Railways and employment: evidence from 19th century England and Wales

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

Preliminary Draft, Dec. 2016[‖]

Abstract

How do transport improvements affect the long-term growth and spatial structure of employment? This paper answers these questions in the context of 19th century England and Wales where urbanization increased and the shares of secondary and tertiary employment rose. To address endogeneity we use an instrument that identifies locations close to railways because they were on a least cost path minimizing elevation changes and distance between large towns. The empirical analysis shows that a one standard deviation increase in distance to 1851 railway stations reduced secondary and tertiary employment growth in a location by 6.0% and 18.4% over the next 30 years. Quantitatively similar results are found for the effect of distance to stations in 1881 on population and employment levels in 1881, and new occupations in the mid-19th century. We also provide evidence that around one-third of the effects of station distance is due to the reorganization of employment from initially low to initially medium employment areas.

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, UC San Diego, NYU, Florida State, and the Economic History Association Meetings. All errors are our own.

1 Introduction

Transport improvements can dramatically alter the population density and employment structure of local economies. When a location gets connected to a national or international transportation network its market access increases relative to other locations nearby. Greater market access encourages workers and firms to move there because of the greater supply of consumer goods and greater demand for products. The result is that population density is generally higher near access points to the transport network like seaports, airports, railway stations, and highway junctions. The effects of gaining greater access to transport may not be the same across all locations however. Some may experience a loss in population and employment to larger or more productive areas. The magnitude of 'reorganization effects' and 'growth effects' is a major issue both in the academic literature and in policy discussions.[1]

This paper examines how railways affected population and employment growth in nineteenth century England and Wales. This historical context is well suited to studying the long-term effects of transport. Steam locomotives were invented in Britain in the 1820s and by the mid nineteenth century railway lines and stations were built throughout the country. Railways marked a major improvement over inland road and water transport. As an illustration rail freight rates were one-tenth the level of road freight rates in the mid-nineteenth century (Bogart 2014). The English context also provides insights because there is rich data on population and male occupational structure at the local level from 1801 onward (see Kitson et. al. 2012, Shaw Taylor and Wrigley 2014). These data show that in the nineteenth century, population density increased, employment in secondary and tertiary occupations increased, and agricultural employment fell. Growth was also uneven across space as population become more concentrated in certain locations, especially in the northern industrial areas.

[1]See Hettigate (2006) for the policy debate and see Redding and Turner (2014) for the academic debate.

This papers makes use of this corpus of historical data. We study population and male employment density in more than 9,000 parishes and townships in England and Wales. The population data are available at every census year from 1801 to 1881. Male employment data, dis aggregated by occupational groups, is available at three dates 1817, 1851, and 1881. There are five general groups: (1) secondary, (2) tertiary, (3) agriculture, (4) extraction or mining, and (5) labourer. Additional specialties within these categories are also studied. The population and occupational data are linked to a shape-file of consistent jurisdictional units between 1801 and 1881, so that their change over time can be studied. For railways we observe the geographic location and the date of opening for all stations. These data are linked to the same jurisdictional units for population and employment.

We employ ordinary least squares (OLS) regression models for studying transport improvements (see Redding and Turner 2014 for an overview). These include regressions of population or employment growth from 1851 to 1881 on distance to railway stations in 1851 and regressions of population or employment levels in 1881 on distance to railway stations in 1881. Besides a variable for railways, our models include a rich set of controls for geographic characteristics, pre-railway infrastructure, and pre-railway trends in population density and occupational structure. Still there may be a concern about omitted variables associated with non-random placement of railway stations. To address this issue we build on related studies in the literature and construct a hypothetical railway network connecting large towns in 1801.[2] The routes are chosen to minimize elevation changes and distance. As we show below distance to our hypothetical network provides a powerful instrumental variable (IV) for distance to the actual railway network.

This paper also explores growth versus reorganization effects in novel ways. One approach focuses on new occupations, like electrical goods, and relatively new occupations where growth was large between 1851 and 1881. In these cases, the effects of railways are close to capturing pure growth effects because there was little initial employment prior to

[2]See Faber (2014), Lipscomb et. al. (2013) for two examples.

railways to reorganize. In a second approach, we use the model of industry location developed by Desmet and Rossi-Hansberg (2009, 2014). They show that initial employment density and employment growth have a negative sloping S-shaped relationship due to technology diffusion, knowledge spillovers, and congestion. We quantify how this structural relationship changes for areas close to railway stations versus areas more distant from railways. These reorganization effects are compared with the direct effect of being close to railway stations.

The main findings show that population, secondary, and tertiary employment levels in 1881 decrease with distance to stations in 1881. We also find that population, secondary, and tertiary employment growth from 1851 to 1881 decreases with distance to stations in 1851. The IV estimates tend to be similar and in some cases larger than the OLS estimates. One of our main specifications shows that a one standard deviation increase in distance to 1851 railway stations reduces secondary and tertiary employment growth in a location by 6.0 and 18.4 percentage points over the next 30 years. In a counter-factual exercise we also show that national secondary and tertiary employment growth is predicted to be 8.5 and 25.1 percentage points lower if the railway network had remained at its 1841 level. In further results, we find that extractive and labourer employment also decreased with station distance complementing the findings for secondary and tertiary. By contrast, distance to stations had a zero or positive effect on agricultural employment levels and growth. These findings are consistent with the prediction that land intensive occupations react less or even negatively to local transport improvements.

Our estimates for new occupations suggest that railways did contribute to pure employment growth in some cases. Distance to railway stations had a negative effect on employment growth for occupations associated with making chemicals, fuels, iron & steel, and machine tools, commerce and administration, and not surprisingly railway workers. Collectively employment growth in new secondary and tertiary occupations was similar or slightly larger than employment growth in all secondary and tertiary occupations. Our analysis of how railways affected the structural relationship between initial employment and employment growth suggests that reorganization was moderately large. We find that between one-third and one-half of the effects of station distance are due to the reorganization of employment from ares of low employment in 1851 to medium employment in 1851.

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 in the 19th century 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 it does not over-turn the idea that railways were one factor among

[3] For studies on the effects of railways on population density or income see Herranz-Loncán (2006), Donaldson (2014) for India, Jedwab et. al. (2015) for Africa, Hornung (2015) for Prussia, Donaldson and Hornbeck (2016) for the US, and Casson (2013) and Alvarez et. al. (2013) for Britain.

many in nineteenth century growth.

Our paper also contributes to the broader literature analyzing transport improvements and local economic development.[4] Despite the broad interest in the topic, there are only a handful of studies which produce comparable estimates to ours. We also provide new estimates on the relative importance of growth versus reorganization effects, and thus address one of the main challenges in the literature aside from endogeneity. Finally ours is one of the few papers to analyze the effects of transport improvements on employment structure. Most studies restrict their attention to population or total employment growth.

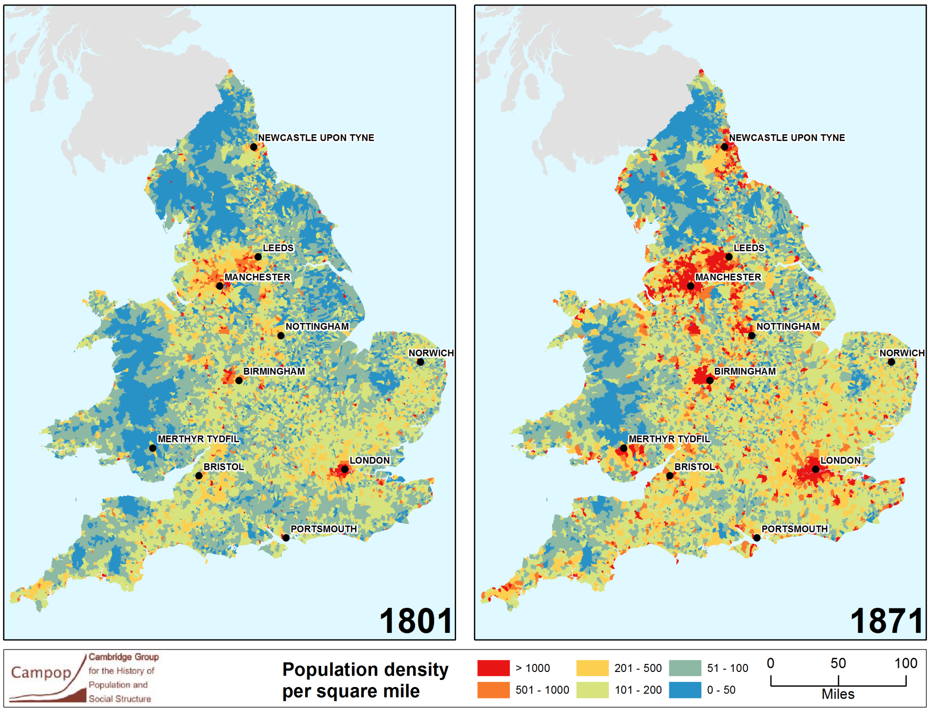
2 Background on population and employment

Building on the innovations of the eighteenth century, the English and Welsh economy experienced rapid economic growth in the early to mid nineteenth century. Census figures show that the English population increased from 8.6 million in 1801 to 17.0 million by 1851 and close to 22.3 million in 1881 (Shaw-Taylor and Wrigley 2014). Total male employment in England and Wales rose from 5.2 million in 1851 to 7.9 million in 1881. There was also significant change in male employment structure. Recently collected data for England and Wales suggests that 36% of males worked primarily in agriculture in 1817, but in 1871 it fell to 19%. From 1817 to 1871 male secondary rose from 44% of total employment in 1817 to 46% in 1871. Male tertiary employment increased from 18% to 28%. Mining employment rose from 3% to 6% over this same period.

The growth of population and employment had an important spatial component. The urban percentage of the population (population in towns of 5000 or more) rose from 29.5% in 1801 to 56.7% in 1871 (Shaw-Taylor and Wrigley 2014). London accounted for some of the urban growth with its percentage of the national population increasing from 11.2% in 1801 to 15.2% in 1871. The rest of urban growth came from outside the capital. The

[4] A selection of such studies includes Baum Snow (2007), Duranton and Turner (2011), Banerjee, Duflo, and Qian (2012), Faber (2014), Garcia-López et. al. (2015), Storeygard (2016).

