings. For example, agricultural shares by 1881 increase marginally with initial distance to stations in 1851, a bit more with distance to stations in 1861, and the same for distance to stations in 1871 and 1881. The only sector where early railway access seems to have a lasting effect is iron and steel. The coefficient for distance to railway stations in 1861 is equivalent to three-fourths of the coefficient for distance to railway stations in 1881. Another notable results concerns labourers. Distance to stations has little effect prior to 1881. We think this reflects labourers role in building railway stations, and is not related to railways permanently attracting labourers to a parish.

4 Endogeneity of railway access

One of the maintained assumptions in the previous models is that railway access is exogenous after including controls. This assumption is fairly strong however, and thus the estimates for railway access may be biased. One argument is that railway companies were forward looking and placed railway stations in parishes that had greater growth potential. Another argument is that along their route, railway lines were put in some parishes because they had lower economic activity and hence the railway companies could obtain land at a lower price (rights of way were weak and gave a lot of power to landowners). Moreover the railway companies were pressured by politicians to put railway lines and stations in their constituency, perhaps more so if the constituency had low growth potential.

There is some evidence to bear on these concerns. First, we have the population of all parish units from 1801 to 1901, and the dates when they got their first station. Thus we can test whether population growth was increasing in a parish in the decades prior to getting a railway. The answer is that population was rising in parishes prior to getting their first station, on the order of 10 to 20%. Second, a simple correlation test shows that a parish unit's distance to railway stations in 1881 is correlated with other developmental variables, most notably the log of population density in 1801 and parish population growth from 1801 to 1821. In short, the data suggest that endogeneity is a concern.

To address these and other endogeneity issues, we follow two approaches. First we analyze the stability of the coefficients after successively adding the control variables. Secondly we propose an instrumental variable for railway access based on geography and location between major towns. The first method focuses on the sensitivity of the estimates after adding controls. A more stable coefficient adds confidence that the estimates are robust. Panel A in table 8 reports results for distance to the nearest railway station after including only the endowment variables (coal, coastal, ruggedness, elevation, and soil types) as controls. Panel B adds the controls for pre-railway employment shares in secondary and tertiary, along with controls for the log of population density in 1801, and the log difference in parish population growth from 1801 to 1821. Panel C adds the county fixed effects and the controls for pre-exisiting transport infrastructure (distance to nearest port, turnpike road, and inland waterway). The coefficients for railway access are quite stable in the models for total employment and agricultural shares. Thus there is less concern about omitted variables in these specifications. However, the coefficients for railway access are more sensitive in the models for secondary and tertiary shares. Here one could argue that omitted variables are more concerning. One solution is to use an instrumental variable for railway access, to which we now turn.

Our instrumental variable is based on the inconsequential place approach. The idea is that railways were designed to link major cities and municipal boroughs who had representatives in the House of Commons. In considering their routes, railway companies (and the government) tried to minimize the construction costs, while still connecting their desired cities. As a consequence many parishes, which one could call inconsequential places, got railways simply because they were on the cost minimizing route.[9]

The main criteria used to plan linear projects has always been the minimization of earth-moving works. Terrains with higher slopes are those in which more earth-moving is required and, in consequence, their construction costs will be higher. According to the literature, increases in tenths of slope lead to great differences in the construction costs. Engineers like Wellington (1877) referred to the impossibility of building railways with slopes over 2%. The power of traction of the locomotives and the potential adherence between wheels and rails could be the main reason. Besides, it is also important to highlight that having slopes over 2% might imply the necessity of building tunnels, cut-and-cover tunnels or even viaducts. The perpendicular slope was also crucial. During the construction of the track section, excavation and filling have to be balanced in order to minimize provisions, waste and transportation of land. Nowadays this method is carried by bulldozers and trailers, but historically it was done manually by workers.

We are examining several approaches with the aim of drawing the least cost railway connections between certain nodes of the territory. For the moment we use a simple model, in which the construction costs of building the next 100 meters of railway line from any point is proportional to slope. We have used ESRI least-cost-path python schema in order to run the spatial analysis across using the SRTM 90 elevation raster of England and Wales,

[9] The inconsequential places approach has been used in other papers see Faber (2014) and Redding and Turner (2014).

which specifies elevation in 90 meter cells. The tool calculates a least cost path (LCP) from a destination point to a source.

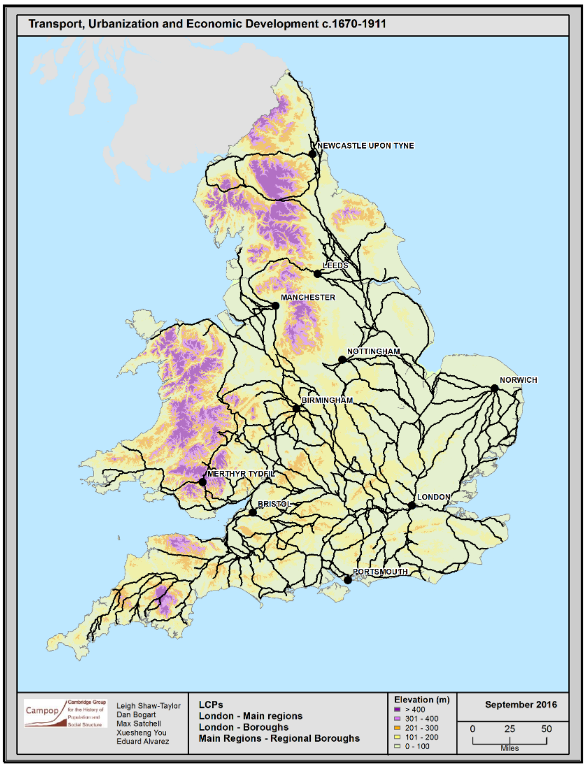
In creating least cost paths (LCPs) there is choice of which locations to connect via a path. Some works in the literature use planned networks to identify the cities to be connected (e.g. Faber 2016). In the case of British railways, there is no obvious planned network. Thus we propose several networks. First, there was clearly a need to connect London to provincial political centers via rail. Towns represented in the House of Commons through municipal boroughs are natural political centers to consider. Our first set of LCPs connect London with all municipal boroughs.

