Railways and growth: evidence from nineteenth 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 growth and spatial structure of population and employment? This paper answers these questions in the context of nineteenth century England and Wales. It analyzes the effect of railways on population and employment growth following a large expansion in the railway network during the 1840s. Endogeneity is addressed using a hypothetical railway network that minimizes elevation changes and distance between large towns in 1801. Results show that population growth and secondary and tertiary employment growth were significantly higher near railway stations. By contrast, agricultural employment growth was significantly lower near stations. Results also show that the growth effects of railways were larger in medium density localities in 1851 and those with coal. Railways appears to magnify agglomeration economies and natural resource advantages.

Keywords: Economic growth, 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

‖Data for this paper was created thanks to grants from the Leverhulme Trust grant (RPG-2013-093) Transport and Urbanization c.1670-1911 and NSF (SES-1260699), Modelling the Transport Revolution and the Industrial Revolution in England. We thank Walker Hanlon for comments on earlier drafts and seminar participants at UC Irvine, UC San Diego, NYU, Florida State, Trinity College Dublin, Queens Belfast, and the Economic History Association Meetings. All errors are our own.

1 Introduction

Large scale transport improvements have the potential to drive economic growth and change the spatial distribution of economic activity. Much of the discussion of transport improvements in the literature focuses on three main questions. (1) How large are the effects of new railways, highways, ports, and other transport improvements on growth? (2) Do locations with better access to transport attract certain types of employment, like manufacturing or services? (3) Do transport improvements increase growth more in high density areas and less in low density areas?[1] This paper examines these questions by studying the effects of railways in nineteenth century England and Wales. This historical context is well suited because railways marked a major improvement over inland road and water transport. They generated large savings through lower freight rates, lower passenger fares, and faster travel speeds. Railways in England are also thought to have attracted population to towns near stations, and possibly depopulated areas more distant from stations.[2]

We provide new evidence by analyzing population and employment data in more than 9000 parish units. They are the smallest administrative unit in England and Wales with consistent boundaries over the nineteenth century. Population totals are observed every 10 years from 1801 to 1881 and the same for all male occupations in 1851 and 1881. Parish male occupations are classified into five general groups: (1) secondary, (2) tertiary, (3) agriculture, (4) extraction or mining, and (5) labourer, which is a non-specific occupation by 1851. Additional specialties within secondary and tertiary are also studied to better understand manufacturing and service employment.

The parish data are linked with new spatial data on transportation networks and endowments.[3] The data include GIS shapefiles of railway lines and stations, ports, the turnpike

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

[2] For literature on the social savings of railways in England and Wales see Hawke (1970) and Leunig (2006). For works on railways and population growth see Gregory and Marti Hennenberg (2010), Alvarez et. al. (2013), and Casson (2013).

[3] For more details see the project on Transport, urbanization and economic development in England and

road network, inland waterways, exposed coal fields, elevation, coastal boundaries, and soils. Combining all the data we measure the distance between parish units and railway stations in 1851 and the distance to other infrastructure and endowments pre-dating railways. The focus on 1851 is useful because the railway network underwent a major expansion in the mid 1840s due to a 'railway mania.' By 1851 the trunk-line network was formed, and much of the expansion in railways over the next few decades occurred near the early trunk lines.

Our empirical model analyzes the effect of distance to stations in 1851 on parish unit population and employment growth from 1851 to 1881. Although we include a rich set of controls, the distance to railwy stations could be correlated with omitted variables related to growth. To address this issue we construct a hypothetical railway network connecting large towns in 1801. The routes are chosen to minimize elevation changes and distance known as the least cost path (LCP).[4] As we show below distance to the LCP provides a strong instrumental variable for distance to the actual railway network.

Our estimates show that proximity to railway stations increased parish population growth, but they attracted different types of employment. Agricultural employment growth increased with distance to the nearest station. By contrast secondary and tertiary employment growth decreased with distance to the nearest station, implying that railways attracted secondary and teritary employment. Extractive employment and labourers also decreased with distance to stations but the estimates are less robust. In terms of magnitudes, our main specification shows that a one standard deviation increase in log distance to the nearest 1851 railway station lowers population, secondary, and tertiary employment growth by 13, 10, and 13 percentage points respectively, and raises agricultural employment growth by 6 percentage points. The effects of railways were even larger in rapidly growing secondary and tertiary occupations, like machine tools, iron & steel, finance, commerce, and administration.

In two key extensions, we test whether railways had different effects depending on ini-

Wales c.1670-1911 at http://www.campop.geog.cam.ac.uk/research/projects/transport/

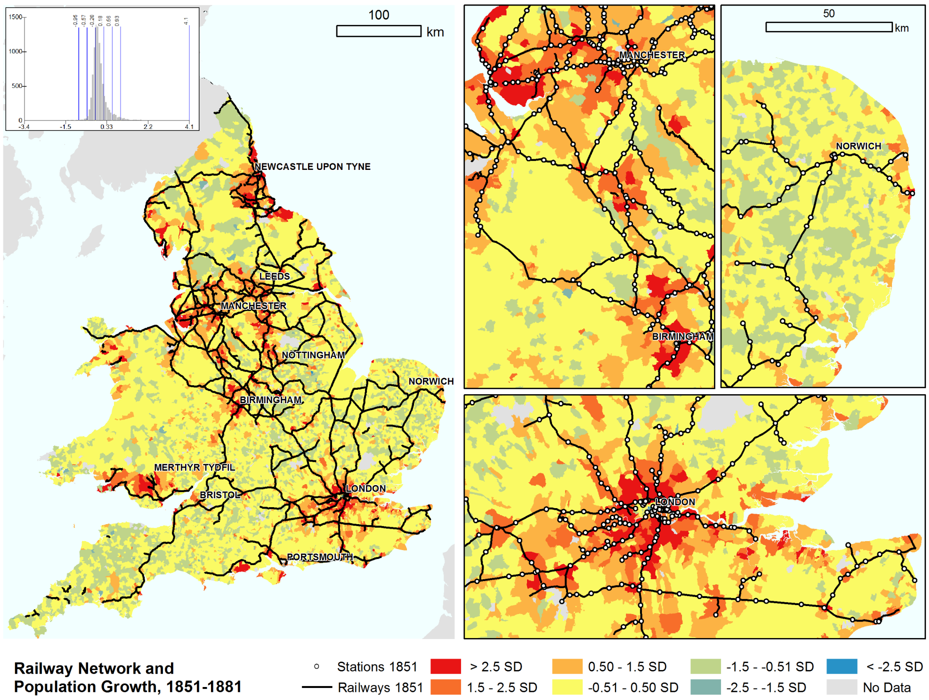
[4]Our methodology draws on other studies which least cost paths as instruments for infrastructure. See Faber (2014) and Lipscomb et. al. (2013) for two examples.

tial density and coal endowments. The largest effects on population growth are found in parishes in the 65th percentile of 1851 population density. The largest effects for secondary employment growth are found in the 45th percentile of 1851 employment density. In both population and secondary employment the effects of railways were also larger in parishes with coal. Railways thus favored growth in 'medium' density units and those with natural resource advantages. The exception is tertiary employment where parishes in the 15th percentile of 1851 employment gained as much from being near railways as the 55th or 85th percentile. Also coal did not significantly influence the effects of railways on tertiary employment growth.

As a visual preview of the results, figure 1 shows the location of railway lines and stations in 1851 along with each parish unit's population growth from 1851 to 1881. At a national level it is clear that areas which grew rapidly were close to railway stations. The same pattern is seen at a regional level near Manchester, Birmingham, and London. However, the East Anglia region in the upper right shows that having railway stations nearby did not guarantee high population growth. Our estimates suggest the effects in East Anglia were different in part because it had no coal and it had lower population density in 1851 than the Manchester and London regions.

How much did railways contribute to growth in England and Wales? To address this question we first compare the estimated effects of railways with the transport networks that preceeded them in history. We find that railways generally had larger effects on population and employment growth than turnpike roads, inland waterways, and general ports. The exception is ports with steamships which often had an equal effect to railways. The effects of railways are also gauged through a counterfactual exercise. Population and employment growth are predicted assuming the railway network had remained at its 1841 level. Railways were still an experimental technology in the 1830s, and had they proved non-economic, railway building might have stopped as early as 1841. Our counter-factual estimates suggest

Figure 1: Railways and local population growth from 1851 to 1881



Sources: see text.

that national population, secondary and tertiary employment growth would be 8.5, 8.9, and 4 percentage points lower with the 1841 network rather than the 1851 network. The effects are large for population and secondary growth which increased by a total of 40 and 52 percent from 1851 to 1881.

The findings contribute to several literatures. The first addresses the spread of industrialization 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 been tested using regions as the unit or they focus on the textile industry (Crafts and Mulatu 2006, Crafts and Wolf 2014). This paper uses much smaller units and covers a wider range of economic activities.

This paper also adds to a large number of historical studies on railways. Some works focus on social savings and the direct benefits of lower transport costs (Fogel 1964, Fishlow 1965, Hawke 1970, Crafts 2004). Others analyze their effects on local population density and agricultural income.[5] 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 findings imply that railways had significant effects, but they were not the sole cause of nineteenth century growth. In fact, some mid-nineteenth century growth is explained by proximity to earlier transport networks, like roads, inland waterways, and ports.

Our paper also contributes to the broader literature analyzing transport improvements and development. First, we provide a historical comparison for estimates from modern settings.[6] Second this paper connects to the theoretical literature on the mechanisms linking transport improvements and externalities. We relate our findings to the model of industry location developed by Desmet and Rossi-Hansberg (2009, 2014). They show how knowledge spillovers can offset congestion effects and drive growth in medium and large density areas. Our finding that railways contributed to growth more in medium density areas suggests transport improvements magnified knowledge spillovers. The further finding that areas with coal had more growth from being close to railways suggests that transport improvements also magnified natural resource advantages.

2 Background on population and employment

In the 1600s and 1700s the English economy experienced the beginnings of industrialization, but more rapid growth did not occur until the 1800s. Census figures show that the

[5]See Herranz-Loncán (2006) for Spain, Donaldson (2014) for India, Jedwab et. al. (2015) for Africa, Berger and Enflo (2015) for Sweden, Tang (2014) for Japan, Hornung (2015) for Prussia, Donaldson and Hornbeck (2016) for the US, and Gregory and Marti Hennenberg (2010), Casson (2013), and Alvarez et. al. (2013) for Britain.