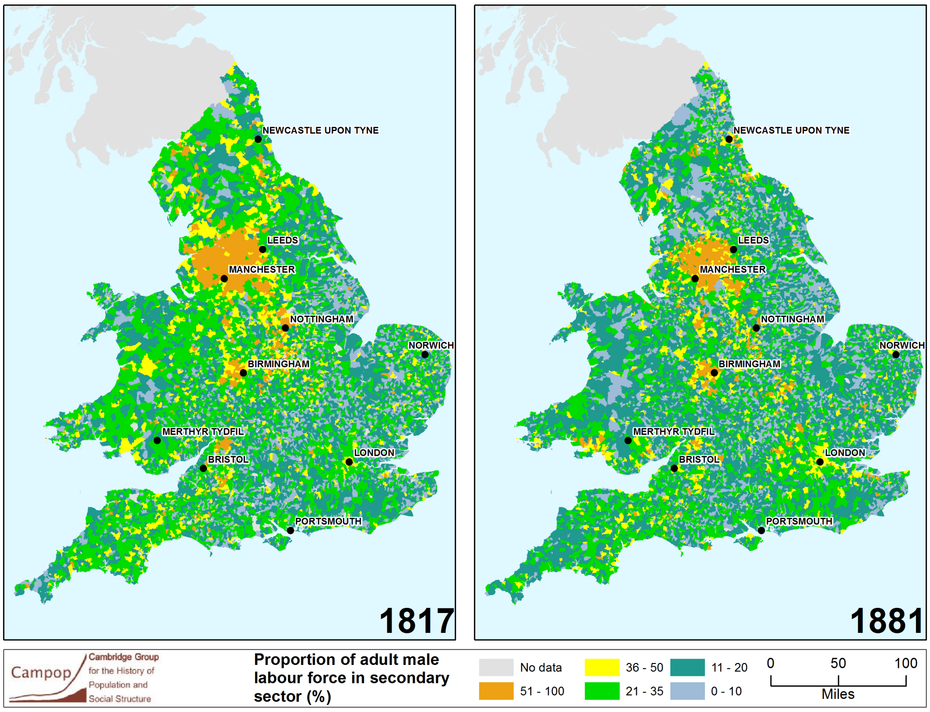
Figure 1: Population Density in England and Wales



most notable were the industrial towns in and near Manchester, Leeds, Nottingham, and Birmingham, and in the coal mining districts in and near Newcastle and Merthyr Tdyfil in South Wales. As an illustration, the population density of what we call parish mappable units are shown in the maps in figure 1 for 1801 and 1881. The units and the underlying data will be discussed later in the data section. The red areas correspond to the highest levels of population density. The growth of urban populations near London and the industrial or mining towns is evident. Also notable is that some areas saw little population growth, especially in north Wales, the southwest, the East Midlands, and the far northwest. In fact 22% of parish units experienced absolute population loss from 1801 to 1881.

The trends in total employment match the trends in population. But there were significant differences in the intensity of secondary and tertiary employment across space and over time. The shares of male secondary employment in each parish in 1817 and 1881 are shown in figure 2 below. These data are drawn from Kitson et. al. (2012) and Shaw-Taylor

Figure 2: Shares of males in secondary in England and Wales

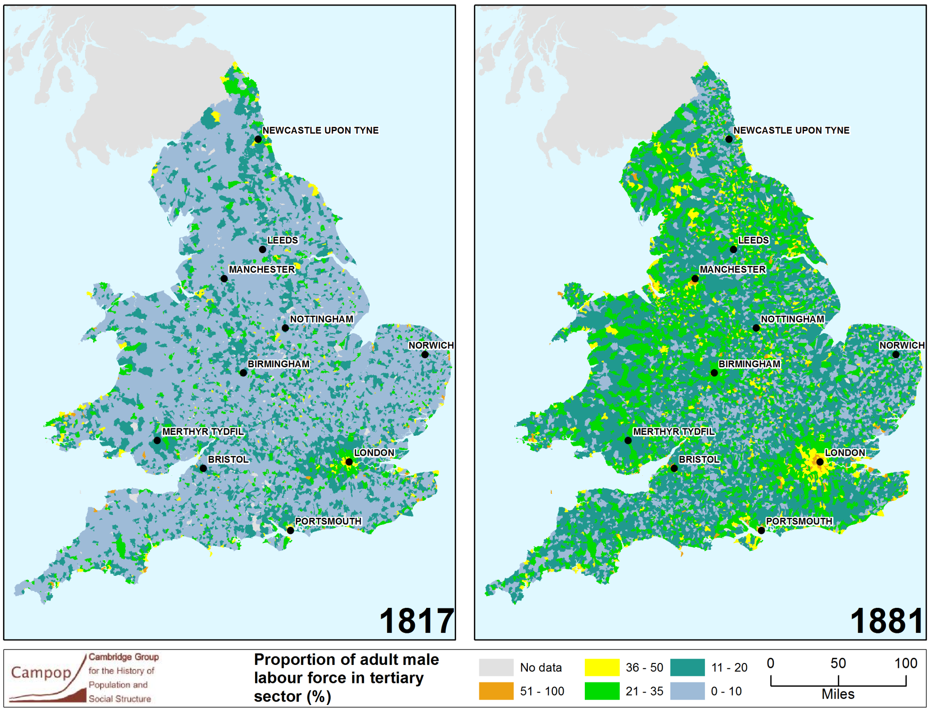


Sources: see text.

and Wrigley (2014) and are explained below. The spatial concentration of secondary employment was notably high near Manchester, Leeds, Nottingham, and Birmingham at both dates. Also notable is the diversification of employment in the north as the shares in secondary fall. 54% of the parish units saw a decline in its secondary share from 1817 to 1881, while the national totals increased around 2 percentage points.

Employment in the tertiary sector was more evenly distributed across regions, but also more concentrated near urban areas. The shares of males employed in the tertiary sector in 1817 and 1881 are displayed in figure 3. 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. The diversification of male employment away from secondary was partly connected with a transition into more tertiary. It is also remarkable that by 1881 tertiary employment became concentrated in similar areas as the secondary sector. Thus

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



Sources: see text.

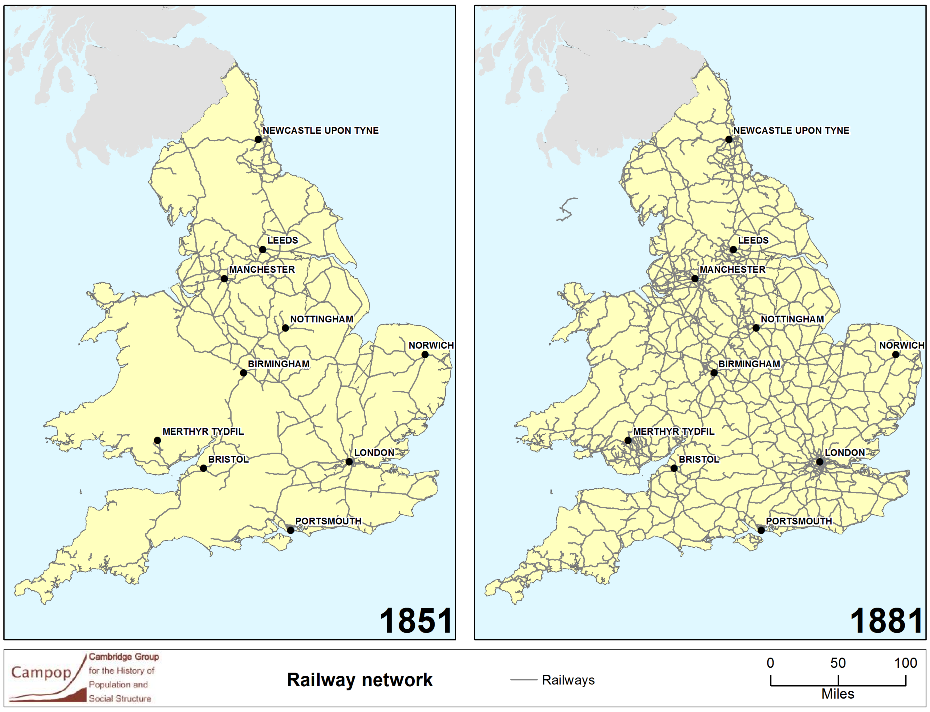
services and manufacturing employment tended to co-locate.

2.1 Background on railways

Britain was a pioneer in railway technology and construction. Inventors like Richard Trevithick and George Stephenson developed steam locomotion in the early 1800s. The first rail line using steam locomotion was opened in 1825 between Stockton and Darlington in the northern coal mining region. In 1830 the Liverpool and Manchester railway was opened to facilitate passenger traffic between the two large towns. It was promoted by local merchants and financiers who formed a joint company. They received authorization from parliament to build their line. Several other railways connecting nearby towns were promoted in the 1830s, but a national network had not yet formed.

In the mid-1840s England and Wales experienced a railway mania. In the span of a few years several hundred railway companies were proposed and approved. The plans called for nearly 15,000 km of railway track to be laid, but only around 10,000 km were built in the

Figure 4: Railway lines and stations in 1851 and 1881



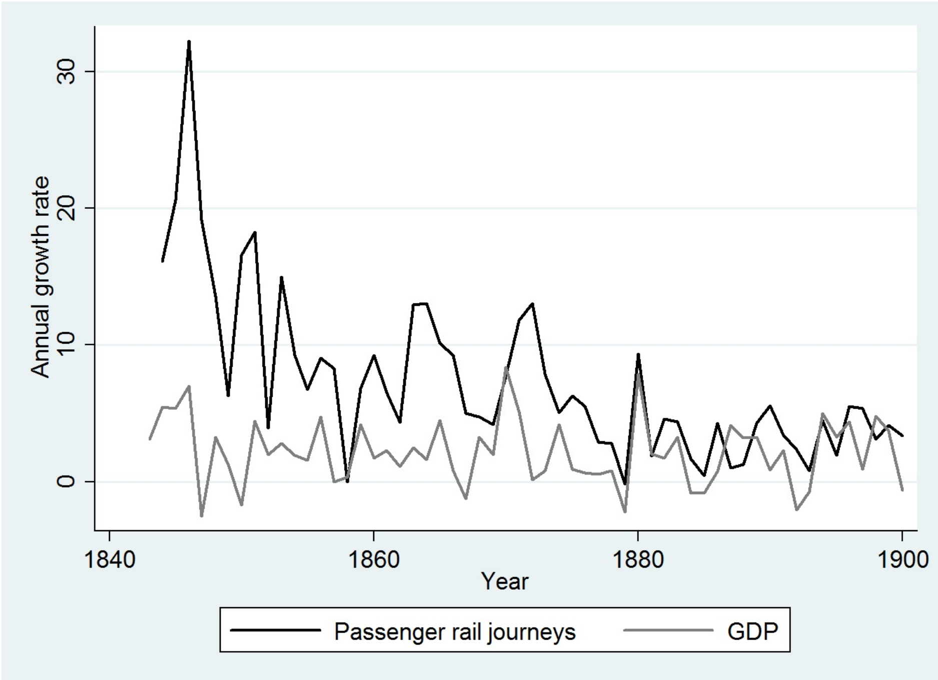
Sources: see text.

following five years. The railway mania was driven by speculation, and many investors lost money on bogus projects. It was also driven by interests of politicians in parliament who wanted to have railway stations in their constituency (Casson 2009, Odlyzko 2010). For our purposes, the railway mania produced the main trunk line network in England and Wales. The network in 1851 is shown in the left panel of figure 4. All of the major towns of England and Wales had a railway by this date, but only a few major cities like London, Birmingham, and Manchester were directly connected to one another.

Between 1851 and 1881 the railway network grew much larger. The network in 1881 is shown in the right hand panel of figure 4. By this date all major towns had connections not only with each other, but also with smaller towns in their region. There was also a process of amalgamation where larger railway companies came to control regions. By the First World War, there were 11 railway companies that accounted for most of the mileage and traffic. They included the likes of the Great Central Railway Company who controlled the northern East Midlands, the Great Eastern Railway Company who controlled the southern East Midlands, and the Great Northern Railway Company that controlled the main north-south routes to London.