Second there was a clear desire to connect London with the major regional cities. We divided England and Wales into nine regions, and identified the most populated cities of each region in 1801, including Portsmouth in the Southeast, Bristol in the Southwest, Merthyr Tydfil in Wales, Birmingham in the West Midlands, Nottingham in the East Midlands, Norwich in East Anglia, Manchester in the Northwest, Leeds in Yorkshire, and Newcastle in the North. We calculate LCPs between London and these 9 regional cities. The largest regional cities also needed to be connected to political centers in their region, and so we calculate LCPs between these cities and municipal boroughs in their region. The LCPs between London and municipal boroughs and the LCPs between London and regional cities and regional cities to boroughs are shown in figure 8 along with elevation levels. The LCPs tend to follow valleys especially in the North, Wales, and Southwest.

Third there was an economic rationale to connect any pair of cities if they were large and close to one another in physical distance. We use a population gravity model to calculate the value of connecting all English cities, and then we consider all values above a threshold. We chose a high threshold so that only connections of value are considered. We then identified the LCPs connecting towns from the gravity model.

The main purpose of the LCPs is to create an instrument for distance to railway stations.

Figure 8: Least Cost Path Instruments connecting London to regional cities and boroughs



Sources: see text.

We construct three instruments based on the distance in km between the parish center and each of the three LCPs. If a parish is close to the LCP then it is more likely to have a railway station or to be close to a railway station. The instrument is more powerful for the distance to railway stations because stations were distributed relatively evenly across the railway line, and thus a town may be close to a station in its neighboring parish, but not actually have a station within its boundary. Therefore below we focus on the effects of distance to railway stations.

The IV estimates for total male employment and the shares in the main occupational categories are reported in table 9. Panels B, C, and D show estimates using each instrument on its own. The F-stats for the first stage are high in all cases suggesting that distance to the LCP network is generally a strong instrument for distance to railway stations. Most of the results show a negative and significant effect on total employment and a positive and significant effect on the share of agricultural employment. The effect on the share of secondary and tertiary employment are generally negative but not always significant. Thus there is a range of estimates. In terms of magnitudes, the IV coefficients for total employment, agricultural shares, and secondary shares are similar to the OLS coefficients shown in panel A. The IV coefficient for tertiary is much larger in a few cases.

Panels E, F, and G in table 7 report IV estimates using two of the three instruments. The over-identification test is of particular interest in these specifications as it gives an indication whether the exclusion restriction for the instruments is satisfied. The model combining the LCP between London and municipal boroughs and the LCP between London and regional cities, and regional cities to boroughs suggests a rejection of the exclusion restriction in the shares regressions, casting doubts on at least one of these instruments. Fortunately, the exclusion restriction is not rejected for agricultural shares in Panel F which uses the LCP between London and regional cities, and regional cities to boroughs, along with the LCP between all cities gravity model. The estimates in this case are three times larger than OLS but remain positive and significant. The exclusion restriction is also not rejected in panel G for secondary shares. The estimates imply a negative but statistically insignificant effect of railway access on the secondary share. However the IV coefficient is very similar to OLS, suggesting it is broadly informative. In no model are the exclusion restrictions satisfied for the tertiary share. Here it is difficult to draw more definitive conclusions. Overall the IV results indicate that railway access increased total employment and it encouraged parishes to alter their occupational structure, especially with fewer workers in agriculture. The share in secondary may be positively affected by greater railway access, but the precision of the estimate is not high.

We also used the instruments to further explore the effects of railways at the occupational sub-category level. In choosing instruments, our general rule was to use all three unless the over-identification test suggests the exclusion restriction is rejected. In that case, we examine specifications using two of the three instruments. Fortunately in all those cases, at least two instruments were identified where the exclusion restriction is not rejected. We focus on those specifications as they are more credible.

Table 10 shows the IV estimates along with the OLS estimates for comparison. Greater distance to railway stations has a positive and significant effect on the agricultural share, and a negative and significant effect on mining shares, textiles shares, clothing shares, and transport shares. The estimates are generally consistent with OLS except that the effect on machine tools share is much larger, although not precisely estimated.

As an extension we also estimated the IV model after dropping all parishes in the top quartile of population density in 1801. A concern is that some nodes in the LCP network had characteristics that made them more likely to experience structural change. In short, they were not inconsequential places. Presumably the nodes in the LCP are the areas with higher population density in 1801, therefore dropping them from the model addressing this concern. The results are not reported to save space. They yield nearly identical results to table 10. In summary, the IV estimates generally support the idea that railways had a large effect in creating structural change in agriculture, in transport where railways were a big employer, and in secondary sectors that were most likely affected by agglomeration economies.