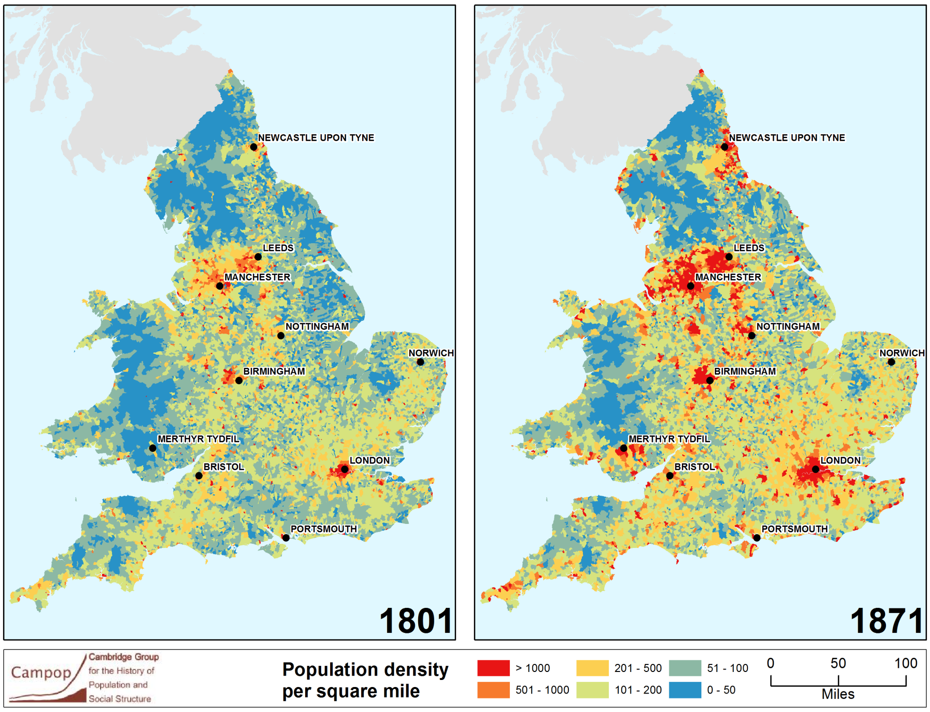
[6]A selection of such studies includes Baum-Snow (2007), Duranton and Turner (2011), Banerjee, Duflo, and Qian (2012), Faber (2014), Garcia-López et. al. (2015), and Storeygard (2016).

English population increased from 8.6 million in 1801 to 17.0 million by 1851 and close to 22.3 million in 1881. Total male employment in England and Wales rose from 5.2 million in 1851 to 7.9 million in 1881. Male employment structure also significantly changed. Agriculture is the main form of employment in most pre-industrial economies. Data for England and Wales shows that 36% of males worked primarily in agriculture in 1817, but in 1871 it fell to 19% (Shaw-Taylor and Wrigley 2014). Secondary is often the sector that absorbs labor leaving agriculture. Secondary refers to the transformation of the raw materials produced by the primary sector into other commodities, whether in a craft or a manufacturing setting. From 1817 to 1871 male secondary employment rose in England and Wales but not by much, increasing from 44% to 46% of the total. Tertiary employment, encompassing all services including transport, shop-keeping, domestic service, and professional activities, is the sector that experienced the most change. Male tertiary employment increased from 18% to 28% from 1817 to 1881. Mining or extractive employment also rose significantly from 3% to 6% in the same period.

The growth of population and employment had a spatial pattern. The urban percentage of the population (people 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 mainly came from the northern industrial towns near Manchester, Leeds, Nottingham, and Birmingham, and in the coal mining districts near Newcastle and South Wales.

As an illustration, the population density of parishes are shown in figure 2 for 1801 and 1871. Parishes shown in red have the highest levels of population density, blue the lowest. The growth of urban populations near London and the industrial towns is evident. So is the stagnant growth in some areas like north Wales, the southwest, the East Midlands, and the far northwest. In fact 22% of parish units experienced absolute population loss from 1801

Figure 2: Population Density in England and Wales



Sources: see section 4 for details.

to 1881.

The trends in secondary employment match the trends in population. Already in the early nineteenth century, secondary employment was concentrated in the northern industrial towns and London. That concentration increased by the late 1800s, especially in the north. Tertiary employment density was low everywhere in the early 1800s, except for London. By the late 1800s century tertiary was more common everywhere, but growth was a bit higher near the large northern towns. Agricultural employment density was low throughout the period and declined in most locations.

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. Railway locomotives were far superior to coaches and wagons in both speed and cost. For example, rail freight rates in 1870 were one-tenth road freight rates in 1800 in real terms (Bogart 2014). Railways also had a competitive edge over barges on inland waterways, especially in speed. The only alternative transport sector that remained competitive with railways was coastal shipping. Steamships were invented in the early nineteenth century and transformed coastal and international shipping much like railways transformed transportation.

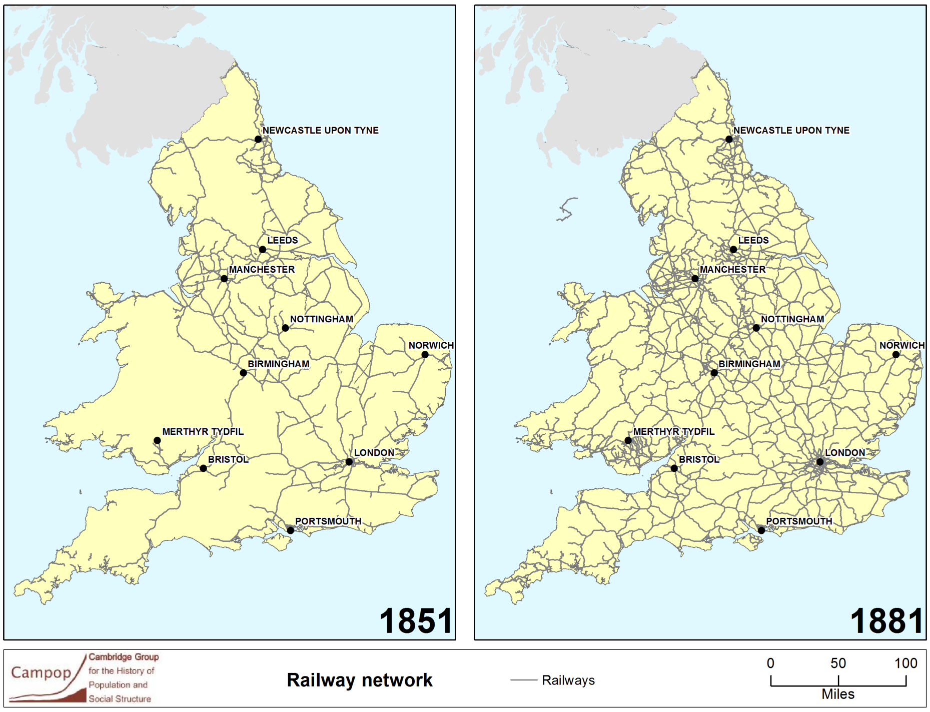
The first rail line using steam locomotion was opened in 1825 between Stockton and Darlington in the northern coal mining region. In 1830, the Liverpool and Manchester railway was opened to facilitate passenger traffic between the two large towns. It was promoted by local merchants and financiers who 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.

England experienced a railway mania in the mid-1840s. Several hundred railway companies were proposed by local groups and approved by parliament. Proposals called for nearly 15,000 km of railway track to be laid, but around 10,000 km were built in the following five years. The expansion of new routes was part of a strategy of the early railways to maintain their incumbent positions. It was also driven by MPs who wanted to have railway stations in their constituency. The investing public was also keen to invest in railway stocks, and those who bought at the peak of the mania lost dearly.[7]

For our purposes, the railway mania is significant because it led to a large expansion in railway building and produced the main trunk line network in England and Wales. The network in 1851 is shown in the left panel of figure 3. All of the major towns of England and Wales had a railway by this date, and most had a train service to London. Below we will focus on the effects of the railway network in 1851 on growth between 1851 and 1881. There are several points to make about this choice. First, we will partly avoid the complication that railways were built in response to growth, which is more likely to be the case in the

[7]For the literature on the railway mania see Casson (2009), Odlyzko (2010), Campbell and Turner (2012, 2015)

Figure 3: Railway lines and stations in 1851 and 1881



Sources: see text.

1860s and 1870s. Second, it is true that more railways were built between 1851 and 1881 (see the right hand panel of figure 3), but areas with a more dense railway network in 1851 tended to have a more dense network in 1881 too. Early railways bred later railways to a degree. Third, the transition to railways was far from complete in 1851. Most locations had recently got their first rail station by 1851. Shippers and passengers that were accustomed to other modes of transport had just began to adjust to railways by 1851. Thus we expect the growth effects of 1851 railways were mainly realized in the following decades.

3 Methodology and empirical strategy

Our empirical strategy is framed by urban and trade models which incorporate costly trade and increasing returns in production.[8] Equilibrium population at a location is generally shown to be increasing in productivity, non-traded amenities, and market access. Following Redding and Turner (2014), equilibrium population can be written in a simplified form as $Y_t^* = A_t M_t^b E^a$ where Y_t^* is the equilibrium population in period t, A_t is productivity in period t, M_t is a function of consumer and firm market access in period t, E are endowments, and a and b are parameters. Consumer market access is a decreasing function of transport costs between the location and each center supplying consumer goods. Producer market access is decreasing in transport costs between the location and centers of demand. Most models imply that increasing market access raises population so that $b > 0$.

In many contexts where infrastructure changes substantially, it is natural to assume that local economies are adjusting to their steady state population level. Specifically we build on an assumption from Duranton and Turner (2012) that the rate of adjustment depends on how far one is out of equilibrium. The adjustment equation is given by $Y_{t+1} = Y_t^{*\lambda} Y_t^{1-\lambda}$ where Y_t is the actual population level in period t, Y_t^* is the steady state population level, and λ is the rate of convergence. After taking logs and rearranging we get the expression: $y_{t+1} - y_t = -\lambda y_t + \lambda y_t^*$, where lower case represents the natural log of a variable. Notice that if $\lambda = 1$ convergence to the steady state is instantaneous: $y_{t+1} = y_t^*$.

The next step is to substitute the natural log of $Y_t^* = A_t M_t^b E^a$ into the adjustment equation. The resulting expression is $y_{t+1} - y_t = -\lambda y_t + \lambda a + \beta m_t + \alpha e$, where the outcome $y_{t+1} - y_t$ is the log difference in population from period t to $t+1$, m_t is the natural log of market access, a is the log of productivity, e is the log of the amenities, and the parameters are $\beta = \lambda b$ and $\alpha = \lambda a$. It implies that higher market access in period t will increase population growth from t to $t+1$. The magnitude depends on the convergence parameter

[8] See Helpman (1998), Redding and Sturm (2008), Redding (2012), Durantan and Turner (2012) for theoretical models.