Railways eventually became the most important form of passenger and freight transportation. Railways were far superior to coaches and wagons in both speed and cost. Railways also had a competitive edge over inland waterways, especially on speed. The only sector that remained competitive with railways was coastal shipping. The transition from road and inland waterway traffic to rail was not immediate however. First, the network had to be constructed, and many locations rail stations after 1851. Second, shippers and passengers that were accustomed to other modes of transport had to adjust to railways. The speed of the transition to rail can be seen by comparing growth rates of rail journeys with the growth in GDP. One would expect that the growth rate of rail journeys would be higher in the transition phase, but once that was complete then the growth rate of journeys should match the growth rate of GDP. The figures on numbers of passenger rail journeys are available annual from 1843 (see Mitchell 1971). The annual growth rate is shown in figure 5 along with the growth rate of British GDP. In the 1840s and 1850s the main transition to railway is underway. The transition process continues in the 1860s and 1870s as the growth rate in rail journeys is still higher. By the 1880s it appears that the transition is over. Rail journeys grow at the same rate as GDP. In our analysis below we take 1881 as the date at which railways had fully penetrated the economy. The year 1851 marks the date when the main network was in place but the transition to rail was still ongoing.

Figure 5: Growth rates of passenger rail journeys and British GDP



Sources: see text.

3 Methodology and empirical strategy

Theoretical models of trade, production, and consumption frame our empirical strategy for analyzing the effects of railways on employment. First, theory suggests that proximity to railway stations should affect population density which is proxy for overall employment density. For example, multi-location general equilibrium models which incorporate costly trade, commuting costs, and increasing returns in the production of differentiated goods yield a surprisingly simple expression for the equilibrium population. (see Helpman 1998, Redding and Sturm 2008, Redding 2012). Population at a location is log linear in functions for its productivity, its commuting technology, its non-traded amenities, and market access. Local transport improvements play a role in these models by changing a location's market access. Market access is the product of market access for consumers and for producers. The former is a weighted sum of the market supply of consumer goods with the weights being an decreasing function of transport costs between the location and each production center. The latter is a weighted sum of the market demands. If a local improvement lowers transport costs to all markets more than lowering other locations transport costs to all markets then it increases market access. Moreover, its share of exports in other markets should rise and its trade share within its own market should fall.

In our empirical analysis, we assume that greater distance to railway stations generally lowers market access for otherwise similar and nearby locations. To see why consider two nearby parish units identical in their productivity and amenities. Suppose there is a railway in the area and the station is closer to unit 1 than unit 2. Unit 2 will have lower market access and hence its population should be lower relative to unit 1. This hypothetical comparison ignores where railways originate and their destination, which is important for market access, but as long as the two units are near one another, then the connections provided by the railway station will be very similar. It also ignores other transport infrastructure like roads and waterways. In our regression we will control for these.

Building on this framework we first estimate the effects of distance to rail stations on population density. The so-called 'levels' equation is

$$y_{it} = \beta_1 \cdot DistRail_{it} + \beta_2 x_i + \alpha_j + \varepsilon_i \quad (1)$$

where y_{it} is the natural of log population density in unit i in time period t, $DistRail_{it}$ is the distance from unit i to its nearest railway station in time period t, x_i is a vector of geographic control variables serving as proxies for productivity and amenities. The controls also include variables for non-rail transport infrastructure like distance to inland waterways, ports, and turnpike roads. These can be thought of as additional factors that enter into market access and commuting costs. In the results, we compare distance to railways with distance to these other transport factors. The vector α_j includes dummy variables identifying whether unit i is in county j or registration district j with the latter being the more narrow jurisdictional classification. They capture unobservable differences across counties or districts, and allow us to control for external differences in market access (say being close to London versus Manchester) as well as other factors.

Their are three immediate concerns in estimating equation (1). First, the effects of distance to stations may not be linear. We address this issue by examine several alternative specifications for railway station access including a 3rd degree polynomial in distance and the log of distance. The second concern is that the vector of controls x_i does not adequately account for differences in productivity and amenities across units. We address this issue by running additional specifications where we include a vector of initial conditions. The most natural is y_{it-1} the log level of population density of the unit in the previous time period. Another is the rate of population growth over the previous two time periods Δy_{it-1}. Another is an indicator for the location being in the top quintile of population density in the previous time period $TopPopQuintile_{it-1}$. We add these three to the list variables serving as initial conditions in other specifications below. The third concern is endogenous placement of railway stations. Our solution is an instrumental variable for distance to railway stations based on geography and location between major towns. The details are discussed below.

A second variant on equation 1 examines population growth. The idea is that at any point in time locations are transitioning to a steady state population (see Duranton and Turner 2012). In this framework, population evolves according to the equation $Y_{t+1} = Y_t^{*\lambda} Y_t^{1-\lambda}$ where Y_t is the actual population level, Y_t^* is the steady state population level, and λ is the rate of convergence. Taking logs and rearranging gives the equation $y_{t+1} - y_t = -\lambda y_t + y_t^*$. If we replace y_t^* with the expression in equation (1) we get the following equation for employment growth between period $t+1$ and t

$$y_{it+1} - y_{it} = -\lambda y_{it} + \beta_1 \cdot DistRail_{it} + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{2}$$

where all variables are defined the same as in (1). Again it is possible to augment this specification with an additional lag of population density y_{it-1} to capture pre-trends. We can also include indicators for the top quintile of population density in year t and $t-1$.

The preceding focus on population density is useful because it sets up our analysis of sector differences in employment. The standard multi-location general equilibrium model does not incorporate multiple differentiated goods sectors. However, there are other spatial models which explicitly deal with two differentiated goods sectors, like agriculture and manufacturing, or manufacturing and services (Fujita et. al. 2001, Rossi-Hansberg 2005, Desmet and Rossi-Hansberg 2014). Most of these models consider locations along a line with the central segments corresponding to central locations, say in the middle of an island, and distant segments corresponding to frontier locations. The two sectors differ in land or labor intensity, and there are productivity spillovers from being located near more employment in the same sector. The spillovers generate employment clusters or areas of specialization. The labor intensive sector is more concentrated and dense in employment than the land intensive sector. Its concentration and density depends on the level of transport costs. In cases of sufficiently high transport costs overall production will be low and there were will be multiple clusters. In cases of sufficiently low transport costs, overall production will be high with fewer employment clusters. The main implication here is that employment density in labor intensive sectors like manufacturing and services should be higher in units closer to railway stations. By contrast, employment density in land intensive sectors like agriculture should be higher in units further from railway stations.

The effects of distance to railway stations on the levels and growth of employment density in different sectors are examined with equations (3) and (4) :

$$y_{it}^k = \beta_1 \cdot DistRail_{it} + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{3}$$

$$y_{it+1}^k - y_{it}^k = -\lambda y_{it}^k + \beta_1 \cdot DistRail_{it} + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{4}$$

where y_{it}^k is the natural of log population density in sector k in location i in time period t.

Additional variables for initial conditions, like log population density in the previous time period, are the same as above. In addition, in some growth specifications we include the employment shares for all but one of the sectors k in period t and the previous period $t-1$. Also we add indicators for employment density being in the top quintile in period t and $t-1$. These variables are meant to capture pre-trends in occupational structure.

3.1 Endogeneity

Endogeneity is a potential concern in the estimating equations considered thus far. The growth equations (2) and (4) have less endogeneity concerns because the distance to railway stations variable is measured in time period t and growth is measured from time period t to $t+1$. Nevertheless there is a still a worry that railways were placed in locations that were more or less likely to growth in the future. We address this issue using an instrumental variable (IV) for distance to railways stations. Our premise is that English railways were designed in large part to link cities and towns that traded the most in the early nineteenth century. Most of the locations along the route, which one could call inconsequential places, were close to railway stations simply because they were on the cost minimizing route connecting major towns.[5] We start with a simple gravity model (GM) equation to calculate the value of connecting English and Welsh cities with a population above 5000 in 1801. Our equation for town pairs i and j is $GM_{ij} = \frac{Pop_i Pop_j}{Dist_{ij}^2}$, where Pop_i is the 1801 population of town i. We consider all town pair connections with $GM_{ij} > 10,000$.

Next we identified a least cost path (LCP) connecting town pairs above the threshold. We assume that in considering their routes, railway companies tried to minimize the construction costs, especially 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. We

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

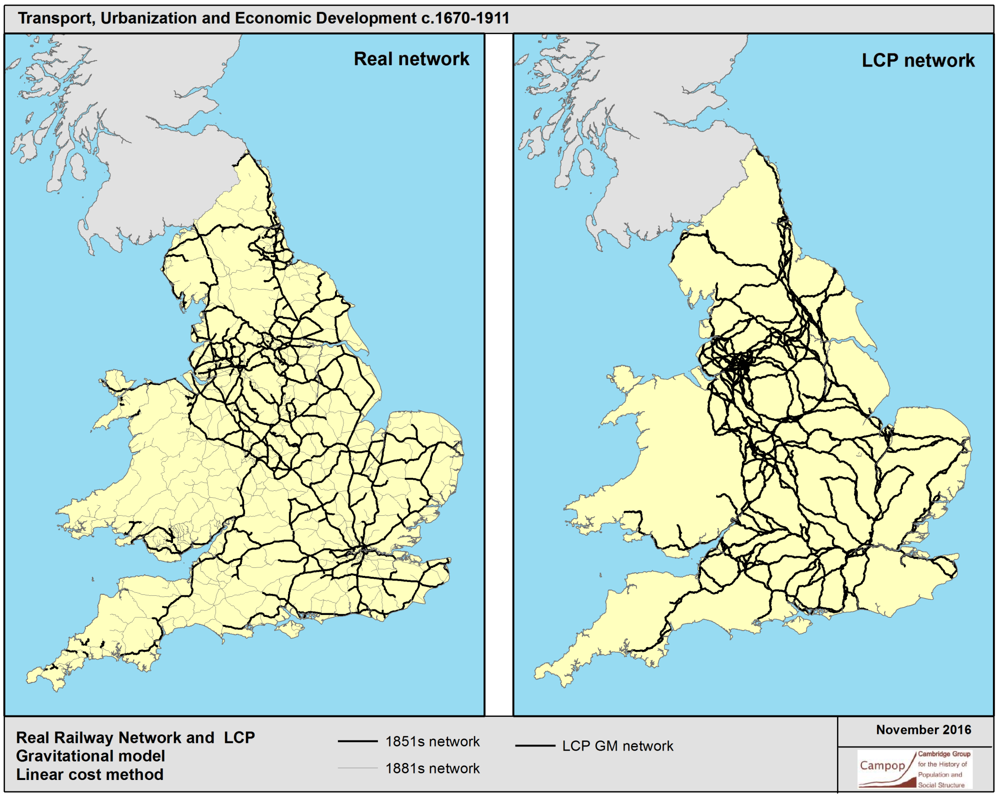
have examined several approaches to modeling the cost of elevation. Our baseline model builds on a 19th century engineer Wellington (1877) who estimated the relationship between construction costs and elevation. Starting with a normalization of construction costs per km at zero slope to be 1, Wellington argued that construction costs increased by a 2.96 for every % increase in slope. We use Wellington's formula to help identify the least cost path connecting our town pairs. Specifically used ESRI least-cost-path python schema in order to run the spatial analysis using the SRTM elevation raster of England and Wales, which specifies elevation in 90 meter cells. The tool calculates a least cost path (LCP) from a destination point to a source. The end result is a network of hypothetical railway lines linking towns, which we call the LCP network. It is shown in the right hand panel of figure 6. The left hand panel of figure 6 shows the real railway network in 1851 in black and the lighter lines are the rail network in 1881. The overlap of the LCP and the rail network is very high, especially in 1851. As we explain later in the data section, we use the distance between an unit's location and the LCP as the instrument for distance to railway stations.

3.2 Reorganization

The OLS and IV estimates reveal how employment density changes as distance to railway stations increases. As noted above, these effects could be due to growth or reorganization. We address the growth vs. reorganization issue using two approaches. First we examine the effects on employment in 'new' sectors or sectors with substantial growth. The idea is that in new sectors reallocation is less relevant because there was little initial employment to reallocate. If the coefficient on distance to railways is close to zero in specifications for new sector employment then growth effects appear less significant than reorganization effects.