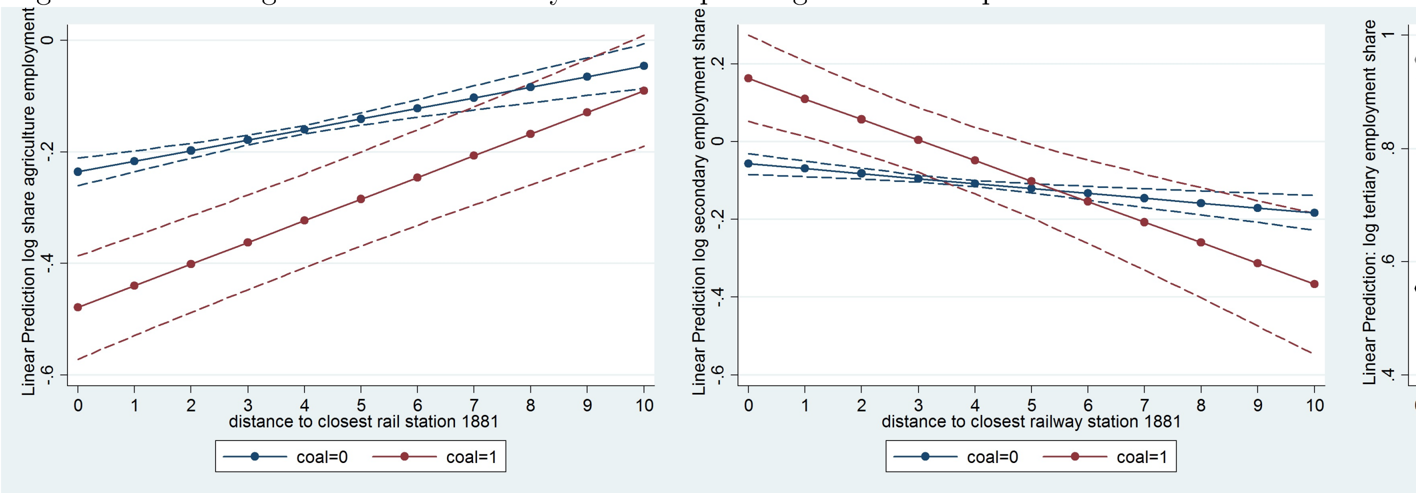
5 Heterogeneity

This last section examines heterogeneity in the effects of railways. We focus on three effects: (1) the interaction between railways and coal endowments in a parish, (2) the interaction between railways and parish population density in 1801, and the (3) the interaction between railways and waterways. One argument is that railways allowed parishes with coal to exploit their cheap energy advantage, perhaps by adopting more capital and energy intensive methods. The positive interaction between railways and coal effect should apply to the secondary sector where energy is a cost, and the interaction effect should be less or perhaps the opposite for tertiary and agriculture.

We test whether railway access differentially affected employment structure in parishes with coal by including in our baseline model an interaction between distance to railway stations and the indicator for a parish having coal. The results are shown in table 11. Coal endowments accentuate the negative effect on total employment of being distant from railway stations. The same is true for the negative effect on the secondary share and the positive effect on the agricultural share from being distant to railways. Coal offsets the negative effect of distance to railways for the tertiary share. The marginal effects for the shares are shown below in figure 9. If a parish has coal and their access to railways is bad (say 5 or 6 km) then their agricultural employment share is close to parishes without coal, but as their access to railways gets better (say 1 or 2 km) their agricultural share falls more rapidly than for a parish without coal. The opposite holds for secondary employment. As parishes with coal get better access to railways their secondary employment rises more rapidly. The combination of coal and better railway access did not have the same effects on the share of tertiary. In this case, better railway access lowers tertiary employment more if a parish has coal. But if the parish has no coal then better railway access raises tertiary employment. These findings suggest that having coal and being close to railway stations made a parish specialize more in the secondary sector where its cheap energy gave manufacturing activities a cost advantage. Parishes without coal tended to get much more tertiary and milder reductions in agriculture as railway station access improved. The lack of an energy cost advantage meant that services were more competitive and secondary was perhaps less competitive.

Another source of heterogeneity concerns the interaction between railways and initial population density. One argument is that railways created a divergence in occupational structure. The most dense areas in 1801 generally had the highest secondary and tertiary employment in the early nineteenth century because of agglomeration economies. Railways

Figure 9: The marginal effects of railway access depending on whether parishes have coal

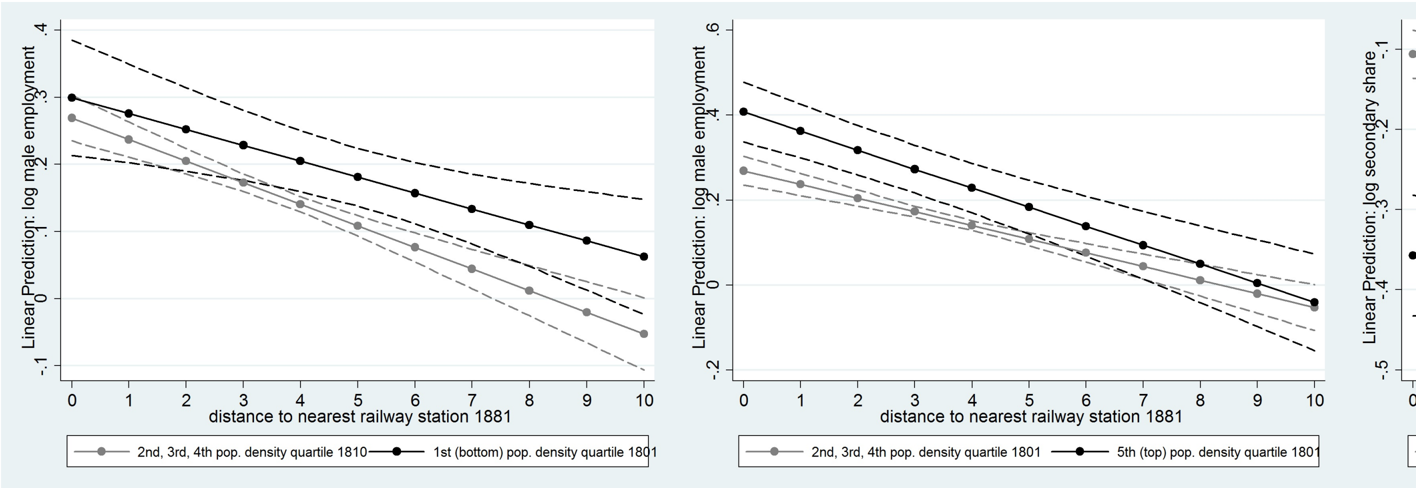


Sources: see text.

potentially strengthened agglomeration by leading to even more secondary and tertiary concentration in the most dense areas. We test this hypothesis by creating five indicator variables for the first, second, third, fourth, and fifth quartiles of log population density (note that the first quartile is the least dense and the fifth quartile is the most dense). The specification drops the log of population density and includes the first four quartile indicator variables along with interactions with distance to railway stations. The omitted group is the fifth quartile which has the greatest density.

The results for the main variables are reported in table 12 and illustrated in figure 10. Greater distance to railway stations decreases total employment as we saw earlier. The effects are a bit larger for parishes in the top quartile of population density, and especially when their distance to railway stations is very small. Thus railway access contributed to the divergence in employment and population already present early in the nineteenth century. We also find that greater distance to stations decreases the share of secondary employment relatively little except for the top quartile of population density in 1801. Thus railway access contributed to further specialization in secondary employment, but only in the most dense locations where agglomeration economies were already present. The share

Figure 10: The marginal effects of railway access depending on quartile of population density 1801



Sources: see text.

of tertiary employment was unaffected by distance to stations at the bottom quartile of population density, and it was more affected at the top quartile. Once again we see that railways contributed to specialization mainly in the most dense locations, already with agglomeration.