λ and the structural parameter b.

In our empirical analysis, we do not calculate market access directly. Instead, an approximation is made with distance to railway stations and other spatial controls, including registration district fixed effects. We assume that if a parish unit has greater distance to a railway station then it has lower market access compared with nearby locations. To see why this assumption is reasonable, consider two nearby parish units. 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 by the model above its population growth should be lower relative to unit 1. Our formulation abstracts from 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 will be very similar and are captured by a district fixed effect. Our empirical analysis will also control for other transport infrastructures. They too will create local differences in market access and are compared with railways.

Building on the preceding framework our baseline estimating equation is the following:

$$y_{i1881} - y_{i1851} = -\lambda y_{i1851} + \beta_1 r_{i1851} + \beta_2 x_i + \alpha_j + \varepsilon_{it} \tag{1}$$

where y_{it} is the natural of log population density in unit i in time period t, r_{i1851} is the log distance from parish unit i to its nearest railway station in 1851, x_i is a vector of geographic control variables including indicators for having coal, being coastal, and ruggedness. They serving as proxies for productivity and amenities. The controls also include variables for distance to the nearest large town in 1801 and distance to inland waterways, ports, and turnpike roads. α_j is a fixed effect for registration district j. The fixed effects control for external differences in market access (say being close to London versus Manchester) as well as other factors like unobserved productivity and amenities common among all units in a district. The standard errors are clustered on registration districts. The clustering addresses correlation in unobservables within districts.

There are four major concerns in estimating equation (1). First, the effects of distance to stations may not be log linear. We address this issue with alternative specifications for railway station access including a 4th degree polynomial in distance to stations. Second, there may be pre-trends in population which are correlated with railway access in 1851. Below we include population growth from 1821 to 1851 ($y_{i1851} - y_{i1821}$) as an additional variable to control for pre-trends. A third concern is that growth is not linear in the log of 1851 population density y_{i1851}. In our estimating equation we allow for non-linearities by including a 3rd degree polynomial in the log of 1851 population. The augmented estimating equation is

$$y_{i1881} - y_{i1851} = \beta_1 r_{i1851} + \sum_{q=1}^{3} \lambda_q \cdot (y_{i1851})^q + \beta_2 x_i + \alpha_j + \varepsilon_{it} \tag{2}$$

where $(y_{i1851})^2$ and $(y_{i1851})^3$ are the added second and third order terms for log of 1851 population.

The fourth concern is 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 larger 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 larger towns.[9] We start with a simple gravity model (GM) equation to calculate the value of connecting English and Welsh towns 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

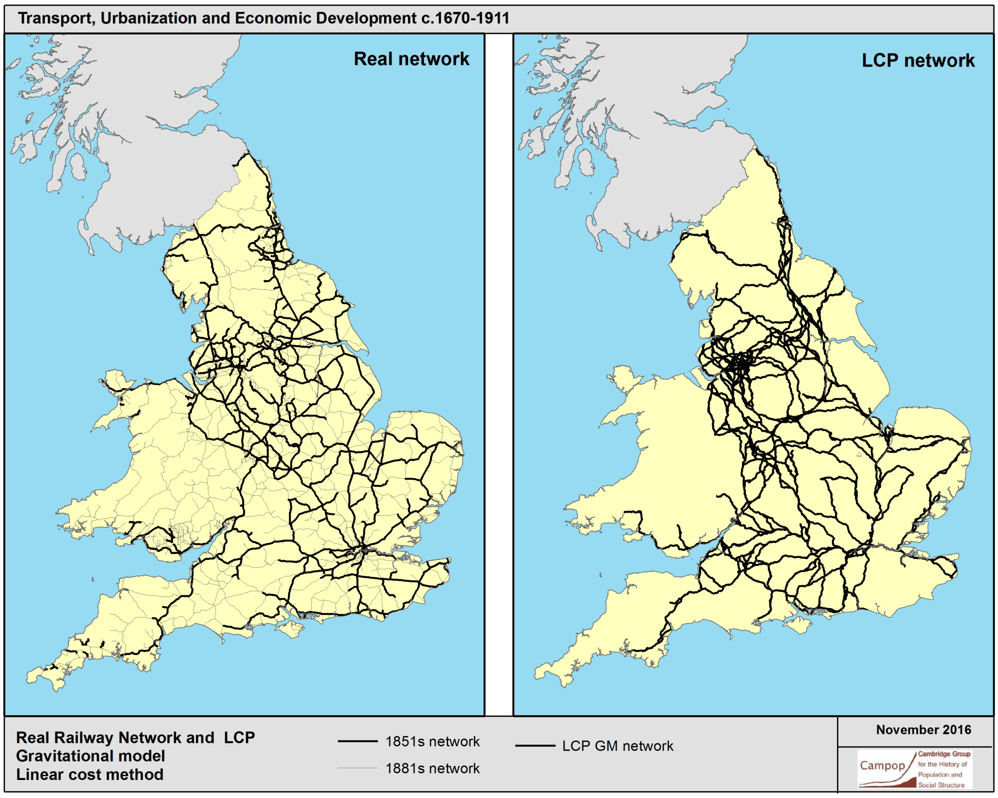
[9] The inconsequential places approach has been used in other papers see Faber (2014) and Lipscombe et. al. (2013).

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 experimented with several approaches to modeling the cost of elevation (see the appendix for more details, EDUARD please supply). Our baseline model builds on a 19th century engineer Wellington (1877) who estimated the relationship between construction costs and elevation slope. Starting with a normalization of construction costs per km at zero slope to be 1, Wellington argued that construction costs per km increased by 2.96 for every 1% increase in slope. We use Wellington's formula to help identify the least cost path connecting our town pairs. Specifically, we used the ESRI least-cost-path python schema in order to run the spatial analysis using an 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 4. The left hand panel of figure 4 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. Below we use the distance between a unit's location and the LCP as the instrument for distance to railway stations. Stations are so numerous by 1851 that distance to railways and distance to stations are similar.

3.1 Employment growth model

Aside from identifying the effects of population growth we are also interested in whether railways attracted certain types of employment. One hypothesis is that employment growth will decrease with railway station distance in labor intensive employment categories, and increase with distance in land intensive employment categories. This prediction is consistent with spatial models with two differentiated goods sectors, and with differing labor production

Figure 4: Rail network and least cost path network



intensities.[10] We test for the effects of distance to railway stations on the five employment categories: (1) secondary, (2) tertiary, (3) agriculture, (4) extractive (mining, fishing, and forestry) and (5) labourer. The last category, labourer, is sector unspecific. It also includes cases where no occupation is stated or the occupational status is uncertain. Our baseline estimating equation is:

$$y^k_{i1881} - y^k_{i1851} = -\lambda^k_1 y^k_{i1851} + \lambda_2 y_{i1851} + \beta^k_1 r_{i1851} + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{3}$$

[10]Fujita et. al. (2001), Rossi-Hansberg (2005), Desmet and Rossi-Hansberg (2009, 2014) write down spatial models which explicitly deal with two differentiated goods sectors, like agriculture and manufacturing, or manufacturing and services. 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.

where y_{it}^k is the log of employment density in category k in period t and as before y_{it} is the log of population density in period t. Our main coefficient of interest is β_1^k which measures the effect of log distance to 1851 railway stations in employment category k . Note the two controls for 1851 density are likely to have different effects. The expectation is that population density in 1851 y_{i1851} will be positively associated with employment growth in category k from 1851 to 1881 if there are 'urbanization' economies. By contrast, own category employment density y_{i1851}^k is predicted to be negatively associated with employment growth in category k due to technology diffusion and congestion forces, which promote dispersion of employment.

As with the population models, the employment growth equation (3) is extended to address various concerns. First, we include population growth from 1821 to 1851 to capture general pre-trends. Second, endogeneity remains a concern and so we again instrument distance to railway stations with distance to the LCP. Third, a 3rd degree polynomial in the log of own category 1851 employment density y_{i1851}^k is included. Drawing on Desmet and Rossi-Hansberg (2009), it is possible that knowledge spillovers are strong enough to create a non-monotonic relationship between a location's initial employment and its subsequent employment growth.[11] Desmet and Rossi-Hansberg document this pattern for services in the late twentieth century and manufacturing in the early twentieth century. They find that U.S. cities with lower initial employment generally grow faster than cities with higher initial employment density, but they also document a reversal in this pattern in the middle of the distribution over initial employment. Below we document a similar pattern in our setting.

[11] Their model relies on three forces: (1) technology diffusion which leads to dispersion, (2) knowledge spillovers which favor concentration, and (3) land congestion which favors dispersion. 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 grow the least. At the medium range of initial employment knowledge spillovers dominate and locations with more initial employment grow more than areas with low initial employment.

3.2 Extenstions allowing heterogeneity

One extension of our baseline model tests whether railways had different effects in parish units that were more dense in population or employment in 1851. This extension is of interest for policy discussions as the distributional effects are transport improvements are often debated. It also has theoretical interest because railways may magnify the effects of knowledge spillovers which potentially increase growth for medium density areas. The following equation is estimated:

$$y^k_{i1881} - y^k_{i1851} = \lambda_2 y_{i1851} + \sum_{j=1}^{3} \lambda^k_j \cdot (y^k_{i1851})^j + \sum_{j=1}^{3} \eta^k_j \cdot (y^k_{i1851})^j \cdot r_{i1851} + \beta_1 r_{i1851} + \beta_2 x_i + \alpha_j + \varepsilon_i \quad (4)$$

where the new terms are the 3rd-degree polynomial in the log of own category employment y^k_{i1851} and a series of interaction terms with distance to railway stations r_{i1851}. The marginal effect of increasing distance to rail stations is given by $\beta_1 + \eta^k_1 y^k_{it} + \eta^k_2 (y^k_{it})^2 + \eta^k_3 (y^k_{it})^3$. We report this value for different levels of 1851 population and employment density and compare with a baseline model that includes no interactions.

A second extension considers whether railways had different effects depending on endowments of coal. Transportation of coal was a problem historically because it has a low value to weight ratio, and thus it did not pay to ship it over long distances (Wrigley 2010). Thus parish units with coal deposits had an energy cost advantage, and were often more likely to adopt early steam engines (Nuvolari et. al. 2011). It is an open question whether the advantages of having coal were enhanced by railways or other transport improvements. We examine this issue by studying interactions between log station distance and an indicator for parishes having exposed coal, which proxies for favorable coal endowments. The specification is otherwise the same as the baseline models (2) and (3).