In our second approach, we use Desmet and Rossi-Hansberg's (2009) model of dispersion and concentration in employment. Their model assumes production along a line with two sectors differing in the rate of technological progress and land intensity. The three important

Figure 6: Rail network and least cost path network



forces are (1) technology diffusion which leads to dispersion, (2) knowledge spillovers which favor concentration, and (3) land congestion which favors dispersion. This model produces a negative sloping S-shape relationship between a location's sector employment in period t and its sector employment growth from period t to $t+1$. At low levels of initial employment technology diffusion dominates making low employment areas growth more than any other. At high levels of initial employment congestion dominates making high employment areas growth the least. At a medium range of initial employment knowledge spillovers dominate and locations with more initial employment have larger increases in productivity from spillovers and grow more rapidly than areas with low initial employment. This last mechanism produces a 'hump' in growth rates in the middle of the distribution over initial employment.

We build on this model and examine how distance to railway stations changed the process of dispersion and concentration. One possibility is that units with the lowest initial employment grow less if they are close to railway stations because they face more competition with medium and large employment areas where knowledge spillovers are stronger. Medium employment areas may also grow more than they would otherwise because they take more more market share from the less productive low employment areas. In our empirical analysis we estimate the effects of railways on dispersion and concentration by modifying our employment growth equation (4). The modified equation is

$$y^k_{it+1} - y^k_{it} = \sum_{j=1}^{J} \lambda_j (y^k_{it})^j + \sum_{j=1}^{J} \eta_j (y^k_{it})^j DistRail_{it} + \beta_1 \cdot DistRail_{it} + \beta_2 x_i + \alpha_j + \varepsilon_i \qquad (5)$$

where the new terms are a J-degree polynomial in the log of initial employment and series of interaction terms between initial employment and distance to railway stations. In the case of $J = 2$ the marginal effect of increasing distance to rail stations is given by $\beta_1 + \eta_1 y^k_{it} + \eta_2 (y^k_{it})^2$. The first coefficient β_1 is what we will call the direct effect of increasing rail distance and $\eta y^k_{it} + \eta_2 (y^k_{it})^2$ is the effect of increasing rail distance at different initial employment levels, or what we will call the interaction effect. Our estimates will measure the size of the interaction effect in absolute value and compare it with the direct effect measured by β_1. If the interaction effect is relatively large then this would imply that railways effect's largely came from altering the structural process of reorganization.

4 Data

Population and male occupational data come from the British census. Population data are available in every decade starting in 1801. Detailed occupational data are available from 1851. This paper uses the 1851 and 1881 census which have been digitized for every parish by the Cambridge Group for the History of Population and Social Structure (CamPop). The occupational groups are classified into 5 broad categories including agriculture, secondary, tertiary, extractive, and a residual category called 'labourer'. 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. Extractive includes mining, fishing, and forestry. Labourer is a sector unspecific occupations, but it also includes cases where no occupation is stated or the occupational status is uncertain. There is a further classifications which includes 38 specific secondary occupations and 25 specific tertiary occupations. We identify 14 new or rapidly growing occupations. These occupations are discussed below.[6]

This paper also uses a quasi census of male occupations c.1817 created by CamPop. The figures are drawn from baptismal records in parish registers. The baptismal records were coded by occupation to best match the classifications in the 1851 census. They provide a very good estimate of male occupational shares c.1817 before railways were introduced.[7] In this paper, we use the male population in each parish in 1821 and an estimate of males over the age of 15 to calculate the number of males in each occupational group and in each parish c.1817. The figures imply that total male employment grew from 3.67 million c.1817 to 5.25 million in 1851 and 7.86 million in 1881.

The population and male occupational data for 1817, 1851, and 1881 are linked to GIS using a consistent set of parish boundaries produced by CamPop. There were some parishes

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

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

that were sub-divided after 1810 and so it was necessary to assemble larger units to maintain a consistent boundary. In total we have 9489 parish units across the three periods 1817, 1851, and 1881. The average size of a parish unit is 15 square km. Maps of population and employment using these units were shown above.

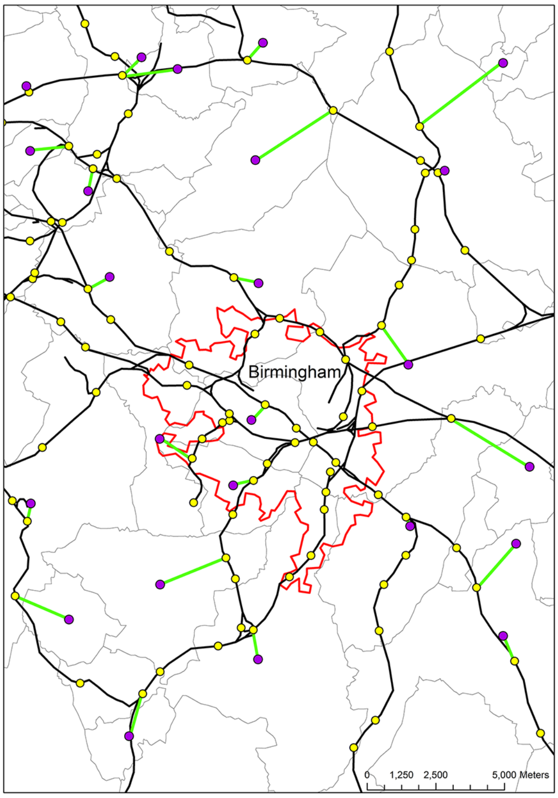
The paper also uses new GIS data on British transport networks.[8] Most importantly the data include the location of rail stations with their opening dates and closing dates by 1851 and 1881. We combine the data on parish units with the rail GIS to calculate the distance between the center of the parish and railway stations. The parish center is defined in two steps. If the parish had a market town at some point between 1600 and 1850 then the market town is taken as the center. This applies to 1425 parishes. 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 7 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.

The control variables include indicators for natural endowments. The endowment data include indicators for soil types, being on exposed coal fields, being on the coast, the average elevation, and ruggedness measures like the average elevation slope in the parish and the standard deviation in the elevation slope in the parish. Soils are divided into 9 general categories. The share in each category is included as a variable. Exposed coalfields are those where coal bearing strata are not concealed by 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.[9] The control variables

[8] See del Río, Martí-Henneberg, and Valentín (2008) for an initial description of the railways shape-file 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.

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

Figure 7: Birmingham area example



Sources: see text.

also include measures of distance to pre-railway transport networks dated around the year 1830. These include all ports, ports with steamship services, waterways, and turnpike roads. There are 54 counties in England and Wales, and each gets its own dummy variable as a control. In some specifications, we drop the county indicators and use registration district dummies. There are 616 registration districts, which implies the district dummies capture much more unobserved variation than the county indicators.

Table 1 reports summary statistics for the main variables. Note the soil shares are omitted and the county and registration district dummies are omitted. The summary statistics reveal two interesting facts: (1) the average distance to railway station declined significantly from 1851 to 1881. In 1851 the average distance to a station was 10.4 km. In 1881 the average distance was 3.8 km. Second the average unit experienced a small negative population growth of -0.8% between 1851 and 1881. It also experienced a somewhat larger negative secondary employment growth of -3% between 1851 and 1881. Given that national population and secondary employment rose this would imply greater spatial concentration of population and secondary employment.

5 Baseline analysis

We begin by analyzing the effects of distance to railway stations in 1881 on the level of population density in 1881. Table 2 reports results for various specifications. Column (1) is the most parsimonious. It shows a significant negative effect of distance to railway stations on population density. Adding 1851 population density and occupational shares in column (2) reduces the coefficient on railways indicating that these pre-trend variables are capturing some unobserved heterogeneity. The coefficient changes little in column (3) which includes registration district fixed effects rather than county fixed effects. In this specification increasing railway distance by 1 km decreases the log of population density in 1881 by -0.018. Using the property of logs this implies approximately a -1.8% effect on population density.

The instrumental variables (IV) results are shown in columns (4) and (5). The latter excludes units with towns used in constructing the LCP to ensure they are not driving the results. The first stage is very strong as indicated by the large F-statistic. The IV estimates are larger than OLS and suggest that increasing railway distance by 1 km decreased population density by 3.2%. It appears there is an upward bias in the OLS specification for railways and population density levels. The same pattern will be repeated in most specifications below. One explanation is that railways were designed to connect large towns, and sometimes railways were placed in other units that were less favorable for economic activity because it minimized the cost of purchasing land. Another explanation is that the railway companies were pressured by politicians to put railway stations in their constituency, perhaps more so if the constituency had low growth potential.

Panel B of table 2 shows results of distance to railway stations in 1851 on the growth of population from 1851 to 1881. In all specifications distance to railway stations significantly decreases the future growth of population. The estimates in column (3) with registration district fixed effects, geographic controls, pre-1830 characteristics, and 1851 population density and occupational shares show that 1 km greater distance to a station in 1851 decreases population growth by 0.55%. The IV estimates are smaller but still show a reduction in population growth of 0.47% moving 1 km distant from a station. The magnitudes of these effects will be discussed more below.

We ran similar specifications on the effect of distance to railway stations in 1851 and the population level in 1851. The results are not shown to save space. They indicate railways had a small effect on population levels in 1851. The coefficient implies a decrease in 1851 population density of -0.4% for every 1 km distance to stations. This finding is perhaps not surprising because the transition to rail transport was still underway in 1851.

We now turn to analyzing employment in different occupational categories. Panel A of table 3 shows regressions of 1881 log secondary employment levels on 1881 distance to railway stations. The 1851 occupational controls now include the log of own occupation employment, or in this case the log of secondary employment in 1851. As with population, including these pre-trends in employment density reduces the coefficient on railways and thus captures some unobserved heterogeneity. The OLS and IV specifications in columns (3) to (5) of panel A show that increasing distance to stations by 1 km reduces secondary employment levels in 1881 between 2.3 and 4.2%. The bottom of table 3 shows the effect on secondary employment growth. In specifications (3) to (5) the coefficients imply a growth effect between -0.46% and -0.63% for 1 km extra distance. Notice these estimates imply that railway access affected secondary employment growth in a similar direction and degree as population growth.