A related issue concerns the interaction between railways and waterways. Britain had an extensive network of waterways when the railway era began. Industries were already concentrated near waterways because they offered cheap transport. One argument is that railways had larger effects in parishes with waterways because of prior agglomeration. Another argument is that waterways provided competition to railways, and perhaps enhanced their effects through lower freight rates. The competitive effect is questionable however because many waterways were taken over by railways by 1881. The interaction effects are estimated using similar regressions as before. They are not reported to save space. Essentially they show that having waterways and railways contributed to employment growth, and more so if a parish had both. However, there is little effect on employment shares, except for tertiary where there is negative effect of waterways and a negative interaction effect. It does not appear that railway access led to more structural change if waterways were present.

6 Conclusion

The transition from agriculture to manufacturing, mining, and services is one of the hallmarks of the British industrial revolution. This paper uses new data to examine the effects of railway access on total employment and occupational structure. The data include numbers of males employed in 12 occupational groups across more than 9500 parishes in England and Wales at two dates 1817 and 1881. The data also draw on new GIS data on transport infrastructure, population, and endowments. The main findings show that proximity to a railway station increased a parish's total employment, increased its share of some aspects of secondary and tertiary employment, and significantly decreased its share of agricultural employment. The effects on secondary employment were also larger in parishes with coal or with dense populations in 1801.

References

1. Alvarez, Eduard, Xavi Franch, and Jordi Martí-Henneberg. "Evolution of the territorial coverage of the railway network and its influence on population growth: The case of England and Wales, 1871–1931." Historical Methods: A Journal of Quantitative and Interdisciplinary History 46.3 (2013): 175-191.

2. Becker, Sascha O., Erik Hornung, and Ludger Woessmann. "Education and catch-up in the industrial revolution." American Economic Journal: Macroeconomics (2011): 92-126.

3. Bogart, Dan. "A global perspective on railway inefficiency and the rise of state ownership, 1880–1912." Explorations in Economic History 47.2 (2010): 158-178.

4. Bogart, Dan. "The Transport Revolution in Industrializing Britain," in Floud, Roderick, Jane Humphries, and Paul Johnson, eds. The Cambridge Economic History of Modern Britain: Volume 1, Industrialisation, 1700–1870. Cambridge University Press, 2014.

5. Casson, Mark. The world's first railway system: enterprise, competition, and regulation on the railway network in Victorian Britain. Oxford University Press, 2009.

6. Casson, Mark. "The determinants of local population growth: A study of Oxfordshire in the nineteenth century." Explorations in Economic History 50.1 (2013): 28-45.

7. Crafts, Nicholas, and Abay Mulatu. "How did the location of industry respond to falling transport costs in Britain before World War I?." The Journal of Economic History 66.03 (2006): 575-607.

8. Crafts, Nicholas, and Nikolaus Wolf. "The location of the UK cotton textiles industry in 1838: A quantitative analysis." The Journal of Economic History 74.04 (2014): 1103-1139.

9. Del Río, Eloy, Jordi Martí-Henneberg, and Antònia Valentín. "La Evolución de la red ferroviaria en el Reino Unido (1825-2000)." Treballs de La Societat Catalana de Geografia 65 (2008): 654-663.

10. Desmet, Klaus, and Esteban Rossi-Hansberg. "Spatial development." The American Economic Review 104.4 (2014): 1211-1243.

11. Donaldson, Dave. Railroads of the Raj: Estimating the impact of transportation infrastructure. No. w16487. National Bureau of Economic Research, 2010.

12. Donaldson, Dave, and Richard Hornbeck. Railroads and American Economic Growth: A" Market Access" Approach. No. w19213. National Bureau of Economic Research, 2013.

13. Faber, Benjamin. "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System." The Review of Economic Studies (2014): rdu010.

14. Fernihough, Alan, and Kevin Hjortshøj O'Rourke. Coal and the European industrial revolution. No. w19802. National Bureau of Economic Research, 2014.

15. Fishlow, Albert. American Railroads and the Transformation of the Ante-bellum Economy. Vol. 127. Cambridge, MA: Harvard University Press, 1965.

16. Fogel, R. "Railways and American Economic Growth." Baltimore: Johns Hopkins Press.(1979):"Notes on the Social Saving Controversy," The Journal of Economic History 39.1 (1964): 1-54.

17. Gutberlet, Theresa. "Cheap Coal versus Market Access: The Role of Natural Resources and Demand in Germany's Industrialization." (2014).

18. Hawke, Gary Richard. Railways and economic growth in England and Wales, 1840-1870. Clarendon Press, 1970.

19. Heblich, Stephan, and Alex Trew. "Banking and Industrialization." (2015).

20. Kelly, Morgan; Mokyr, Joel; Ó Gráda, Cormac. “The Roots of the Industrial Revolution.” UCD Centre for Economic Research Working Paper Series 2015, WP15/24.

21. Kitson, Peter M., et al. "The creation of a ‘census’ of adult male employment for England and Wales for 1817." Work in Progress. For details of the wider project see http://www. geog. cam. ac. uk/research/projects/occupations (2012).

22. Odlyzko, Andrew. "Collective hallucinations and inefficient markets: The British Railway Mania of the 1840s." University of Minnesota (2010).

23. Redding, Stephen J., and Matthew A. Turner. Transportation costs and the spatial organization of economic activity. No. w20235. National Bureau of Economic Research, 2014.

24. Shaw-Taylor, L. and Wrigley, E. A. “Occupational Structure and Population Change,” in Floud, Roderick, Jane Humphries, and Paul Johnson, eds. The Cambridge Economic History of Modern Britain: Volume 1, Industrialisation, 1700–1870. Cambridge University Press, 2014.

25. Wellington, A.M. The Economic Theory of the Location of Railways: An Analysis of the Conditions Controlling the Laying Out of Railways to Effect the Most Judicious Expenditure of Capital. Ed. J. Wiley & sons, 1877.