4 Data

4.1 Population and employment data

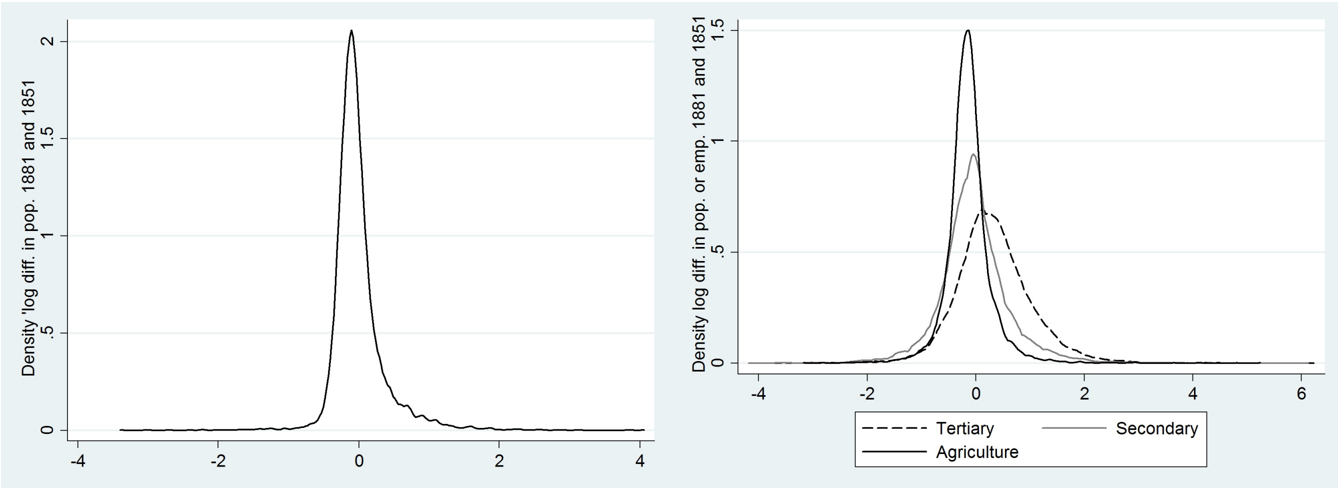
Population data from the census are available in every decade starting in 1801. They have been digitized at the parish level by the Cambridge Group for the History of Population and Social Structure. Detailed occupational data from the census are available at the parish level for all males in 1851 and 1881. Occupations are classified into 5 broad categories noted above. There is a further classifications which includes 38 specific secondary occupations and 25 specific tertiary occupations.[12] We identify 14 rapidly growing occupations within these for special analysis.

The population data from 1801 to 1881 and the male occupational data in 1851 and 1881 are linked to GIS using a consistent set of parish boundaries. Some parishes were subdivided after 1801 and so it is necessary to assemble larger units to maintain a consistent boundary. In total we have 9489 parish units across time. The distribution of the log difference in population from 1851 to 1881 across units is plotted in the left hand side of figure 5. The median parish experienced negative population growth with a log difference of -0.073. Other parishes experienced significant population gains. At the top or 90th percentile the log difference is 0.377. Considering that English population increased from 17.0 million in 1851 to 22.3 million in 1881 (a log difference of 0.27) it is clear that population was concentrating in smaller numbers of units.

The distributions of the log difference in employment for 3 main categories are shown in the right hand panel of figure 5. Agricultural employment declined for most parishes, and the rate of decline was broadly similar across parishes as indicated by the tight distribution. Secondary employment also declined for most parishes, but there were some that experienced high secondary employment growth. The log difference in secondary employment is 0.69 at the 90th percentile, which implies a 100% growth in secondary employment. Tertiary

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

Figure 5: Distribution of population and employment growth across parish units 1881 and 1851



Sources: see text.

employment grew in the median parish, but there was a wide dispersion. At the 90th percentile the log difference was 1.19 which implies a 228% growth.

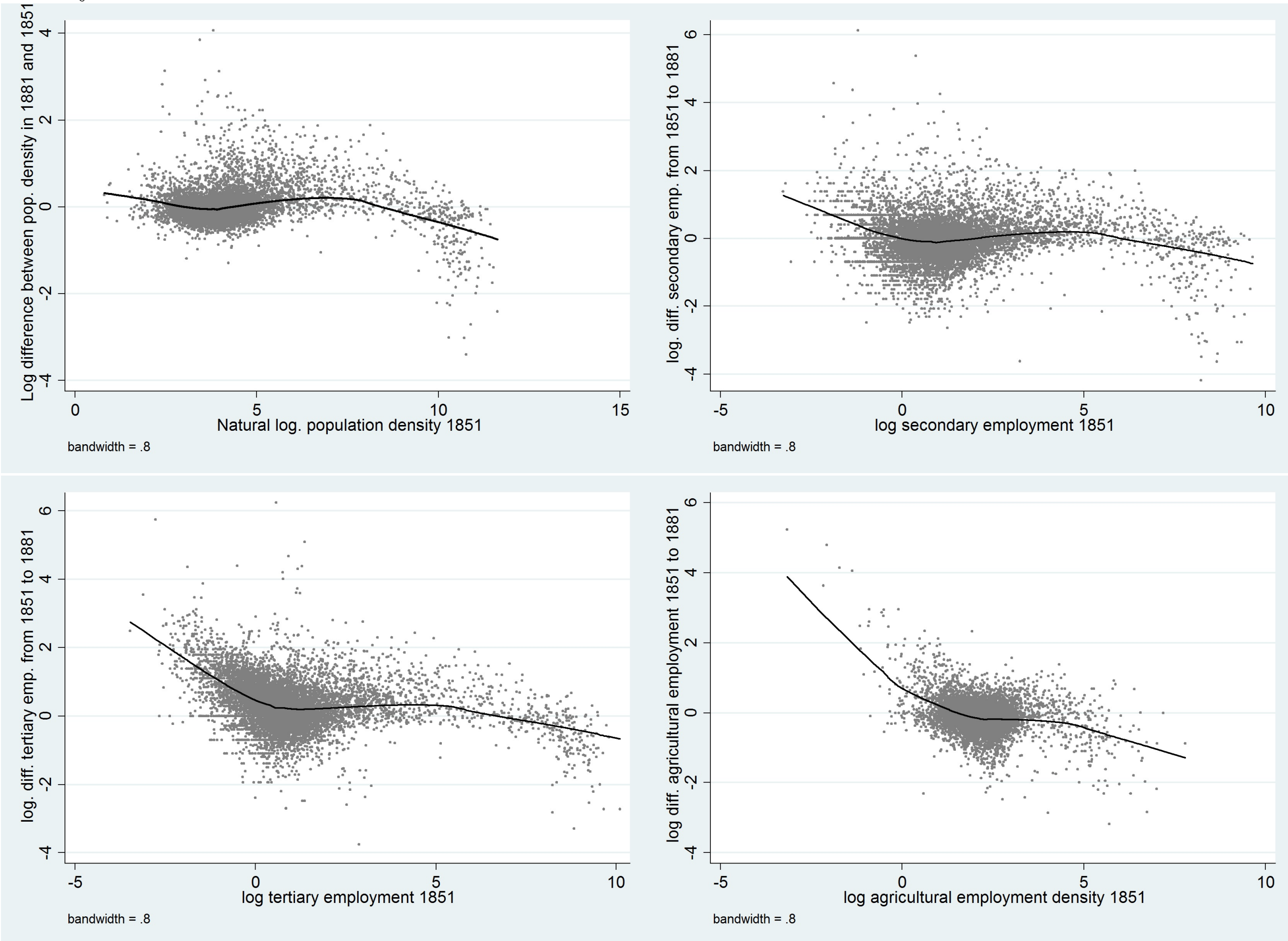
It is also useful to document the correlation between population or employment growth from 1851 to 1881 and their respective levels in 1851. Figure 6 plots these data and the locally weighted scatter plot smoothing curve (LOWESS). In the middle of the distribution for 1851 population there is a reversal where the log differences in populaton tend to rise with higher 1851 population density. The same pattern is evident for secondary and tertiary growth in the middle of the distribution for 1851 own employment density. As expected the same pattern is largely absent in agriculture where knowledge spillovers are weaker.

4.2 Data on transport networks and parish unit centers

The paper uses new GIS data on transport networks in England and Wales.[13] These include all ports c.1840, ports with steamship services c.1840, inland waterways c.1830, and turnpike roads c.1830. Most importantly the data include the location of railway lines and stations

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

Figure 6: Plots of population and employment growth on 1851 population and employment density



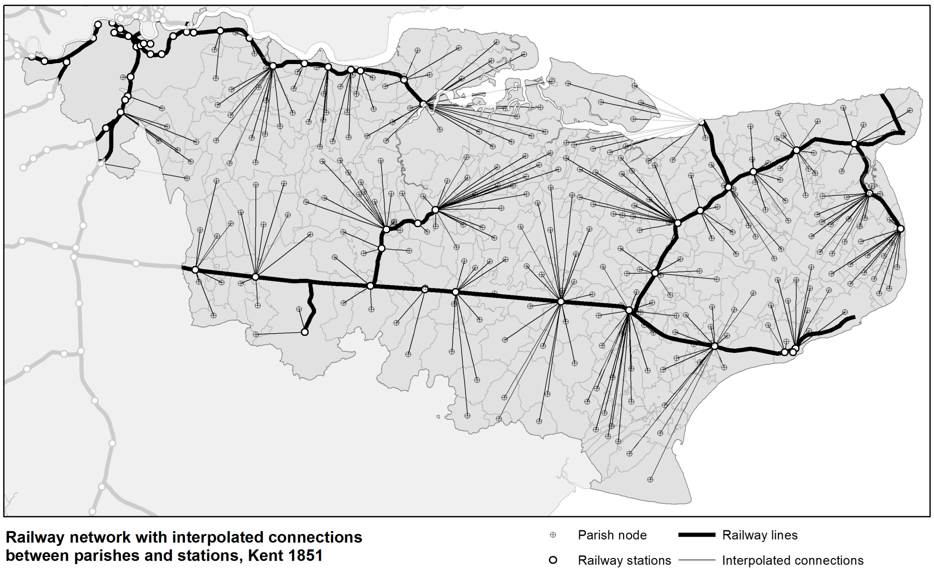
Sources: see text.

with their opening and closing dates starting in the early nineteenth century.[14] Here we focus on the railway network of 1851 which is shown in figure 3 above.

Our empirical analysis studies the effects of distance between parishes and transport networks. For this purpose, it is necessary to define a parish unit center. Ideally the center would represent the main location of production and consumption in a parish. In many cases, the market place is ideal for this purpose. Therefore, if the parish had a market

[14]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.