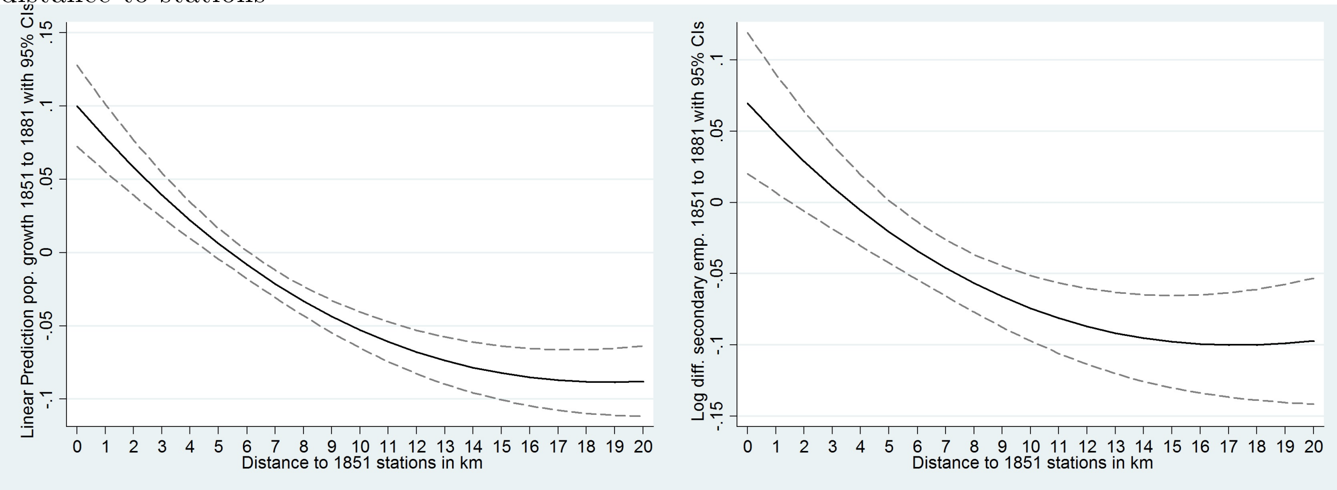
Table 4 shows regressions of tertiary employment on distance to railway stations. In the levels regression, the OLS coefficients indicate a -3.8% to -5.6% decline for every 1 km extra distance. The IV estimates are smaller, especially when units with towns that serve as nodes in the LCP network are dropped. The coefficients indicate a -3.0% to -3.2% decline for every 1 km extra distance, but the p-values are close to 0.10 indicating less precision. The estimates for the tertiary employment growth equation are shown in panel B. As in the levels equation, the coefficients vary and the IV estimates are less precise. The smallest IV coefficient implies a -0.45% decrease in population growth for every 1 km distance.

Agricultural employment should exhibit a very different relationship with railway access because this sector is more land intensive. Table 5 shows the same regressions for agricultural employment. In the levels regression the estimates vary, but they all imply close to a zero effect from distance to railway stations. The growth in agricultural employment is also largely unaffected by distance to railway station, although there is a positive and significant effect in column 3.

5.1 Alternative specifications of railway distance

The effects of railway distance may not be linear as assumed in previous specifications. In this subsection we consider alternative specifications. We start with a 3rd degree polynomial in distance to railway stations. We focus on the population and employment growth models and illustrate the results through graphs. The predicted log difference in population density is shown in the left panel of figure 8 over the range 0 to 20 km. The two end values approximately correspond to one standard deviation below and above the mean distance in 1851. Population growth is predicted to be 9.4% higher if a unit is exactly next to the station versus 5 km from the station. It is predicted to be 18.8% higher if the unit is exactly next to the station rather than 20 km distance.

Figure 8: Effects on population and secondary growth using 3rd degree polynomial for distance to stations



Sources: see text.

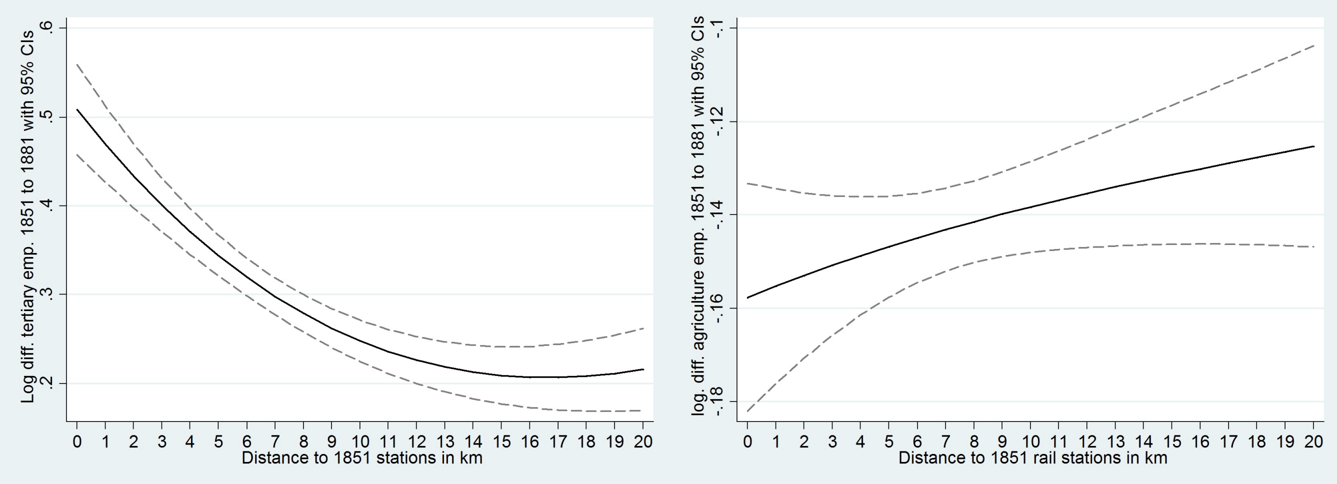
The predicted log difference in secondary employment from 1851 to 1881 over the range of 0 to 20 km is shown in the right panel of figure 8. Secondary employment growth is predicted to be 9% higher if a unit is exactly next to the station versus 5 km from a station, and 16.6% higher if the unit is exactly next versus 20 km distance. Notice these effects are larger in magnitude than for the linear specifications reported above. They suggest extra growth effects from being very close to the station.

Larger effects are also found on tertiary employment growth. The left panel of figure 9 shows the predicted log difference in tertiary employment from 1851 to 1881 over the range of 0 to 20 km. Tertiary employment is predicted to grow by 16% more if a unit is exactly next to the station versus 5 km from the station, and by 29.2% more if the unit is exactly next to the station versus 20 km distant.

Agriculture exhibits different results as expected (see figure 9 right-side graph). The prediction is that agricultural employment would be 3.2% lower for units exactly next to a station compared to those 20 km away but the confidence intervals imply the estimated change is imprecise.

The estimated effect of railway distance on extractive employment and labourer employ-

Figure 9: Effects on tertiary and agriculture growth using 3rd degree polynomial for distance to stations

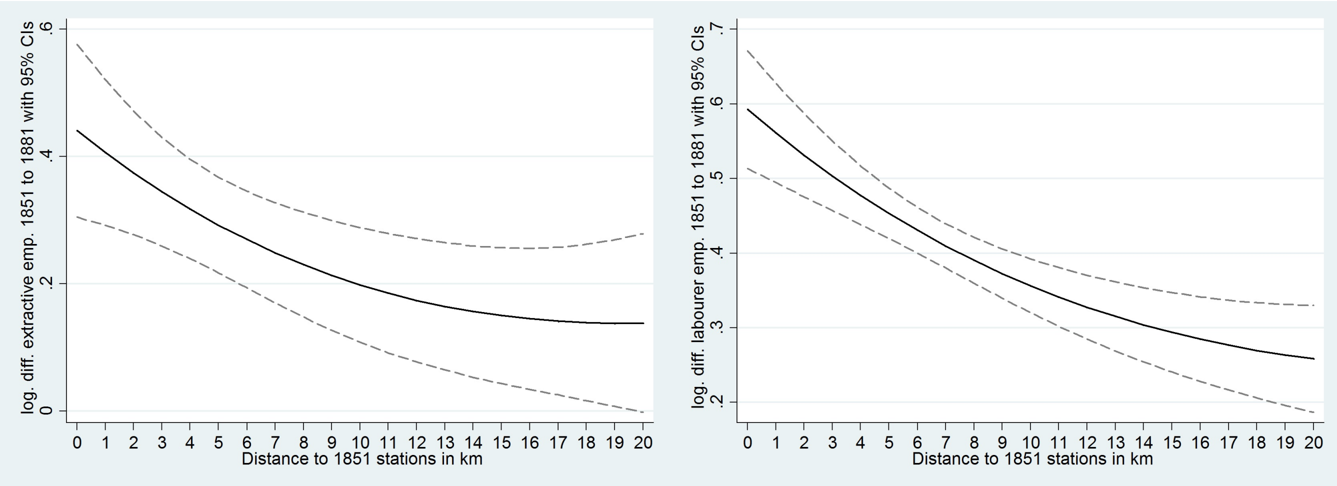


Sources: see text.

ment are summarized in figure 10. The effects on extractive employment, like mining, are quantitatively large although not precise. Units exactly near stations in 1851 had 30.2% more extractive employment growth than units 20 km from the station. The effects of railway access on labourer employment were quantitatively large and relatively precise. Units exactly near stations in 1851 had 33% more labourer employment growth than units 20 km from the station. The estimates for labourers are consistent with the general idea that railways had their largest effect on employment in labor intensive sectors.

We also estimate specifications using the log of distance to railway stations in km. These coefficients can be interpreted as an elasticity which gives additional interpretation to the results. Table 6 summarizes the population and employment levels results for the OLS specification with county fixed effects and the IV specification with county fixed effects. The conclusions are broadly similar to before with the IV estimates being larger and precisely estimated for population, secondary, and tertiary employment levels. Focusing on the IV estimates, a 10% increase in distance to 1881 rail stations would reduce population levels by 2.67%, it would reduce secondary employment levels by 1.35%, and tertiary employment levels by 4.09%. Table 7 summarizes the population and employment growth results for the

Figure 10: Effects on extractive and labourer growth using 3rd degree polynomial for distance to stations



Sources: see text.

same OLS and IV specifications. The IV estimates imply that a 10% increase in distance to 1851 rail stations would reduce population growth by 1.1% over the next 30 years, it would reduce secondary employment growth by 0.5% over the next 30 years, and tertiary employment growth by 1.7%. Alternatively, a one standard deviation increase in log distance to 1851 railway stations would reduce a location's secondary employment growth from 1851 to 1881 by 6.0% and it would reduce tertiary employment growth by 18.4%.

5.2 Summary and comparisons

It is useful to summarize what the estimates thus far suggest about the effects of railways. The mean secondary employment growth rate was -2.9% and the standard deviation was 91%. The mean tertiary employment growth rate was 39% and the standard deviation was 101%. Thus being next to a railway station raised a parish unit's growth substantially relative to the mean, but relative to the total variation in secondary and tertiary growth railway distance explains a small portion.

How important were railways compared to geographic factors or pre-railway infrastructure networks? We answer this question by comparing standardized coefficients.[10] The geographic factors are the indicators for having coal and being coastal. The pre-railway infrastructure variables are distance to inland waterways, ports with steamships, all ports, and turnpike roads. Table 8 reports the results for specifications using linear distance to stations and including registration district fixed effects, geographic controls, pre-1830 characteristics, and 1851 population density and occupational levels and shares. Distance to railway stations have the largest standardized coefficient in the regression on population growth, tertiary employment growth, and agricultural employment growth. Distance to railway station has the second largest standardized coefficient after distance to steamship ports for secondary employment growth. In other words, distance to railway stations was generally the most important factor among our main variables in explaining growth from 1851 to 1881. Note that railways are of moderate importance in explaining the growth of extractive employment. Having coal, being near a port, and being coastal explains most of the variation in extractive employment growth.

Is also useful to consider how outcomes would have been different if railway history unfolded differently. Railways were an experimental technology in the 1820s. It was only by the 1830s that railways were clearly superior to horse drawn wagons and canal boats (Dyos and Aldcroft 1969). Instead suppose that early railways proved to be too costly to operate and build, and no more railways were built after 1841. How would growth in England and Wales been different from 1851 to 1881? We answer this question by comparing our model's predicted level of growth given the railway network of 1851 with our model's predicted level of growth assuming England and Wales kept its 1841 network.[11] The results

[10] For example, the standardized coefficient for distance to railway stations is the distance to 1851 stations coefficient multiplied by the standard deviation of railway distance divided by the standard deviation in the growth variable. The same standardized coefficient is calculated for distance to inland waterways, ports with steamships, all ports, and turnpike roads, and also for the indicators having coal or being coastal.