26. Wrigley, Edward Anthony. Energy and the English industrial revolution. Cambridge University Press, 2010.

Table 1: Occupational Sub-groups: Part I

Categories	Agriculture	Labourer	Mining	Clothing	Footware	Textiles
1	Farming and market gardening		Miner	Clothing manufacture	Boots and shoes	Textile occupations
2	Animal husbandry		Coal mining	Hats, gloves, stockings	footwear, other	Woollen and worsted manufacture
3	Gardening and estate work		Lead mining	Clothing other	Direction, supervision	Cotton manufacture
4			Tin mining	Direction, supervision		Linen, jute manufacture
5			Copper mining			Silk manufacture Lace manufacture Textiles, other
6			Iron mining			Direction, supervision
7			Salt mining			
8			Quarrying			
9			Mining and quarrying, other			
10			Direction, supervision of mining, quarrying			

Notes: for more details see the PST system documentation,
http://www.campop.geog.cam.ac.uk/research/projects/occupations/categorisation/pst.pdf

Table 1: Occupational Sub-groups: part 2

Categories	Iron and steel	Machine making, tool	Building and construction	Dealers and sellers in	Professions	Transport
1	Smith	Machine making	Building	products of land and water	Legal profession	Road transport
2	Iron manufacture	Tool, implement making	Bricklaying	coal and minerals	Medical profession	Canal, river transport
3	Steel manufacture	Engineering	Masonry	food, drink, tobacco	Ministers, priests, etc.	Sea transport
4	Iron and steel products, major categories	Direction, supervision in machine making, etc.	Carpentry	clothing and footwear	Teaching	Rail transport
5	Iron and steel processing		Slater, tiler, thatcher	textile products	Scientific professions	Docks, warehouses, storage facilities
6	Iron and steel products, other		Plasterer, painter, decorator	wood products	Artistic professions	Porter
7	Direction, supervision		Plumber, glazier	leather, hair, bone, straw products	Professions relating to construction	Messenger
8			Construction, public works, etc.	furnishings		Others in transport
9			Other	paper and paper products		
10			Direction, supervision	printing and publishing products		
11				earthenware, pottery		
12				glass and glass products		
13				precious metals and jewelry		
14				instruments		
15				chemicals and chemical products		
16				iron and steel, and iron and steel products		
17				non-ferrous metal products		
18				machines, tools		
19				marine stores		
20				bricks, tiles		
21				stone and stone products		
22				building materials		
23				coke, tar, water, etc.		
24				minor manufacturing products		
25				other		

Notes: for more details see the PST system documentation,
http://www.campop.geog.cam.ac.uk/research/projects/occupations/categorisation/pst.pdf

Table 2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
log diff employment	9586	0.181	0.539	-4.064	5.734
log diff share agriculture	9521	-0.178	0.342	-2.605	2.186
log diff share secondary	8871	-0.094	0.619	-3.296	3.327
log diff share tertiary	8452	0.840	0.797	-3.157	4.337
Rail station indicator	9588	0.249	0.432	0.000	1.000
Nearest station 1881 in km	9586	3.841	2.956	0.023	26.005
coastal indicator	9586	0.148	0.355	0.000	1.000
coal indicator	9586	0.080	0.271	0.000	1.000
average slope to neighboring parishes	9581	22.057	17.782	1.371	179.99
stand. Dev. slope to neighboring parishes	9581	4.777	3.631	0.487	37.436
average elevation parish	9581	89.815	74.143	-1.255	525.30
log pop. Density 1801	9582	-3.030	1.310	-6.424	4.530
log diff pop. 1801 to 1821	9582	0.196	0.191	-1.059	1.910
share secondary emp. In 1817	9586	0.215	0.152	0.000	1.000
share tertiary emp. In 1817	9586	0.101	0.118	0.000	1.000
nearest waterway in 1830 in km	9586	7.237	6.496	0.006	48.387
nearest port in km	9586	34.571	23.786	0.079	104.72
nearest turnpike road in 1830 in km	9586	1.381	1.759	0.000	20.296

Table 3: Difference in mean difference analysis, Part 1

Panel A	Mean log difference from c.1817 to 1881 in				
	(1)	(2)	(3)	(4)	(5)
Station variables 1881	Total population	Employment	Male Agric. &lab. employment	Male Secondary employment	Male Tertiary employment
Station indicator=0	0.049	0.071	-0.059	-0.029	0.893
Station indicator=1	0.494	0.513	0.224	0.447	1.430
Diff. in mean diff.	0.445	0.442	0.284	0.476	0.536
P-value, two-tail test	0.00	0.00	0.00	0.00	0.00
N	9586	9586	9519	8868	8442
Distance to station above median	0.023	0.048	-0.082	-0.063	0.913
Distance to station below median	0.297	0.313	0.107	0.251	1.157
Diff. in mean diff.	0.274	0.264	0.190	0.315	0.243
P-value, two-tail test	0.00	0.00	0.00	0.00	0.00
N	9586	9586	9519	8868	8442

Panel B	Mean log difference from c.1817 to 1881 in		
	(6)	(7)	(8)
Station variables 1881	Agric. & lab. Share	Secondary share	Tertiary share
Station indicator=0	-0.141	-0.103	0.814
Station indicator=1	-0.289	-0.067	0.909
Diff. in mean diff.	-0.147	0.035	0.095
P-value, two-tail test	0.00	0.016	0.00
N	9521	8871	8452
Distance to station above median	-0.130	-0.113	0.855
Distance to station below median	-0.226	-0.075	0.825
Diff. in mean diff.	-0.095	0.038	-0.029
P-value, two-tail test	0.00	0.003	0.084
n	9521	8871	0

Table 3: Difference in mean difference analysis, part 2

Panel C Station variables 1881	(1) Agriculture share	(2) Labourer Share	(3) Mining Share	(4) Builders Share
Mean share c.1817 if >0	0.199	0.481	0.103	0.073
N	9311	9375	1272	7878
	Mean difference in share from c.1817 to 1881 if either share >0			
Station indicator=0	0.342	-0.433	0.032	-0.0017
Station indicator=1	0.165	-0.301	0.050	0.0028
Diff. in mean diff.	-0.176	0.147	0.018	0.0045
P-value, two-tail test	0.00	0.00	0.016	0
N	9524	9508	3035	9001
	Mean difference in share from c.1817 to 1881 if either share >0			
Distance to station above median	0.359	-0.449	0.029	-0.0024
Distance to station below median	0.235	-0.351	0.047	0.0012
Diff. in mean diff.	-0.124	0.098	0.018	0.0036
P-value, two-tail test	0.00	0.00	0.003	0.00
N	9524	9508	3035	9001