Figure 7: Parish centers and distance to 1851 railway stations in Kent



Sources: see text.

town at some point between 1600 and 1850 then the market town is taken as the center. This applies to 746 of the 9489 units (see appendix for details. Xuesheng please include explanation of market towns in appendix describing continuous units). Parishes that had no market town were likely to be rural and the geographic centroid is taken as the parish center. It should be noted that little error is introduced by using the centroid since the average parish unit was 15 square km. Figure 7 focuses on railways and stations in Kent to the southeast of London. It shows parish unit boundaries, centers (open circles), railway lines (black), and stations (white circles). There is clearly much variation in the distance between parish centers and 1851 railway stations.

4.3 Data on endowments and summary statistics

The endowment data include shares of land in 11 soil types and indicators for being on exposed coal fields, and being on the coast. Soil categories are from Avery (1980) and Clayden and Hollis (1985) and were digitized by the National Soils Map of England and Wales.[15] The share in each category except one is included as a variable. Exposed coalfields are those where coal bearing strata are not concealed by rocks laid down during the Carboniferous Period.[16] Coastal units are identified using shapefiles for parish boundaries in England and Wales.

The endowment data also include ruggedness measures like the average elevation, the average elevation slope in the parish, and the standard deviation in the elevation slope in the parish. The appendix provides a description of these variables (Eduard please provide an appendix on methods for calculating average and standard deviation of slopes, and sources for elevation). A final control variable of note is the distance to the nearest large town in 1801 (Birmingham, Bristol, Leeds, Liverpool, London, Manchester, Newcastle, Plymouth, Portsmouth, Sheffield). It captures growth effects from having nearby urban centers.

Table 1 reports summary statistics for the main variables. They are labeled as railways variables (1), population and occupation dependent variables (2), and the controls (3)-(6) described above. Note the 54 county and 616 registration district dummies are omitted.

5 Results

We begin by analyzing the effects of log distance to railway stations in 1851 on the log difference in population in 1881 and 1851. As a preliminary calculation, the the correlation coefficient between the two is -0.203 (p-value 0.00). Table 2 reports results for various regression specifications. Coefficients for geography and pre-railway infrastructure are not reported but will be described later. Column (1) is the most parsimonious and includes

[15]The 11 soil categories are (1) Terrestrial raw, (2) Raw gley, (3) Lithomorphic, (4) Pelosols, (5) Brown, (6) Podzolic, (7) Surface-water gley, (8), Ground-water gley, (9) Man made, (10) peat soils, and (11) other. See http://www.landis.org.uk/downloads/classification.cfm#Clayden_and_Hollis.

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

the log of 1851 population density, geography and pre-railway transport infrastructure, and county fixed effects. It shows a significant negative effect of log distance to railway stations on population growth. The coefficient implies that a 10% increase in station distance reduces population growth from 1851 to 1881 by 0.42 percentage points. A complete discussion of the magnitudes is provided later in the paper. Column (2) adds the log difference in population growth from 1821 to 1851 to address concerns about pre-trends. Population growth from 1821 to 1851 is found to be positively related to population growth from 1851 to 1881 as expected. Including this variable diminishes the effect of station distance, but it remains significant. The specification in column (3) replaces county fixed effects with registration fixed effects. The coefficient on station distance becomes more negative, implying that the narrower control for unobserved heterogeneity at the local level increases the magnitude of the station distance coefficient. Column (4) adds the two higher order polynomial terms for log of 1851 population density. They raise the R-square, but the estimated effect of station distance changes little. Column (5) reports the IV results for the same specification as in (4). The IV estimate is larger than OLS by a substantial amount. The larger magnitude is not due to a weak instruments problem. The Kleibergen-Paap Wald F statistic indicates that the first stage is very strong. Columns (6) and (7) repeat the OLS and IV estimates dropping units with towns used in the construction of the LCP. The IV estimates get larger in magnitude but the overall conclusion is the same.

One explanation for the difference between OLS and IV 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. In any case, there appears to be a downward bias in the OLS coefficient. A similar conclusion will be found in other specifications below.

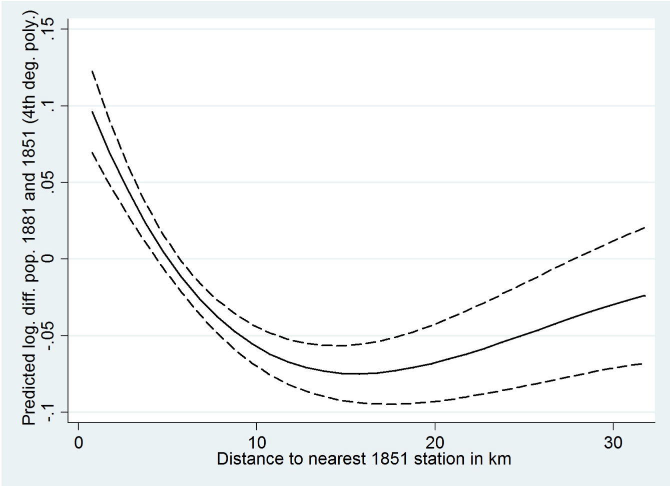
Railway access can be defined in other ways than distance to nearest 1851 station. One alternative is the average distance to the first and second nearest station in 1851. This approach addresses whether it was important to have more than one nearby station. Results using the log of the average distance to the first and second nearest station in 1851 are presented in table 3. They are qualitatively similar in all specifications, although the magnitudes differ.

Another alternative specification replaces log distance to stations with a 4th degree polynomial in station distance. The advantage here is that the marginal effects can be different as station distance increases. The left hand panel of figure 8 plots the predicted log difference in population between 1881 and 1851 using the model with 4th degree polynomial. The prediction is made for distances up to 32 km which represents the 95th percentile. The model predicts a 15 percentage point decline in population growth as distance increases from 0.75 km to 10 km, which is approximately a one standard deviation change in distance. Beyond 15 km the polynomial model predicts a positive relationship between growth and station distance but the wide confidence interval implies that the reversal in effects is not statistically significant. For comprison, the log linear model analyzed earlier predicts a 14 percent decline in population from 0.75km to 15km. The 4th degree polynomial produces a similar estimate of population change as the log linear model over distances less than 15km, so in many comparisons the two models agree. For the next sections we continue to use the log linear model because of its ease of interpretation.

5.1 Employment growth results

The results for male employment growth are summarized in table 4. Column (1) analyzes the effects on the log difference in total male employment in 1881 and 1851 using OLS. The specification includes the full set of controls. Besides the log distance to the nearest 1851 station, it has a 3rd degree polynomial in log 1851 male employment, the log of 1851

Figure 8: Effects on population and secondary growth using 4th degree polynomial for distance to stations



Sources: see text.

population density, the log difference in population in 1851 and 1821, registration district fixed effects, geographic controls, and pre-railway infrastructure controls. The results show a negative and significant effect of station distance on male employment growth. Column (2) reports similar results using IV. The distance to station coefficient becomes more negative and remains significant, albeit with less precision. These two findings are nearly identical to the similar model for population growth (columns 4 and 5 in table 2), which is reassuring because male employment and population growth should respond similarly to station distance.

The advantage of studying employment is that we can test which types of occupations were attracted to locations near railway stations. Column (3) in table 4 examines the effects on the log difference in secondary employment using OLS and column (4) does the same using IV. In both models a 3rd degree polynomial in 1851 secondary employment is included. The estimates are of similar magnitude to total male employment, although the standard error is larger in IV and makes the coefficient statistically insignificant. The most that we can say is that distance to railway stations had a negative, but imprecise effect on total secondary employment.

Columns (5) and (6) in table 4 repeat the same specifications for the log difference in tertiary employment. The coefficients are negative and significant in OLS and IV and are larger in magnitude than for total male employment. It appears that tertiary occupations were especially attracted to railway stations. Columns (7) and (8) show results for the log difference in agricultural employment. In OLS the coefficient on log station distance is positive and significant. In the IV the coefficient is larger although less precise. It is clear though that agricultural employment growth exhibits a very different relationship to railway access compared to secondary and tertiary employment growth.

The final two employment categories are extractive and labourer. Results for these are shown in columns (9) to (12). Labourer employment growth is broadly similar to secondary and tertiary although again the standard errors are larger in the IV. The effect of railways is less consistent on extractive employment growth with the sign of the IV coefficient changing. Extractive is arguably a unique sector because its location depends greatly on endowments like coal, and it was heavily dependent on transport. It is perhaps less straightforward to think of an inconsequential place getting a railway station and also having coal to mine.

Secondary and tertiary are the most diverse of the 5 employment categories and so it is useful to examine their sub-categories. Rather than review all we chose to focus on secondary and tertiary employment categories that were growing more rapidly than the average from 1851 to 1881. Our list of rapidly growing 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.

Table 5 reports the estimates for new and rapidly growing secondary categories. The specification is the same used to analyze the log difference in secondary employment in 1881 and 1851 (see table 4). The difference is that the controls include a 3rd degree polynomial in log 1851 own category employment, not a polynomial in log 1851 secondary employment. The results show that the log difference in employment is negatively related to station distance for 9 out of the 10 sub-categories. However, the coefficients are statistically significant in only two of the sub-categories, iron and steel and machine tools. One reason may be the small sample size for some of these new secondary categories. In column (11) we combine all males occupied in high growth secondary occupations into one sub-group. The effect of distance to stations has a significantly negative sign in the combined group. The same finding is reached using IV as shown in column (12).

Our list of rapidly growing tertiary occupations includes media, financial services, commercial and administrative services, and railway transport services. 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. Table 6 reports the estimates for rapidly growth tertiary categories. There is a negative and significant effect on the growth of railway transport services, which makes sense given that railway stations required employees to operate. More notable is the negative and significant sign on employment growth in financial services, and in commerce and administration. Railways also contributed to non-rail tertiary employment growth.

5.2 Heterogeneous effects based on initial density

In this section, we extend our baseline model to examine how the effects of station distance differ depending on a parish unit's 1851 population or employment levels. Estimates are reported for equation (4) with interactions between log distance to stations and the 3rd degree polynomial in either log 1851 population density or log 1851 own-category employment density. Estimates are also reported for a baseline model that includes the 3rd degree polynomial but does not have interactions with log station distance. We call the former model 1 and the latter model 2.

The results are summarized in table 7. The first panel reports the predicted change in population for a one standard deviation increase in log station distance at the 5th percentile of 1851 population density up to the the 95th percentile. The predicted values in italics are statistically significant at the 95% confidence level. Models 1 and 2 produce somewhat different estimates as indicated by the last column showing the ratio of the predicted effects. In model 1, the largest effect of log station distance on population growth in terms of magnitude is -0.0731 log points, which applies to the 65th percentile of 1851 population density. Smaller, but still similar effects are found at 25th percentile and 85th percentile. Notice also that the predicted effects of greater distance to stations has a negative and significant sign as low as the 15th percentile of 1851 population density. This finding suggests that many low density parishes still gained population from being near railways.