[11] Specifically we calculate $\widehat{y_{i1881} - y_{i1851}}(rail1851)$ which is the predicted log difference in growth for each unit using the rail network of 1851. We then take exponential of the predicted growth which gives the

are summarized in table 9. With the 1851 network population is predicted to grow by 36.04 percentage points, and with the 1841 network population is predicted to growth by 21.38 percentage points. The difference in population growth, 14.65 percentage points less, is significant. We also find that with the 1841 network secondary employment is predicted to grow 8.52 percentage points less, tertiary employment 25.1 percentage points less, and agriculture 2.89 percentage points more. Based on these estimates economic growth, and especially tertiary employment growth would have been substantially different if railway technology proved non-economic in the 1830s.

More insights can be gained by comparing our estimates with others in the literature. Storeygard (2016) estimates an elasticity of African city population with respect to transport costs at different distances to ports. At the median distance, Storeygard estimates an elasticity of -0.25. It is similar to our IV estimate of the elasticity of population density with respect to log distance to stations (-0.2679), which is a proxy for transport costs. Banerjee, Duflo, and Qian (2012) estimate the elasticity of local economic activity and employment in China with respect to distance to railways. Their elasticities for GDP, secondary, tertiary, and primary are -0.067, -0.096, -0.078, and -0.029 respectively. They are consistent with our OLS estimates, but smaller than our IV estimates. Banerjee, Duflo, and Qian (2012) also estimate the elasticity of annualized GDP growth with respect to distance to railways in China. Their estimates are not significantly different from zero, but still they provide a useful comparison. We convert their estimates into a growth effect after 30 years. Their estimates for population growth, secondary growth, tertiary growth, and primary growth are -0.06, -0.194, -0.036, and -0.009. These estimates are again comparable to our OLS estimates, but we find a larger effect of railways on tertiary employment growth and a

predicted ratio for population or employment: $\widehat{\frac{Y_{i1881}}{Y_{i1851}}}$. We then multiply by the 1851 value $Y_{i1851}$to get each unit's predicted population or employment level in 1881 $\widehat{Y_{i1881}}$. Finally we sum over all units to the national predicted population or employment. The same calcluations is done for$\widehat{y_{i1881} - y_{i1851}}(rail1841)$ which is the predicted log difference in growth for each unit using the rail network of 1841.

smaller effect on secondary employment growth.

Duranton and Turner study the effects of highways on population growth of U.S. cities from the 1980s to the 2000s. They find that a 10% increase in highway density near a city would raise population growth by 1.5% over the next ten years. Extrapolating to 30 years their estimates imply a 2.3% increase in population growth for a 10% increase in highway density. Our estimates for a 10% increase in distance to stations are smaller but of a similar magnitude. These comparisons suggest broad similarities in the effects of transport improvements across different contexts, but also some differences especially concerning secondary, tertiary, and agricultural employment growth.

6 Railways and the growth of new employment

In this section, we report the effects of railway distance on 'new' occupations growth. New occupations, like electricity, are interesting in their own right, but they also give clues about the relative importance of growth and reorganization effects. The idea is that reorganization effects are likely to be smaller in new occupations because little employment existed prior to railways. In our definition of 'new', the occupation could exist before railways but it had to experience substantial technological change or rapid employment growth from 1851 onward. Our list of new secondary occupations includes printing, glass making, instrument making, chemicals, fuel, iron & steel, machine tools, electrical goods, gas equipment, and railway vehicles. Together these occupations represented 17% of secondary employment in 1851 and 26.5% of secondary employment in 1881. They accounted for 32% of all secondary employment growth from 1851 to 1881. Our list of new tertiary occupations includes media, financial services, commercial and administrative services, and railway transport workers. Together they represented 11.8% of tertiary employment in 1851 and 22.9% of tertiary employment in 1881, and account for 28.3% of all tertiary employment growth from 1851 to

1881.

We estimate the effect of linear distance to 1851 stations on new occupation employment growth using the OLS specification with registration district fixed effects, geographic controls, pre-1830 characteristics, 1851 population density, and 1851 occupational levels and shares. The exception are occupations where employment is found in less than 500 units. Here district fixed effects are dropped because of the small number of observations. The results for new secondary occupations are reported in table 10. For chemical, fuel, iron & steel, and machine tools the estimates show a negative and significant effect of distance to stations on employment growth. The coefficients range between -0.010 for iron & steel and -0.079 for chemical, which implies that employment growth fell by 1 and 8% respectively for every 1 km distance from stations. In the remaining six new occupations the coefficient is negative in five, although the precision is lower.

We also combine all ten new secondary occupations into a single category and rerun the regressions. Table 11 reports the results for several specifications. The OLS coefficients for new secondary employment growth all show a negative and significant effect with the coefficients ranging from -0.049 to -0.011. The IV coefficients are in the middle of this range, although not as precisely estimated. For comparison the coefficients in the same specification using all secondary employment growth are shown at the bottom along with p-value for a hypothesis test of the new secondary coefficient being different from the all secondary coefficient. For the OLS models the effect of distance to railway stations on new secondary employment growth is significantly larger than all secondary employment growth. The coefficients are not significantly different in the IV models.

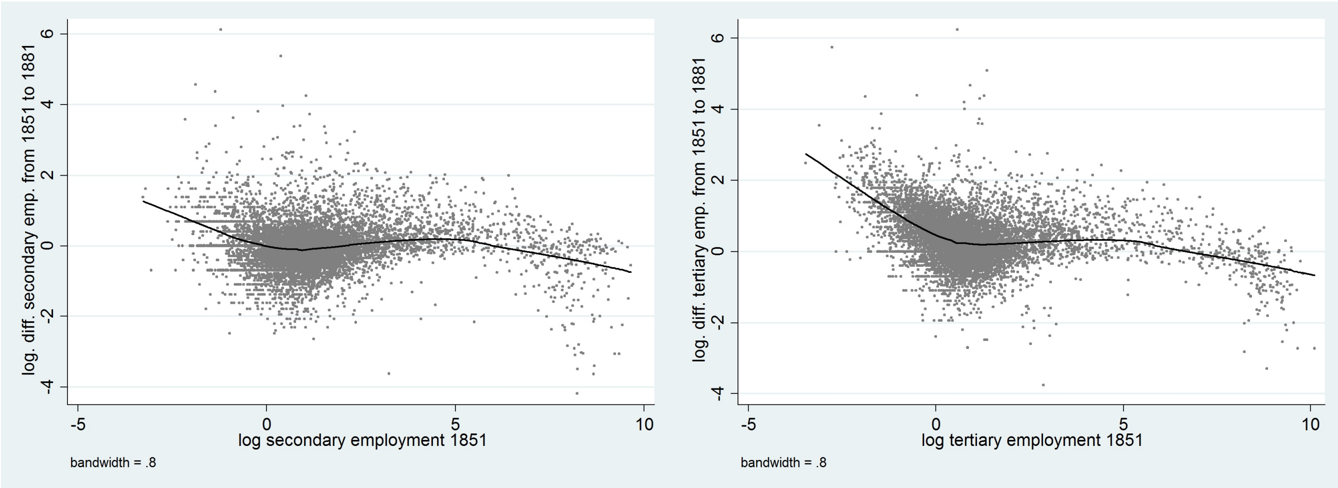
The results for new tertiary occupations are shown in table 12. For commerce & administration and railway workers the effect of distance to stations on employment growth is negative and statistically significant. For media and financial services the effects are also negative but not precisely estimated. The last column shows results combining all new tertiary occupations. The coefficient is significantly negative and larger than the coefficient in the same specification for all tertiary employment growth. The overall conclusion is that railways had some growth effects. They added some new employment to nearby parishes without drawing it away from other locations.

7 Railways and the reorganization of employment

Desmet and Rossi-Hansberg's model of employment growth shows how technology diffusion and congestion contribute to reorganization of employment from initially high to initially low employment areas. Their model also shows how knowledge spillovers can reverse the direction of reorganization in initially medium employment areas. In this section, we use their model to estimate how railways altered this 'structure' of reorganization. To begin, the relationship between secondary employment growth from 1851 to 1881 and the log of secondary employment in 1851 is fitted using locally-weighted least squares (lowess). The results are shown in the left panel of figure 11. The patterns fit what Desmet and Rossi-Hansberg (2009) find for new industries in the twentieth century U.S. The most rapid growth in secondary employment occurs in the lowest employment areas in 1851 but then at the medium levels the trend reverses. In areas of medium initial employment, employment growth is higher if initial employment is higher. The same relationships between tertiary employment growth from 1851 to 1881 and the log of tertiary employment in 1851 are fitted with lowess (see the right panel of figure 11). The patterns are broadly similar.

The specifications analyzed in the previous sections assumed a linear relationship between employment growth from 1851 to 1881 and the log of employment in 1851. We modify those specifications to include a 3rd degree polynomial in the log of 1851 employment in order to fit the S-shaped relationship displayed in the previous figure. An interaction between each term of the 3rd degree polynomial and the log of distance to railway stations in 1851

Figure 11: Plots of Employment growth and initial employment



Sources: see text.

is also included. The interaction terms are of special interest as they measure how railways changed the reorganization process at each initial employment level.

A summary of the elasticities for secondary employment are shown in panel A of table 13. The total elasticity from increasing rail distance one standard deviation above the mean is reported, along with the direct effect and the interaction effect. The total elasticity varies with the average value being -0.0029 close to what was reported earlier. The interaction effect is generally positive and largest at high and low levels of initial employment. This implies that reorganization effects tended to depress the employment enhancing effect of being close to stations at low and high initial employment levels. Also reported is the absolute value of the interaction effect and the ratio of the absolute value to the direct effect. The ratio is useful because it shows the relative importance of railway reorganization effects compared to direct effects. In the reported range, the average value of the ratio of interaction to direct effects is 0.63, but for most of the observations between -1 and 4 it was 0.29.

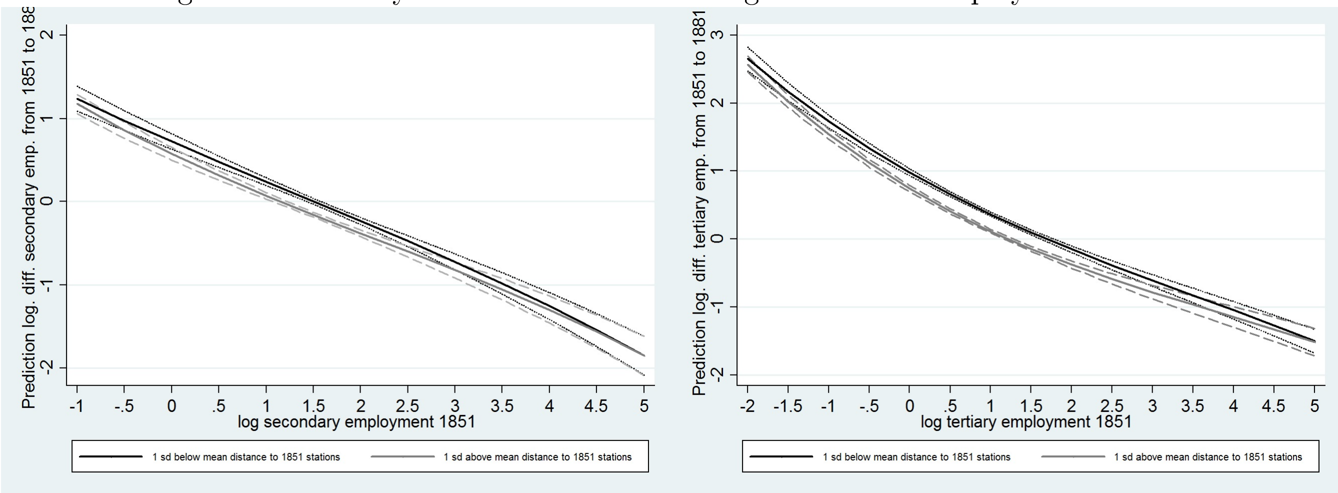
Reorganization can account for more of the effect in the case of tertiary. A summary is provided in panel B of table 13. The elasticity from increasing distance one standard deviation above the mean varies with the average value being -0.0051, which is a bit smaller than what we found earlier. The interaction effect is generally positive and again largest at low and high initial levels of employment. The average reported value of the ratio of interaction and direct effects is 0.57, and for most of the observations between -1 and 4 it was 0.20. The conclusion from both secondary and tertiary is that reorganization effects can account for one-fifth to three-fifths of the effect of railway access.