Station variables 1881	(5) Machine tools share	(6) Iron-steel tools Share	(7) Textile Share	(8) Footware Share
Mean share c.1817 if >0	0.016	0.033	0.066	0.040
N	1570	6405	3074	6702
	Mean difference in share from c.1817 to 1881 if either share >0			
Station indicator=0	0.010	-0.0011	-0.027	-0.012
Station indicator=1	0.013	0.0024	-0.038	-0.014
Diff. in mean diff.	0.003	0.0035	-0.011	-0.002
P-value, two-tail test	0.00	0.00	0.016	0.009
N	4750	8298	3822	7964
	Mean difference in share from c.1817 to 1881 if either share >0			
Distance to station above median	0.009	-0.0022	-0.024	-0.011
Distance to station below median	0.012	0.0018	-0.035	-0.014
Diff. in mean diff.	0.0037	0.0041	-0.011	-0.002
P-value, two-tail test	0.00	0.00	0.00	0.005
N	4750	8298	3822	7964

Table 3: Difference in mean difference analysis, part 3

Panel C continued Station variables 1881	(9) Clothing share	(10) Transport Share	(11) Professions & Services Share	(12) Dealers & Sellers Share
Mean share c.1817 if >0	0.038	0.048	0.074	0.032
N	5022	4435	8025	4720
	Mean difference in share from c.1817 to 1881 if either share >0			
Station indicator=0	-0.014	0.028	0.060	0.107
Station indicator=1	-0.015	0.046	0.057	0.106
Diff. in mean diff.	-0.001	0.018	-0.003	0.000
P-value, two-tail test	0.35	0.00	0.153	0.657
N	6360	8689	9515	9476
	Mean difference in share from c.1817 to 1881 if either share >0			
Distance to station above median	-0.011	0.022	0.061	0.101
Distance to station below median	-0.017	0.043	0.057	0.111
Diff. in mean diff.	-0.006	0.021	-0.004	0.009
P-value, two-tail test	0.000	0.000	0.025	0.000
N	6360	8689	9515	7964

Table 4: Baseline regression results for railway access

	Log Difference in						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	All male employ.	Agric. & lab. Employ.	Secondary employ.	Tertiary employ.	Agric. &lab. share	Secondary share	Tertiary share
Panel A							
Rail station=1	0.293***	0.119***	0.441***	0.433***	-0.164***	0.158***	0.152***
	(0.031)	(0.020)	(0.041)	(0.029)	(0.013)	(0.017)	(0.022)
Observations	9,577	9,517	8,863	8,436	9,517	8,864	8,444
R-squared	0.301	0.328	0.262	0.401	0.259	0.300	0.461
Panel B							
Nearest station in km	-0.035***	-0.014***	-0.048***	-0.056***	0.020***	-0.014***	-0.022***
	(0.005)	(0.002)	(0.007)	(0.007)	(0.003)	(0.003)	(0.004)
	Standardized coefficient						
	-0.190	-0.091	-0.160	-0.174	0.169	-0.067	-0.083
Observations	9,577	9,517	8,863	8,436	9,517	8,864	8,444
R-squared	0.283	0.325	0.242	0.390	0.245	0.293	0.460

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: A comparison of the baseline coefficients for railway access

VARIABLES	(1) log diff. male employment	(2) log diff. agric. lab. share	(3) log diff. sec. share	(4) log diff. tert. share
Rail station=1	0.293***	-0.164***	0.158***	0.152***
	(0.031)	(0.013)	(0.017)	(0.022)
Coastal=1	0.191***	-0.029	0.064***	-0.007
	(0.027)	(0.018)	(0.024)	(0.035)
Coal=1	0.325***	-0.173***	0.093**	-0.294***
	(0.068)	(0.045)	(0.046)	(0.033)
Observations	9,577	9,517	8,864	8,444
R-squared	0.301	0.259	0.300	0.461
Nearest station 1881 in km	-0.035***	0.020***	-0.014***	-0.022***
	(0.005)	(0.003)	(0.003)	(0.004)
Nearest waterway in km	-0.004**	0.002	-0.004**	0.003
	(0.001)	(0.001)	(0.002)	(0.003)
Nearest turnpike in km	-0.003	0.004	-0.025***	0.008
	(0.008)	(0.004)	(0.005)	(0.006)
Observations	9,577	9,517	8,864	8,444
R-squared	0.283	0.245	0.293	0.460

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Railway proximity and percentage change occupational shares by sub-group

VARIABLES	(1) ΔAgricultural share	(2) ΔLabourer share	(3) ΔMining share	(4) ΔBuilder share	(5) ΔMachine tools share	(6) ΔIron-Steel share
Nearest station 1881 in km	1.234***	-0.295***	-0.379**	-0.026	-0.055***	-0.039**
	(0.129)	(0.081)	(0.146)	(0.028)	(0.009)	(0.018)
			Standardized coefficient			
	0.147	-0.036	-0.088	-0.013	-0.069	-0.030
All controls	Y	Y	Y	Y	Y	Y
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.414	0.577	0.227	0.135	0.099	0.063

VARIABLES	(7) ΔTextile share	(8) ΔFootware share	(9) ΔClothing share	(10) ΔTransport share	(11) ΔProfession Services share	(12) ΔDealers Sellers share
Nearest station 1881 in km	-0.099*	0.004	-0.028	-0.487***	0.047	-0.052
	(0.057)	(0.016)	(0.029)	(0.061)	(0.051)	(0.047)
			Standardized coefficient			
	-0.035	0.002	-0.017	-0.222	0.015	-0.017
All controls	Y	Y	Y	Y	Y	Y
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.431	0.074	0.339	0.189	0.304	0.280

Notes: Occupational shares are in percentages. The sample only includes parishes that had a positive occupational share in 1817 or 1881. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Railway proximity and percentage change occupational shares by sub-group