There is more heterogeneity in the effects of station distance on secondary employment growth (see panel B, table 7). The largest effect in magnitude in -0.099 log points, which applies to the 45th percentile of 1851 secondary employment density. By comparison the effect is -0.0826 log points at the 75th percentile. Below the 45th percentile the estimated effects are similar but are not stasticially different from zero. Thus most parishes with low 1851 secondary employment density did not clearly gain secondary employment from being near railways stations.

The effects of railways on tertiary employment growth showed less heterogeneity with respect to initial density (see panel C, table 7). The largest effects are found at the 45th percentile of 1851 tertiary employment, but they are very similar to the effects at the 15th and 85th percentiles.

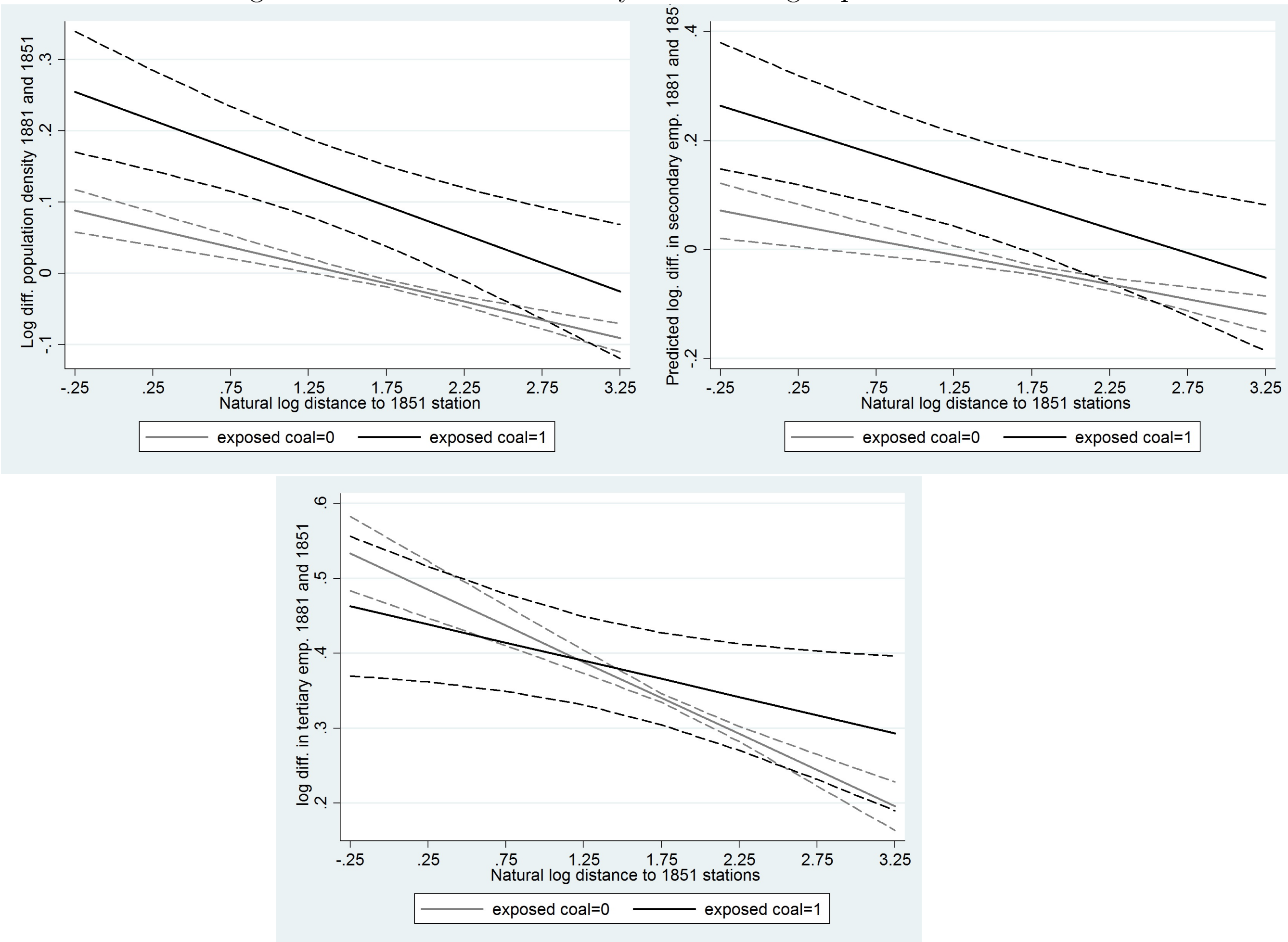
5.3 Heterogeneous effects based on coal

The baseline results are extended further in this section to examine whether railways affected parish units with coal differently. The extended model including interactions between log station distance and the indicator for exposed coal. For simplicity the specification is based on equation (5) which omits the interactions between station distance and 1851 population or employment density. The results are shown graphically in figure 9 for population growth, and secondary and tertiary employment growth. They provide several insights. First, having coal endowments significantly increases population and secondary employment growth. On average population growth was 12 percentage points higher in parishes with coal, and secondary employment growth was 13 percentage points higher in parishes with coal. By contast, teritary employment growth was not any higher in parishes with coal. This makes sense because coal was not an important input in the production of services. Second, the estimates some some heterogeneity in the effects of railways and coal. For example, the effect of increasing station distance on population growth was one-half larger if parish units had a coal. The effects on secondary growth are two-thirds larger if a parish unit had coal. For tertiary there is no significant heterogeneity, but it is notable that the ordering of the slopes is different. The magnitude of the effect of railway station distance is smaller if a parish had coal.

5.4 Summary

What do our estimates imply about the effects of railways on population and employment growth in England and Wales? There is evidence that population growth was lower for a parish if it was more distant to a railway station. Our IV estimates in tables 2 and 4 imply that a one standard deviation increase in log distance to the nearest 1851 railway station would lower population growth by 13 percentage points. Larger effects are implied by the model using the log of average distance to the first and second nearest 1851 railway station. Here the IV estimates would imply that a one standard deviation increase in log average distance would reduce population growth by 17 percentage points. One interpretation of these results is that railways increased market access, which made parishes near railway

Figure 9: The effects of railways and having exposed coal



Sources: see text.

stations more desirable for workers and firms to locate.

Parishes near railway stations had higher employment growth as they increased population, but the growth was uneven across employment categories. Proximity to railways increased secondary employment growth, but the estimated effects vary. Railways had their clearest effects in rapidly growing secondary sectors, like iron and steel and machine tools. Railways also increased secondary growth more in parishes with coal and in parishes with medium 1851 secondary employment density (e.g. the 45th to the 75th percentile). The connection between railways and higher growth in medium density areas is telling of the mechanisms at work. They suggest that railways enhanced or reinforced knowledge spillovers, the main drivers of higher growth in medium density areas according Desmet and Rossi-Hansberg's model. The precise mechanisms require more research, but one theory is that railways facilitated information flows which encourage knowledge spillovers. A complementary argument is that units with the lowest initial employment density grow less if they are close to railway stations because they face more competition with medium and large employment areas where productivity is higher.

Railways had large and more uniform effects on tertiary employment growth. The IV estimates imply that a one standard deviation increase in log distance to the nearest 1851 railway station would lower tertiary employment growth by 13 percent. Proximity to railways also led to more growth in the rapidly growing railway service sector, financial services, and in commerce and administration. The last two service occupations are important because they contributed to growth in England and Wales well into the 20th century.

Finally, railways had very different effects on the spatial distribution of agricultural employment. Parishes near railway stations maintained or most likely lost employment in agriculture. A one standard deviation increase in distance to stations is predicted to raise agricultural growth by 6 percentage points. As railways attracted labor intensive industries it likely became more difficult for agriculture to remain competitive in the land market near stations. Many agricultural jobs were eliminated and many that remained shifted to parishes more distant from stations.

5.5 Comparisons

How do the effects of railways compare with transport improvements in other contexts? To address this question we compare our estimates with similar models 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. Our IV estimate of the elasticity of population density with respect to log distance to stations is -0.119. The comparison is not exact, but it does suggest that access to transport is a more significant driver of growth in modern-day Africa than 19th century England and Wales.

Another comparison is with modern-day China. Banerjee, Duflo, and Qian (2012) estimate the elasticity of annualized GDP growth in Chinese localities with respect to distance to railways in China. For 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. We find a larger effect of railways on population and tertiary employment growth and a smaller effect on secondary employment growth. Perhaps one reason is that England had already started its structural transformation in secondary employment prior to railways, but it was still in the midst of its early transition to tertiary. In China the transition to secondary occurred at the same time as the large increase in railways.

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 twenty 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 average distance to first and second nearest stations implies a 1.8% increase in population growth over the next 30 years. Thus density of railway stations in nineteenth century England and Wales did not generate as much growth as the density of highways in the late twentieth century US, but the differences are not large.

It is also revealing to compare the effects of railways with the transport networks pre-ceding them in history. With the exception of coastal shipping, railways were signficantly more cost efficient than earlier forms of transport on roads and inland waterways. However, earlier transport networks, turnpike roads, inland waterways, and ports, may have had persistent effects into the nineteenth century, even after their transport modes became outdated. Thus it is not obvious that railways had a larger effect on mid-nineteenth century growth than other transport networks. Also interesting is the comparison with ports serving steamships, as they were a recent innovation like railways. We measure the relative effects by calculating standardized coefficients for the distance to each transport network. For example, the standardized coefficient for the railways variable is its coefficient multiplied by the standard deviation of log distance to stations divided by the standard deviation in the log difference of population or employment. The comparison is informative because the distributions of the transport variable appear broadly similar, albeit with differences in means and variances.

Table 8 reports the results for specifications using the baseline model with a 3rd degree polynomial in log 1851 population density or log 1851 own category employment density. In other words, there are no interactions for heterogeneity in transport variables. In general, log distance to railway stations has the first or second largest standardized coefficient and is the most consistent factor in explaining growth. Log distance to ports with steamships also has a large effect on growth, and emphasizes the importance of steampower for mid-nineteenth century growth. However, it is also surprising that distance to earlier transport networks have similar or larger effects in some cases. For example, distance to turnpike roads (in logs) is nearly as important as distance to railways in explaining secondary employment growth. Distance to inland waterways is nearly as important as distance to railways in explaining agricultural employment growth. Finally, the distance to general ports has a larger effect than railways in explaining extractive employment growth. It would appear that earlier forms of transport had persistent effects, perhaps by creating employment clusters which lead to subsequent growth in the nineteenth century.