The results can be seen graphically. The left panel in figure 12 shows the predicted secondary employment growth from 1851 to 1881 as a function of 1851 secondary employment for two levels of railway access. The black curve shows the predicted growth when distance to railway stations is one standard deviation below the mean distance in 1851 and the gray curve shows the predicted growth when distance to railway stations is one standard deviation above the mean. At middle levels of initial secondary employment, secondary employment growth is predicted to be approximately 15% higher for locations at one standard deviation below the mean distance to stations in 1851 compared to locations one standard deviation above the mean distance to stations. The confidence intervals show that the predicted difference is precisely measured in the middle ranges of initial employment. The right hand panel of figure 12 shows the same for tertiary employment growth from 1851 to 1881. At the middle levels of tertiary employment, locations one standard deviation below the mean distance to stations are predicted to have approximately 24% higher tertiary employment growth.

8 Conclusion

How do transport improvements affect the long-term growth and spatial structure of employment? This paper answers these questions in the context of 19th century England and Wales where urbanization increased and the shares of secondary and tertiary employment rose. To address endogeneity we use an instrument that identifies locations close to railways

Figure 12: Railways and the structural reorganization of employment



Sources: see text.

because they were on a least cost path minimizing elevation changes and distance between large towns. The empirical analysis shows that a one standard deviation increase in distance to 1851 railway stations reduced secondary and tertiary employment growth in a location by 6.0% and 18.4% over the next 30 years. Quantitatively similar results are found for the effect of distance to stations in 1881 on population and employment levels in 1881, and new occupations in the mid-19th century. We also provide evidence that around one-third of the effects of station distance is due to the reorganization of employment from initially low to initially medium employment areas.

A Data Appendix

To be filled.

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. Banerjee, Abhijit, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in China. No. w17897. National Bureau of Economic Research, 2012.

3. Baum-Snow, Nathaniel. "Did highways cause suburbanization?." The Quarterly Journal of Economics (2007): 775-805.

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

5. 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.

6. 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.

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

8. 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.

9. 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.

10. 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.

11. 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.

12. Desmet, Klaus, and Esteban Rossi-Hansberg. "Spatial growth and industry age." Journal of Economic Theory 144.6 (2009): 2477-2502.

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

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

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

16. Duranton, Gilles, and Matthew A. Turner. "Urban growth and transportation." The Review of Economic Studies 79.4 (2012): 1407-1440.

17. Dyos, Harold James, and Derek Howard Aldcroft. British transport: an economic survey from the seventeenth century to the twentieth. Leicester University, 1969.

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

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

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

21. 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.

22. Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. The spatial economy: Cities, regions, and international trade. MIT press, 2001.

23. Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. "Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities." Journal of Urban Economics 85 (2015): 52-67.

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

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

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

27. Herranz-Loncán, Alfonso. "Railroad impact in backward economies: Spain, 1850–1913." The Journal of Economic History 66.04 (2006): 853-881.

28. Hettige, Hemamala. "When do rural roads benefit the poor and how?: an in-depth analysis based on case studies." (2006).

29. Hornung, Erik. "Railroads and growth in Prussia." Journal of the European Economic Association 13.4 (2015): 699-736.

30. Jedwab, Remi, Edward Kerby, and Alexander Moradi. "History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya." The Economic Journal (2015).

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

32. 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).

33. Lipscomb, Molly, Mushfiq A. Mobarak, and Tania Barham. "Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil." American Economic Journal: Applied Economics 5.2 (2013): 200-231.

34. Mitchell, Brian R. Abstract of British historical statistics. CUP Archive, 1971.

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

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

37. 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.

38. Storeygard, Adam. "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa." The Review of Economic Studies (2016): rdw020.

39. 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.

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

Table 1. Summary statistics, part 1

Variable	Obs	Mean	Std. Dev.	Min	Max
railway variables					
Distance to rail station km 1881	9489	3.8195	2.9253	0.0231	26.0047
Distance to rail station km 1851	9489	10.4564	11.0657	0.0215	73.1296
Distance to GM LCP km (IV)	9489	11.8619	16.5488	0.0001	116.3862
Population and occupation dependent variables					
Ln Pop. Density 1881	9489	4.2338	1.4215	0.6433	11.3002
Ln diff Pop. Density 1851 to 1881	9489	-0.0087	0.3765	-3.4044	4.0585
Ln secondary emp. 1881	9202	1.2762	1.8534	-3.7511	9.5371
Ln diff secondary emp. 1851 to 1881	9061	-0.03	0.6569	-4.1897	6.1203
Ln tertiary emp. 1881	9435	1.2974	1.7148	-2.6525	9.6347
Ln diff tertiary emp. 1851 to 1881	9321	0.3328	0.7035	-3.7612	6.2385
Ln agriculture emp. 1881	9418	2.107	0.7489	-1.7112	7.1897
Ln diff agriculture emp. 1851 to 1881	9403	-0.1382	0.4228	-3.1781	5.2364
Ln extractive emp. 1881	4810	-0.6337	1.9541	-4.8235	6.7842
Ln diff extractive emp. 1851 to 1881	3385	0.2752	1.1165	-3.912	7.5374
Ln labourer emp. 1881	8938	0.5392	1.7535	-3.8989	8.4803
Ln diff labourer emp. 1851 to 1881	8231	0.4024	1.1718	-3.8067	5.247

Sources: see text.

Table 1. Summary statistics, Part 2

Variable	Obs	Mean	Std. Dev.	Min	Max
1851 population and occupation controls					
Ln Pop. Density 1851	9489	4.2425	1.3673	0.8088	11.6253
Ln secondary emp. 1851	9222	1.3039	1.7556	-3.2755	9.6566
Ln tertiary emp. 1851	9362	0.9765	1.7621	-3.4681	10.1004
Ln agriculture emp. 1851	9449	2.2543	0.7663	-3.1699	7.7996
Ln extractive emp. 1851	4358	-0.7515	1.9174	-4.8644	6.62
Ln labourer emp. 1851	8586	0.1948	1.7672	-3.7992	8.7426
share secondary emp. 1851	9489	0.1964	0.1239	0	0.8
share tertiary emp. 1851	9489	0.1497	0.1095	0	0.9412
share extractive emp. 1851	9489	0.0255	0.0767	0	0.7451
share labourer emp. 1851	9489	0.075	0.0902	0	0.7603
Pre-railway occupation and population controls					
Ln diff Pop. Density 1801 to 1821	9485	0.1956	0.1918	-1.0594	1.9102
Ln Pop. Density 1801	9485	3.8776	1.3105	0.4833	11.4381
share secondary emp. 1817	9489	0.2017	0.1446	0	1
share tertiary emp. 1817	9489	0.0949	0.1141	0	1
share extractive emp. 1817	9489	0.0161	0.0672	0	0.8647
share labourer emp. 1817	9489	0.5048	0.2157	0	1
Geographic controls					
Distance to nearest large city in 1801 km	9489	136.3901	67.9921	0	418.7408
Indicator exposed coal	9489	0.0802	0.2716	0	1
Indicator coastal unit	9489	0.1479	0.355	0	1
average elevation slope within unit	9489	4.7675	3.6157	0.4849	37.4272
SD elevation slope within unit	9489	3.4324	2.7174	0	23.1755
elevation unit	9489	89.7215	74.0256	-1.243	524.3845
Pre-railway transport infrastructure					
Distance to nearest inland waterway 1830 km	9489	7.2316	6.5016	0	48.3873
Distance to nearest steamship port 1840 km	9489	85.0676	44.058	0	267.7452
Distance to nearest general port km	9489	30.2513	22.9766	0.0592	99.7121
Distance to nearest turnpike road km	9489	1.2302	1.4749	0	15.3485

Sources: see text.

Table 2: Distance to railway stations and population density

	Panel A: 1881 log population per sq. km.				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1881	-0.0299	-0.0168	-0.0183	-0.0321	-0.0316
	(0.00)	(0.00)	(0.00)	(0.004)	(0.004)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occ. shares	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				119.704	117.777
R-square	0.9135	0.9567	0.9643		
N	9485	9485	9485	9485	9386
	Panel B: log difference pop. density 1851 to 1881				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1851	-0.0022	-0.0015	-0.0055	-0.0047	-0.0046
	(0.003)	(0.037)	(0.00)	(0.003)	(0.003)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occ. shares	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				1007.55	997.85
R-square	0.3315	0.3713	0.4835		
N	9485	9485	9485	9485	9485

Notes: Standard errors are clustered on county or registration district in OLS. Robust stand. err. are reported in IV. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer. For explanation of instrument see text.

Table 3: Distance to railway stations and secondary employment density

	1881 secondary employment per sq. km.				
	1	2	3	4	5
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1881	-0.039	-0.0196	-0.0234	-0.0417	-0.0415
	(0.00)	(0.00)	(0.00)	(0.028)	(0.03)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				121.746	119.753
R-square	0.8524	0.9163	0.9261		
N	9199	9199	9199	9199	8959
	log difference secondary employment 1851 to 1881				
	6	7	8	9	10
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1851	-0.0021	-0.0001	-0.0046	-0.0063	-0.0062
	(0.109)	(0.882)	(0.018)	(0.026)	(0.027)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				988.672	978.472
R-square	0.1802	0.3346	0.4126		
N	9058	9058	9058	9058	8959

Notes: Standard errors are clustered on county or registration district in OLS. Robust stand. err. are reported in IV. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log secondary employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer. For explanation of instrument see text.

Table 4: Distance to railway stations and tertiary employment density

	1881 tertiary employment per sq. km.				
	1	2	3	4	5
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1881	-0.0561	-0.0382	-0.0436	-0.0327	-0.0305
	(0.00)	(0.00)	(0.00)	(0.091)	(0.114)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				120.859	118.820
R-square	0.8524	0.9083	0.9215		
N	9317	9317	9317	9317	9218
	log difference tertiary employment 1851 to 1881				
	1	2	3	4	5
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1851	-0.0004	-0.0033	-0.0077	-0.0049	-0.0045
	(0.809)	(0.048)	(0.00)	(0.088)	(0.111)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				998.532	988.356
R-square	0.1802	0.4376	0.5180		
N	9317	9317	9317	9317	9218

Notes: Standard errors are clustered on county or registration district in OLS. Robust stand. err. are reported in IV. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log tertiary employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer. For explanation of instrument see text.