VARIABLES	(1) ΔAgricultural share	(2) ΔLabourer share	(3) ΔMining share	(4) ΔBuilder share	(5) ΔMachine tools share	(6) ΔIron-Steel share
A. Nearest station 1851 in km	0.164***	-0.057	0.033	-0.009	-0.007	0.001
	(0.057)	(0.049)	(0.069)	(0.014)	(0.006)	(0.010)
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.401	0.577	0.222	0.135	0.096	0.062
B. Nearest station 1861 in km	0.472***	-0.112	-0.122	-0.008	-0.029***	-0.028**
	(0.108)	(0.094)	(0.095)	(0.014)	(0.007)	(0.014)
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.414	0.577	0.227	0.135	0.099	0.063
C. Nearest station 1871 in km	0.632***	-0.008	-0.264*	-0.030	-0.046***	-0.026*
	(0.219)	(0.163)	(0.147)	(0.018)	(0.008)	(0.014)
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.406	0.576	0.225	0.135	0.099	0.063
Nearest station 1881 in km	1.234***	-0.295***	-0.379**	-0.026	-0.055***	-0.039**
	(0.129)	(0.081)	(0.146)	(0.028)	(0.009)	(0.018)
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.414	0.577	0.227	0.135	0.099	0.063

VARIABLES	(7) ΔTextile share	(8) ΔFootware share	(9) ΔClothing share	(10) ΔTransport share	(11) ΔProfession Services share	(12) ΔDealers Sellers share
A.. Nearest station 1851 in km	-0.036*	0.009	0.009	-0.087***	0.023	-0.015
	(0.021)	(0.006)	(0.009)	(0.021)	(0.026)	(0.022)
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.431	0.074	0.339	0.160	0.304	0.280
B. Nearest station 1861 in km	-0.059	-0.000	-0.009	-0.196***	0.045	-0.006
	(0.039)	(0.011)	(0.013)	(0.030)	(0.037)	(0.031)
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.431	0.074	0.339	0.189	0.304	0.280
C. Nearest station 1871 in km	-0.090	-0.012	-0.006	-0.294***	0.017	-0.055
	(0.055)	(0.013)	(0.019)	(0.061)	(0.044)	(0.033)
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.432	0.074	0.338	0.175	0.304	0.280
D. Nearest station 1881 in km	-0.099*	0.004	-0.028	-0.487***	0.047	-0.052
	(0.057)	(0.016)	(0.029)	(0.061)	(0.051)	(0.047)
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.431	0.074	0.339	0.189	0.304	0.280

Notes: Occupational shares are in percentages. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include the full set of controls.

Table 6: Robustness I: Results after adding control variables successively

	Log difference between 1817 and 1881 in			
	(1)	(2)	(3)	(4)
VARIABLES	All male employment	Agriculture & labourer share	Secondary share	Tertiary share
Panel A: few controls				
Nearest station 1881 in km	-0.037***	0.014***	-0.005*	-0.004
	(0.005)	(0.003)	(0.003)	(0.006)
Controls for coal, coastal, ruggedness, elevation, and soil types	Y	Y	Y	Y
Controls for pre-railway employment shares, pop. density, and growth	N	N	N	N
Controls for pre-railway transport infrastructure and county FE's	N	N	N	N
Observations	9,581	9,521	8,866	8,447
R-squared	0.168	0.081	0.028	0.069
Panel B: more controls				
Nearest station 1881 in km	-0.038***	0.022***	-0.017***	-0.023***
	(0.005)	(0.003)	(0.004)	(0.004)
Controls for coal, coastal, ruggedness, elevation, and soil types	Y	Y	Y	Y
Controls for pre-railway employment shares, pop. density, and growth	Y	Y	Y	Y
Controls for pre-railway transport infrastructure and county FE's	N	N	N	N
Observations	9,577	9,517	8,864	8,444
R-squared	0.231	0.205	0.265	0.421
Panel C: all controls				
Nearest station 1881 in km	-0.035***	0.020***	-0.014***	-0.022***
	(0.005)	(0.003)	(0.003)	(0.004)
Controls for coal, coastal, ruggedness, elevation, and soil types	Y	Y	Y	Y
Controls for pre-railway employment shares, pop. density, and growth	Y	Y	Y	Y
Controls for pre-railway transport infrastructure and county FE's	Y	Y	Y	Y
Observations	9,577	9,517	8,864	8,444
R-squared	0.283	0.245	0.293	0.460

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Instrumental variables estimates using Least Cost Paths

VARIABLES	(1) Log diff. total employment	(2) Log diff. agric. & lab. share	(3) Log diff. secondary share	(4) Log diff. tertiary share
A. OLS estimate	-0.034***	0.020***	-0.014***	-0.022***
B. Instrument: LCP between London and municipal boroughs				
Nearest station 1881 in km	-0.029	0.030***	-0.022	-0.080***
	(0.020)	(0.014)	(0.023)	(0.024)
Cragg-Donald Wald F statistic	84.91	84.11	70.63	71.91
Observations	9577	9,517	8,864	8,444
C. Instrument: LCP between London and regional cities, regional cities to boroughs				
Nearest station 1881 in km	-0.046***	0.062***	-0.056***	-0.132***
	(0.017)	(0.013)	(0.021)	(0.028)
Cragg-Donald Wald F statistic	117.65	116.62	109.12	104.56
Observations	9577	9,517	8,864	8,444
D. Instrument: LCP between all cities based on gravity model				
Nearest station 1881 in km	-0.054***	0.057***	-0.014	-0.015
	(0.017)	(0.015)	(0.020)	(0.026)
Cragg-Donald Wald F statistic	97.55	96.38	87.68	82.41
Observations	9577	9,517	8,864	8,444
E. Instrument: LCP between London and municipal boroughs, London and regional cities, reg. cities to boroughs				
Nearest station 1881 in km	-0.041***	0.053***	-0.049**	-0.120***
	(0.017)	(0.012)	(0.019)	(0.027)
Cragg-Donald Wald F statistic	63.29	62.73	56.82	55.18
Overid. test of all instruments	1.24	9.51	3.037	5.407
P-value overid. test	0.264	0.002	0.081	0.020
Observations	9577	9,517	8,864	8,444
F. Instrument: LCP between London and regional cities, reg. cities to boroughs, between all cities gravity model				
Nearest station 1881 in km	-0.049***	0.060***	-0.039**	-0.085***
	(0.015)	(0.012)	(0.017)	(0.023)
Cragg-Donald Wald F statistic	76.66	75.89	70.57	67.01
Overid. test of all instruments	0.163	0.159	3.717	17.38
P-value overid. test	0.686	0.69	0.053	0.000
Observations	9577	9,517	8,864	8,444
G. Instrument LCP between London and municipal boroughs, LCP between all cities based on gravity model				
Nearest station 1881 in km	-0.043**	0.045***	-0.017	-0.044**
	(0.016)	(0.012)	(0.018)	(0.024)
Cragg-Donald Wald F statistic	64.02	63.32	55.66	53.99
Overid. test of all instruments	1.401	3.23	0.127	5.55
P-value overid. test	0.236	0.072	0.721	0.018
Observations	9577	9,517	8,864	8,444