5.6 Counterfactual with 1841 network

As a final exercise we consider how growth would have been different if history unfolded differently. Railways were an experimental technology in the 1820s and 30s. It was only around 1840 that railways were proven to be cost effective substitutes for horse drawn wagons and canal boats (Dyos and Aldcroft 1969). Instead suppose that 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 have 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.[17] In this case, we think it is appropriate to use a model that flexibily captures the effects of railways. Our model for predicting population growth comes from equation (6) below.

$$y_{i1881} - y_{i1851} = \rho(y_{i1851} - y_{i1821}) + \textstyle\sum_{j=1}^{3} \lambda_j^k \cdot (y_{i1851})^j$$

$$\sum_{j=1}^{3} \eta_j \cdot (y_{i1851})^j \cdot r12_i + \beta_1 r12_i + \nu \cdot r12_i \cdot coal_i + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{5}$$

where $r12_i$ is the log average distance to the first and second nearest 1851 station. The specification also includes the log difference in population in 1851 and 1821, registration district fixed effects α_j, the 3rd degree polynomial in 1851 log population density, and it allows for interactions between station distance $r12_i$ and the polynomial for 1851 log population density and between station distance and having exposed coal.

Equation (6) below is used to predict employment growth in category k. It is similar to (5) but it includes interactions between $r12_i$ and the polynomial for log 1851 own category

[17] 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 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 calculation 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.

employment.

$$y_{i1881}^k - y_{i1851}^k = \rho(y_{i1851} - y_{i1821}) + \lambda_2 y_{i1851} + \textstyle\sum_{j=1}^3 \lambda_j^k \cdot (y_{i1851}^k)^j +$$

$$\sum_{j=1}^{3} \eta_j^k \cdot (y_{i1851}^k)^j \cdot r12_i + \beta_1 r12_i + \nu \cdot r12_i \cdot coal_i + \beta_2 x_i + \alpha_j + \varepsilon_i \tag{6}$$

The results are summarized in table 9. With the 1851 network, population is predicted to grow by 40.3 percentage points, and with the 1841 network, population is predicted to growth by 31.8 percentage points. The difference in population growth is minus 8.5 percentage points. Turning to employment, we find that secondary employment is predicted to grow 8.9 percentage points less, tertiary employment 4.0 percentage points less, and agriculture 1.5 percentage points more with the 1841 network. Based on these estimates population and employment growth would have continued if railway technology proved non-economic in the 1830s. However, their growth would have been significantly lower, indicating that railways were a key driver of development in mid-nineteenth century England and Wales.[18]

6 Conclusion

How do transport improvements affect the growth and spatial structure of employment? This paper answers these questions in the context of nineteenth century England and Wales where urbanization increased and secondary and tertiary employment rose. To address endogeneity of railway placement, we use an instrument that identifies locations close to railways because they were on a least cost path minimizing elevation changes and distance

[18] Hawke (1970) estimated that without railways freight and passengers would have faced higher transport costs resulting in approximately a 6 percent loss in GDP by 1865. Our estimates are more in line with Leunig (2006) who revises Hawke's estimates for time and money savings to railway passengers and finds them to be around 7 percent of GDP by 1880. Note that we arrive at our estimate from a very different methodology by using parishes more distant from railway stations as our counter-factual for parishes close to stations.

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 10 and 13 percentage points over the next 30 years. Distance to railway stations had a different and positive effect on agricultural employment growth. In further analysis the largest effects of station distance on population and employment growth are found at medium 1851 density units and in units with coal. Our estimates also show that proximity to railways had a moderately larger effect on growth than proximity to other forms of transport. Lastly, we find that population growth in England and Wales between 1851 and 1881 would have been around 8 to 9 percent lower if the railway network did not expand in the 1840s. This accounts for about one-fifth of the total growth in population.

Our findings go beyond the implications for growth in nineteenth century England and Wales. They contribute to a better understanding of the mechanisms by which transport improvements affect growth. Our finding that railways had the largest effect on secondary employment growth in medium density areas suggests that railways enhanced knowledge spillovers, which are a key factor in many urban and trade models. There are also policy implications from the findings that medium density areas had the largest growth effects from railways. Transport improvements are likely to have a more limited growth effect in lower density areas. If the aim of policy makers is to enhance growth in areas with less population or employment. then perhaps something other than transport should be considered. There is one caveat to this conclusion. Tertiary employment grew more evenly across the initial density distribution in response to railways. Transport may still be an effective tool to raise tertiary employment in areas that have less tertiary employment to begin with.

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. Avery, Brian William. Soil classification for England and Wiles: higher categories. No. 631.44 A87. 1980.

3. 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.

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

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

6. Berger, Thor, and Kerstin Enflo. "Locomotives of local growth: The short-and long-term impact of railroads in Sweden." Journal of Urban Economics (2015).

7. 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.

8. Campbell, Gareth, and John D. Turner. "Dispelling the Myth of the Naive Investor during the British Railway Mania, 1845–1846." Business History Review 86.01 (2012): 3-41.

9. Campbell, Gareth, and John D. Turner. "Managerial failure in mid-Victorian Britain?: Corporate expansion during a promotion boom." Business History 57.8 (2015): 1248-1276.

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

11. 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.

12. Clayden, Benjamin, and John Marcus Hollis. Criteria for differentiating soil series. No. Tech Monograph 17. 1985.

13. Crafts, Nicholas. "Productivity growth in the industrial revolution: A new growth accounting perspective." The Journal of Economic History 64.02 (2004): 521-535.

14. 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.

15. 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.

16. 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.

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

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

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

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

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

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

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

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

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

26. 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.

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

28. 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.

29. Gregory, Ian N., and Jordi Martí Henneberg. "The railways, urbanization, and local demography in England and Wales, 1825–1911." Social Science History 34.2 (2010): 199-228.

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

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

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

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

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

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

36. 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).

37. 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.

38. Nuvolari, Alessandro, Bart Verspagen, and Nick Von Tunzelmann. "The early diffusion of the steam engine in Britain, 1700–1800: a reappraisal." Cliometrica 5.3 (2011): 291-321.

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

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

41. 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.

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

43. Tang, John P. "Railroad expansion and industrialization: evidence from Meiji Japan." The Journal of Economic History 74.03 (2014): 863-886.

44. 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.

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

Table 1. Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
(1) railway vars.					
Distance to 1851 rail station in km	9489	10.4564	11.0657	0.0215	73.1296
Log distance to 1851 rail station in km	9489	1.8401	1.0979	-3.8407	4.2922
Distance to GM LCP km (IV)	9489	11.8619	16.5488	0.0001	116.3862
(2) Population and occupation dependent vars.					
Ln diff Pop. Density 1851 to 1881	9489	-0.0087	0.3765	-3.4044	4.0585
Ln diff secondary emp. 1851 to 1881	9061	-0.03	0.6569	-4.1897	6.1203
Ln diff tertiary emp. 1851 to 1881	9321	0.3328	0.7035	-3.7612	6.2385
Ln diff agriculture emp. 1851 to 1881	9403	-0.1382	0.4228	-3.1781	5.2364
Ln diff extractive emp. 1851 to 1881	3385	0.2752	1.1165	-3.912	7.5374
Ln diff labourer emp. 1851 to 1881	8231	0.4024	1.1718	-3.8067	5.247
(3) 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
(4) Pre-trend controls					
Ln diff Pop. Density 1821 to 1851	9485	0.1695	0.2690	-1.0492	4.7950
(5) 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
Perc. of land with Raw gley soil	9489	0.0847	1.3279	0	76.4964
Perc. of land with Lithomorphic soil	9489	8.6151	19.8301	0	100
Perc. of land with Pelosols soil	9489	8.2038	20.6374	0	100
Perc. of land with Brown soil	9489	41.5641	33.1188	0	100
Perc. of land with Podzolic soil	9489	4.6249	14.3262	0	99.5655
Perc. of land with Surface-water gley soil	9489	24.6329	29.4604	0	100
Perc. of land with Ground-water gley soil	9489	10.1871	20.1177	0	100
Perc. of land with Man made soil	9489	0.36384	3.2621	0	94.9904
Perc. of land with Peat soil	9489	1.1875	5.2798	0	91.4403
Perc. of other soil	9489	0.5354	1.9668	0	65.1538
(6) 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: Effect of distance to nearest railway station on log difference population density 1881 and 1851

	(1) OLS coeff. (p-value)	(2) OLS coeff. (p-value)	(3) OLS coeff. (p-value)	(4) OLS coeff. (p-value)	(5) IV coeff. (p-value)	(6) OLS coeff. (p-value)	(7) IV coeff. (p-value)
Log dist. to nearest 1851 rail station km	-0.0419	-0.0377	-0.0652	-0.0548	-0.1191	-0.0573	-0.1597
	(0.000)	(0.000)	(0.000)	(0.000)	(0.013)	(0.005)	(0.003)
Log pop. density 1851	0.0041	-0.0169	-0.0222	-0.9459	-0.8903	-0.9617	-0.8797
	(0.548)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)
ΔLog pop. density 1821 to 51		0.2778	0.2094	0.1923	0.1794	0.1868	0.1659
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log pop. density 1851* Log pop. density 1851*				0.1752	0.1641	0.1772	0.1609
				(0.000)	(0.000)	(0.002)	(0.002)
Log pop. density 1851* Log pop. density 1851* Log pop. density 1851*				-0.0099	-0.0094	-0.0099	-0.0091
				(0.000)	(0.000)	(0.002)	(0.002)
County fixed effects	Y	Y	N	N	N	N	N
Registration district fixed effects	N	N	Y	Y	Y	Y	Y
Controls for geography and pre-railway transport	Y	Y	Y	Y	Y	Y	Y
Include units with town nodes in LCP	Y	Y	Y	Y	Y	N	N
Kleibergen-Paap rk Wald F statistic					58.271		49.042
R-square	0.3008	0.3308	0.4550	0.4818	0.4724	0.4816	0.4336
N	9489	9489	9489	9489	9489	9390	9390

Notes: Standard errors are clustered on county in specifications (1)-(2) and on registration district in (3)-(7). Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, distance to nearest large city and share of soil types. Pre-railway transport includes distance to nearest inland waterway, port, steamship port, and turnpike road. The instrument is distance to the LCP connecting large towns in 1801. see text for more details on instrument.