Table 5: Distance to railway stations and agricultural employment density

	1881 agricultural employment per sq. km.				
	1	2	3	4	5
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1881	-0.0027	-0.0002	0.0032	0.0075	0.0051
	(0.087)	(0.869)	(0.024)	(0.449)	(0.605)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				117.666	115.594
R-square	0.7898	0.8409	0.8658		
N	9414	9,399	9,399	9,399	9302
	log difference agricultural employment 1851 to 1881				
	6	7	8	9	10
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1851	-0.0030	-0.0000	0.0017	0.0011	0.0007
	(0.068)	(0.977)	(0.067)	(0.449)	(0.605)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				993.583	983.367
R-square	0.0945	0.5066	0.5836		
N	9,399	9,399	9317	9317	9302

Notes: Standard errors are clustered on county or registration district in OLS. Robust stand. err. are reported in IV. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log agricultural employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer. For explanation of instrument see text.

Table 6: Distance to stations, population and employment levels: elasticities

Panel A: OLS				
	1	2	3	4
	Ln Pop.	Ln Sec. emp.	Ln Tert. emp.	Ln Agric. emp.
	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)
Ln Distance to rail station in km 1881	-0.0666	-0.0752	-0.1500	0.0069
	(0.000)	(0.000)	(0.000)	(0.185)
County fixed effects	Y	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y

Panel B: IV estimates				
	5	6	7	8
	Ln Pop.	Ln Sec. emp.	Ln Tert. emp.	Ln Agric. emp.
	Stand. coeff. (p-value)	Stand. coeff (p-value)	Stand. coeff (p-value)	Stand. coeff (p-value)
Ln Distance to rail station in km 1881	-0.2679	-0.1357	-0.4097	0.0618
	(0.000)	(0.089)	(0.000)	(0.187)
County fixed effects	Y	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y

Notes: for definitions of controls see tables… Standard errors are clustered on the county in OLS and robust standard errors in the IV.

Table 7: Distance to stations, population and employment growth: elasticities

Panel A: OLS				
	1	2	3	4
	ΔLn Pop.	ΔLn Sec. emp.	ΔLn Tert. emp.	ΔLn Agric. emp.
	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)
Ln Distance to rail station in km 1851	-0.0314	-0.0180	-0.0666	0.0084
	(0.000)	(0.048)	(0.000)	(0.111)
County fixed effects	Y	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y

Panel B: IV estimates				
	5	6	7	8
	ΔLn Pop.	ΔLn Sec. emp.	ΔLn Tert. emp.	ΔLn Agric. emp.
	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)	Coeff. (p-value)
Ln Distance to rail station in km 1851	-0.1098	-0.0552	-0.1680	0.0257
	(0.000)	(0.088)	(0.000)	(0.186)
County fixed effects	Y	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y

Notes: for definitions of controls see tables… Standard errors are clustered on the county in OLS and robust standard errors in the IV.

Table 8: Determinants of population and employment growth: standardized coefficients

	1 Pop. growth Stand. coeff. (p-value)	2 Sec. growth Stand. coeff (p-value)	3 Tert. Growth Stand. coeff (p-value)	4 Ag. Growth Stand. coeff (p-value)	5 Extr. Growth Stand. coeff (p-value)
Distance to rail station in km 1851	-0.162	-0.078	-0.121	0.046	-0.090
	(0.000)	(0.004)	(0.000)	(0.032)	(0.074)
Distance to inland waterway in km 1830	-0.034	-0.012	0.001	0.007	-0.016
	(0.025)	(0.489)	(0.945)	(0.637)	(0.631)
Distance to steamship port in km 1840	-0.046	-0.105	-0.020	0.036	0.053
	(0.451)	(0.112)	(0.729)	(0.523)	(0.699)
Distance to any port in km 1840	0.041	0.034	0.047	-0.006	-0.122
	(0.289)	(0.438)	(0.241)	(0.852)	(0.142)
Distance to turnpike in km 1830	-0.048	-0.066	-0.045	-0.002	0.021
	(0.000)	(0.000)	(0.000)	(0.758)	(0.278)
Indicator for having exposed coal	0.059	0.054	0.007	-0.033	0.197
	(0.000)	(0.000)	(0.502)	(0.001)	(0.000)
Indicator for having being a coastal unit	0.075	0.078	0.069	-0.022	0.099
	(0.000)	(0.000)	(0.000)	(0.054)	(0.001)
County fixed effects	N	N	N	N	N
Reg. district fixed effects	Y	Y	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	Y

Notes: for explanations of variables see tables 2-9.

Table 9: National Counter-factual assuming England &Wales kept rail network of 1841

Model predicted population growth 1851 to 1881 in % with 1851 network	36.04
Counter-factual population growth 1851 to 1881 in % with 1841 network	21.38
change in pop. Growth in %	-14.65
Model predicted secondary emp. growth 1851 to 1881 in % with 1851 network	61.15
Counter-factual secondary emp. growth 1851 to 1881 in % with 1841 network	52.63
change in sec. emp. growth in %	-8.52
Model predicted tertiary emp. growth 1851 to 1881 in % with 1851 network	68.64
Counter-factual tertiary emp. growth 1851 to 1881 in % with 1841 network	43.4
change in tert. emp. growth in %	-25.19
Model predicted agricultural growth 1851 to 1881 in % with 1851 network	-11.81
Counter-factual agricultural growth 1851 to 1881 in % with 1841 network	-8.92
change in agric. emp. growth in %	2.89

Notes: for sources and details of calculations see text.

Table 10: Distance to railway stations and growth in 'new' secondary employment from 1851 to 1881

	1	2	3	4	5
	Printing	Glass making	Instrum. making	Chemical	Fuel
	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)
Distance to rail station in km 1851	-0.0027	0.0072	-0.0021	-0.0797	-0.0456
	(0.858)	(0.774)	(0.791)	(0.087)	(0.098)
OLS	Y	Y	Y	Y	Y
Reg. district fixed effects	Y	N	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y	Y
R-square	0.6924	0.3333	0.6700	0.7389	0.7511
N	1223	264	1424	594	820
	6	7	8	9	10
	Iron & steel	Machine tool	Electrical goods	Gas equip.	Rail vehicle
	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)
Distance to rail station in km 1851	-0.0105	-0.0173	-0.0053	-0.0125	-0.0734
	(0.00)	(0.039)	(0.597)	(0.133)	(0.524)
OLS	Y	Y	Y	Y	Y
Reg. district fixed effects	Y	Y	N	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y	Y
R-square	0.4463	0.5651	0.3659	0.4843	0.6405
N	7356	2617	243	369	99

Notes: Standard errors are clustered on county or registration district in OLS except in column 2, 8, 9, and 10, where robust standared errors are used. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log own sector employment (e.g. log fuel employment) log secondary employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer.

Table 11: Distance to railway stations and growth of all new secondary employment

	log difference *new* secondary employment 1851 to 1881.				
	1	2	3	4	5
	OLS	OLS	OLS	IV	IV
	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Distance to rail station in km 1881	-0.0062	-0.0049	-0.0116	-0.0056	-0.0051
	(0.003)	(0.005)	(0.000)	(0.144)	(0.177)
County fixed effects	Y	Y	N	Y	Y
Reg. district fixed effects	N	N	Y	N	N
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	N	Y	Y	Y	Y
Include units with towns in LCP	Y	Y	Y	Y	N
Cragg-Donald Wald F statistic				815.109	805.535
R-square	0.1801	0.3698	0.4602		
N	7645	7645	7645	7645	7546
	log difference secondary employment 1851 to 1881				
	coeff.	coeff.	coeff.	coeff.	coeff.
Distance to rail station in km 1851	-0.0021	-0.0001	-0.0046	-0.0063	-0.0062
P-value for hypothesis test of coefficient in New secondary model different from coefficient in all secondary model	0.041	0.006	0.007	0.859	0.794

Notes: Standard errors are clustered on county or registration district in OLS. Robust stand. err. are reported in IV. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log secondary employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer. For explanation of instrument see text.

Table 12: Distance to railway stations and growth in ‘new’ tertiary employment from 1851 to 1881

	1	2	3	4	5
	Media	Financial services	Commerce and admin.	Railway transp.	All new tertiary
	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)	coeff. (p-value)
Distance to rail station in km 1851	-0.0119	-0.0089	-0.0166	-0.0283	-0.0263
	(0.703)	(0.244)	(0.031)	(0.008)	(0.000)
OLS	Y	Y	Y	Y	Y
Reg. district fixed effects	Y	N	Y	Y	Y
Controls for geography and pre-1830 char.	Y	Y	Y	Y	Y
Controls for 1851 pop. density and occupation	Y	Y	Y	Y	Y
R-square	0.8010	0.6516	0.6352	0.6108	0.5072
N	737	1937	2565	2758	4496

Notes: Standard errors are clustered on county or registration district in OLS except in column 2, 8, 9, and 10, where robust standared errors are used. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, and share of soil types. Pre-1830 characteristics include pop. density in 1801, pop. growth from 1801 to 1821, an indicator for top quintile of pop. density in 1801, and distance to nearest large city, inland waterway, port, steamship port, and turnpike road. Pre-1830 also includes employment shares in secondary, tertiary, extractive, and labourer, and also indicators for top quintile of employment in secondary, tertiary, extractive, agriculture, and labourer. Controls for 1851 include log own sector employment (e.g. log fuel employment) log tertiary employment, pop. density, top quintile for pop. density, secondary, tertiary, agriculture, extractive, and labourer employment, and finally employment shares in secondary, tertiary, extractive, and labourer.

Table 13: Railways and reorganization of employment

Panel A: Secondary employment

log secondary emp. in 1851	Elasticity from increasing 1851 rail distance 1 SD	direct effect of increasing distance	interaction effect of increasing distance	interaction effect abs. value	ratio abs. interaction effect to direct effect
-2	0.05142	-0.07438	0.1258	0.1258	1.69135
-1	-0.03198	-0.07438	0.0424	0.0424	0.5701
0	-0.07438	-0.07438	0	0	0
1	-0.08518	-0.07438	-0.0108	0.0108	0.14515
2	-0.07375	-0.07438	0.00063	0.00063	0.00846
3	-0.04949	-0.07438	0.02489	0.02489	0.33465
4	-0.02178	-0.07438	0.0526	0.0526	0.70721
5	0	-0.07438	0.07438	0.07438	0.99996
6	0.00645	-0.07438	0.08083	0.08083	1.08669
7	-0.01181	-0.07438	0.06257	0.06257	0.84122

Panel B: Tertiary employment

log tertiary emp. in 1851	Elasticity from increasing 1851 rail distance 1 SD	direct effect of increasing distance	interaction effect of increasing distance	interaction effect abs. value	ratio abs. interaction effect to direct effect
-2	-0.03955	-0.12021	0.08067	0.08067	0.67103
-1	-0.09308	-0.12021	0.02713	0.02713	0.2257
0	-0.12021	-0.12021	0	0	0
1	-0.12514	-0.12021	-0.00493	0.00493	0.04099
2	-0.11206	-0.12021	0.00815	0.00815	0.06781
3	-0.08517	-0.12021	0.03504	0.03504	0.29148
4	-0.04867	-0.12021	0.07154	0.07154	0.59512
5	-0.00676	-0.12021	0.11346	0.11346	0.94378
6	0.03637	-0.12021	0.15659	0.15659	1.30257
7	0.07652	-0.12021	0.19674	0.19674	1.63655

Notes: see text.