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Railway proximity and percentage change occupational shares by sub-group: IV estimates

VARIABLES	(1) Δagric. share	(2) Δlabourer share	(3) Δmining share	(4) Δbuilder share	(5) Δmachine tool share	(6) Δiron steel share
			OLS estimates			
Nearest station 1881 in km	1.234***	-0.295***	-0.379**	-0.026	-0.055***	-0.039**
	(0.129)	(0.081)	(0.146)	(0.028)	(0.009)	(0.018)
			IV estimates			
Nearest station 1881 in km	2.576***	0.226	-1.918*	-0.085	-0.117	0.129
	(0.681)	(0.569)	(1.023)	(0.179)	(0.107)	(0.113)
LCP between London and regional cities, regional cities to boroughs	Y	Y	Y	Y	Y	Y
LCP between London and municipal boroughs	Y	Y	Y	N	Y	Y
LCP between all cities gravity model	Y	Y	Y	Y	Y	Y
Cragg-Donald Wald F statistic	51.64	51.57	9.846	70.78	23.81	46.11
Overid. test of all instruments	2.466	1.39	2.174	1.151	1.987	2.70
P-value overid. test	0.291	0.497	0.337	0.283	0.370	0.259
Observations	9,522	9,505	3,035	8,997	4,747	8,294
R-squared	0.396	0.574	0.143	0.134	0.094	0.051

VARIABLES	(7) Δtextile share	(8) Δfootware share	(9) Δclothing share	(10) Δtransport share	(11) Δprofession service share	(12) Δdealer seller share
			OLS estimates			
Nearest station 1881 in km	-0.099*	0.004	-0.028	-0.487***	0.047	-0.052
	(0.057)	(0.016)	(0.029)	(0.061)	(0.051)	(0.047)
			IV estimates			
Nearest station 1881 in km	-0.579**	-0.160	-0.447**	-1.129***	-0.182	-0.383
	(0.283)	(0.114)	(0.216)	(0.241)	(0.256)	(0.240)
LCP between London and regional cities, regional cities to boroughs	N	Y	Y	Y	N	N
LCP between London and municipal boroughs	Y	Y	Y	Y	Y	Y
LCP between all cities gravity model	Y	Y	N	N	Y	Y
Cragg-Donald Wald F statistic	20.98	43.73	28.11	65.79	63.53	62.06
Overid. test of all instruments	0.012	1.569	0.111	1.097	0.164	0.233
P-value overid. test	0.912	0.456	0.738	0.295	0.685	0.629
Observations	3,820	7,960	6,352	8,680	9,506	9,467
R-squared	0.412	0.062	0.293	0.127	0.300	0.271

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to units that had positive occupational shares in 1817 or 1881.

Table 9: Inter-action effect between railway access and coal endowments

VARIABLES	(1) log diff. male employment	(2) log diff. agriculture & labourer share	(3) log diff. secondary share	(4) log diff. tertiary share
Nearest station 1881 in km	-0.031***	0.019***	-0.013***	-0.024***
	(0.004)	(0.003)	(0.003)	(0.005)
(Nearest station 1881 in km)x(Coal)	-0.094***	0.020***	-0.040***	0.048***
	(0.017)	(0.005)	(0.012)	(0.014)
Coal	0.604***	-0.243***	0.220***	-0.404***
	(0.094)	(0.049)	(0.057)	(0.055)
Observations	9,577	9,517	8,864	8,444
R-squared	0.293	0.246	0.294	0.461

Notes: Robust standard errors in parentheses, clustered on counties. *** p<0.01, ** p<0.05, * p<0.1

Table 12: Inter-action effect between railway access and population density in 1801

VARIABLES	(1) log diff. male employment	(2) log diff. agricultural & labourer share	(3) log diff. secondary share	(4) log diff. tertiary share
Nearest station 1881 in km	-0.045***	0.023***	-0.028***	-0.026***
	(0.007)	(0.005)	(0.007)	(0.008)
(Nearest station 1881 in km)x(1quart.)	0.022**	-0.008	0.023***	0.023*
	(0.010)	(0.006)	(0.008)	(0.012)
(Nearest station 1881 in km)x(2quart.)	0.017**	-0.006	0.018***	0.001
	(0.007)	(0.005)	(0.007)	(0.009)
(Nearest station 1881 in km)x(3quart.)	0.015**	-0.006	0.022***	-0.011
	(0.007)	(0.006)	(0.007)	(0.011)
(Nearest station 1881 in km)x(4quart.)	0.006	0.002	0.013*	-0.002
	(0.007)	(0.006)	(0.007)	(0.009)
1st quartile log pop. Den 1801	-0.115*	0.186***	-0.635***	0.208***
	(0.067)	(0.058)	(0.060)	(0.064)
2nd quartile log pop. Den 1801	-0.158***	0.166***	-0.500***	0.216***
	(0.047)	(0.044)	(0.057)	(0.057)
3rd quartile log pop. Den 1801	-0.176***	0.151***	-0.422***	0.145***
	(0.044)	(0.041)	(0.048)	(0.050)
4th quartile log pop. Den 1801	-0.098***	0.081**	-0.326***	0.038
	(0.033)	(0.039)	(0.034)	(0.041)
Observations	9,577	9,517	8,864	8,444
R-squared	0.270	0.256	0.324	0.470

Notes: Robust standard errors in parentheses, clustered on counties. *** p<0.01, ** p<0.05, * p<0.1