Table 3: Effect of average distance to first and second nearest railway station on log difference population density 1881 and 1851

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	IV	OLS	IV
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log av. dist. to first and second nearest 1851 rail station km	-0.0538	-0.0482	-0.0962	-0.0846	-0.1833	-0.0894	-0.2382
	(0.000)	(0.000)	(0.000)	(0.000)	(0.012)	(0.005)	(0.002)
Log pop. density 1851	0.0047	-0.0163	-0.0205	-0.9522	-0.9044	-0.9684	-0.9031
	(0.492)	(0.010)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
ΔLog pop. density 1821 to 51		0.2781	0.2088	0.1902	0.1750	0.1845	0.1612
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log pop. density 1851* Log pop. density 1851*				0.1770	0.1680	0.1790	0.1669
				(0.000)	(0.000)	(0.000)	(0.002)
Log pop. density 1851* Log pop. density 1851* Log pop. density 1851*				-0.0100	-0.0096	-0.0100	-0.0095
				(0.000)	(0.000)	(0.000)	(0.002)
County fixed effects	Y	Y	N	N	N	N	N
Registration district fixed effects	N	N	Y	Y	Y	Y	Y
Controls for geography and pre-railway transport	Y	Y	Y	Y	Y	Y	Y
Include units with town nodes in LCP	Y	Y	Y	Y	Y	N	N
Kleibergen-Paap rk Wald F statistic					50.421		45.223
R-square	0.3002	0.3303	0.4545	0.4818	0.4724	0.4821	0.4602
N	9489	9489	9489	9489	9489	9390	9390

Notes: Standard errors are clustered on county in specifications (1)-(2) and on registration district in (3)-(7). Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, distance to nearest large city and share of soil types. Pre-railway transport includes distance to nearest inland waterway, port, steamship port, and turnpike road. The instrument is distance to the LCP connecting large towns in 1801. see text for more details on instrument.

Table 4: Effect of average distance to nearest railway station on log difference male employment density 1881 and 1851

	All		Sec.		Tert.	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log dist. nearest 1851 rail station km	-0.0445	-0.0967	-0.0589	-0.0900	-0.0897	-0.1173
	(0.000)	(0.076)	(0.000)	(0.185)	(0.000)	(0.066)
3rd deg. poly. in log. 1851 own emp. den.	Y	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y	Y
Registration district FE	Y	Y	Y	Y	Y	Y
Geography and pre-rail transp. controls	Y	Y	Y	Y	Y	Y
Units with town nodes in LCP included	Y	Y	Y	Y	Y	Y
Kleibergen-Paap rk Wald F stat.		56.035		59.358		58.801
R-square	0.4386	0.4337	0.3983	0.3976	0.5199	0.5194
N	9488	9488	9061	9061	9321	9321
	Agric.		Extract		Labour	
	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	IV	OLS	IV	OLS	IV
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log dist. nearest 1851 rail station km	0.0152	0.0513	-0.0456	0.1978	-0.0990	-0.0773
	(0.019)	(0.133)	(0.110)	(0.211)	(0.000)	(0.377)
3rd deg. poly. in log. 1851 emp.	Y	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y	Y
Registration district FE	Y	Y	Y	Y	Y	Y
Geo. and transport controls	Y	Y	Y	Y	Y	Y
Units with town nodes in LCP	Y	Y	Y	Y	Y	Y
Kleibergen-Paap rk Wald F stat.		62.721		40.420		66.264
R-square	0.5613	0.5589	0.4765	0.4578	0.5687	0.5686
N	9403	9403	3385	3385	8231	8231

Notes: Standard errors are clustered on registration district in all specifications. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, distance to nearest large city and share of soil types. Pre-railway transport includes distance to nearest inland waterway, port, steamship port, and turnpike road. Own employment applies to male employment in columns (1) and (2), and to the employment category in all other columns. (The instrument is distance to the LCP connecting large towns in 1801. see text for more details.

Table 5: Effect of average distance to nearest railway station on log difference male employment density in rapidly growing secondary categories

	Printing	Glass making	Instrum. making	Chemical	Fuel	Iron & steel
	(1)	(2)	(3)	(4)	(5)	(6)
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log dist. nearest 1851 rail station km	-0.0373	-0.1523	-0.0374	-0.1298	-0.0982	-0.0727
	(0.422)	(0.502)	(0.238)	(0.165)	(0.158)	(0.000)
3rd deg. poly. in log. 1851 own emp. den.	Y	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y	Y
Registration District FE	Y	Y	Y	Y	Y	Y
Geography and pre-rail transp. controls	Y	Y	Y	Y	Y	Y
Units with town nodes in LCP included	Y	Y	Y	Y	Y	Y
R-square	0.7132	0.7502	0.6511	0.6905	0.7168	0.4460
N	1223	264	1424	594	820	7358
	Machine tool	Electrical goods	Gas equip.	Rail vehicle	All high gr. sec. OLS	All high gr. sec. IV
	(7)	(8)	(9)	(10)	(11)	(12)
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log dist. nearest 1851 rail station km	-0.1006	0.1945	-0.0233	-0.2171	-0.1039	-0.2081
	(0.001)	(0.457)	(0.901)	(0.302)	(0.000)	(0.021)
3rd deg. poly. in log. 1851 own emp.	Y	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y	Y
Registration district FE	Y	Y	Y	N	Y	Y
Geo. and transport controls	Y	Y	Y	Y	Y	Y
Units with town nodes in LCP	Y	Y	Y	Y	Y	Y
Kleibergen-Paap rk Wald F stat.						58.756
R-square	0.5451	0.8941	0.8487	0.3659	0.4510	0.4452
N	2617	243	369	99	7647	7647

Notes: Standard errors are clustered on registration district in all specifications. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, distance to nearest large city and share of soil types. Pre-railway transport includes distance to nearest inland waterway, port, steamship port, and turnpike road. Own employment applies to male employment in columns (1) and (2), and to the employment category in all other columns. (The instrument is distance to the LCP connecting large towns in 1801. see text for more details.

Table 6: Effect of average distance to nearest railway station on log difference male employment density in high growth tertiary sectors

	Media	Financial services	Comm. and admin.	Railway transp.	All high gr. tert. OLS	All high gr. tert. IV
	(1)	(2)	(3)	(4)	(5)	(6)
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
Log dist. nearest 1851 rail station km	-0.0337	-0.0565	-0.0893	-0.1510	-0.1781	-0.2980
	(0.688)	(0.046)	(0.000)	(0.000)	(0.000)	(0.040)
3rd deg. poly. in log. 1851 own emp.	Y	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y	Y
Registration district FE	Y	Y	Y	Y	Y	Y
Geo. and transport controls	Y	Y	Y	Y	Y	Y
Units with town nodes in LCP	Y	Y	Y	Y	Y	Y
Kleibergen-Paap rk Wald F stat.						37.358
R-square	0.7652	0.6108	0.5980	0.6070	0.4998	0.4949
N	737	1938	2566	2758	4498	4498

Notes: Standard errors are clustered on registration district in all specifications. Geographic controls include indicators for exposed coal, coastal, elevation, average elevation slope and standard deviation within parish, distance to nearest large city and share of soil types. Pre-railway transport includes distance to nearest inland waterway, port, steamship port, and turnpike road. Own employment applies to male employment in columns (1) and (2), and to the employment category in all other columns. (The instrument is distance to the LCP connecting large towns in 1801. see text for more details.

Table 7: Estimates for Heterogeneous effects of railway distance depending on 1851 population and employment density

Panel A			
percentile 1851 population density	model 1: Interactions effect of 1SD change in log distance	model 2: no interactions effect of 1SD change in log distance	ratio coefficient 1 to coefficient 2
5	-0.0385	-0.05974	0.6445
15	*-0.0577*	-0.05974	0.9659
25	*-0.0645*	-0.05974	1.0797
35	*-0.0684*	-0.05974	1.145
45	*-0.0709*	-0.05974	1.1868
55	*-0.0724*	-0.05974	1.2119
65	*-0.0731*	-0.05974	1.2237
75	*-0.0728*	-0.05974	1.2186
85	*-0.069*	-0.05974	1.155
95	-0.0205	-0.05974	0.3432

Panel B			
percentile 1851 secondary emp. density	model 1: Interactions effect of 1SD change in log distance	model 2: no interactions effect of 1SD change in log distance	ratio coefficient 1 to coefficient 2
5	-0.0338	-0.0642	0.5265
15	-0.0738	-0.0642	1.1495
25	-0.0846	-0.0642	1.3178
35	-0.0885	-0.0642	1.3785
45	*-0.0899*	-0.0642	1.4003
55	*-0.0896*	-0.0642	1.3956
65	*-0.0876*	-0.0642	1.3645
75	*-0.0826*	-0.0642	1.2866
85	-0.0673	-0.0642	1.0483
95	0	-0.0642	0

Panel C			
percentile 1851 tertiary emp. density	model 1: Interactions effect of 1SD change in log distance	model 2: no interactions effect of 1SD change in log distance	ratio coefficient 1 to coefficient 2
5	-0.096	-0.09787	0.9808
15	*-0.1178*	-0.09787	1.2036
25	*-0.1241*	-0.09787	1.2679
35	*-0.1269*	-0.09787	1.2966
45	*-0.1281*	-0.09787	1.3088
55	*-0.1281*	-0.09787	1.3088
65	*-0.1269*	-0.09787	1.2966
75	*-0.1235*	-0.09787	1.2618
85	*-0.1122*	-0.09787	1.1464
95	-0.0273	-0.09787	0.2789

Notes: Italics represent statistical significance at the 5% level or below. Sources: see text.

Table 8: A comparison of the effect of different transport variables on population and employment growth 1851 to 1881: standardized coefficients

	1	2	3	4	5
	Pop. growth	Sec. growth	Tert. Growth	Agric. Growth	Extract. Growth
	Stand. coeff. (p-value)	Stand. coeff (p-value)	Stand. coeff (p-value)	Stand. coeff (p-value)	Stand. coeff (p-value)
Log distance to nearest 1851 rail station in km	-0.1530	-0.0927	-0.0630	0.0242	-0.0551
	(0.000)	(0.000)	(0.000)	(0.066)	(0.050)
Log distance to nearest 1830 inland waterway in km	-0.0350	-0.0240	0.0017	0.0250	0.0132
	(0.008)	(0.091)	(0.080)	(0.002)	(0.616)
Log distance to nearest 1840 steamship port in km	-0.1760	-0.1056	0.0839	0.0450	-0.0264
	(0.003)	(0.074)	(0.679)	(0.313)	(0.767)
Log distance to nearest general port in km 1840	-0.0537	-0.0061	0.0116	0.0344	-0.0737
	(0.062)	(0.825)	(0.189)	(0.164)	(0.037)
Log distance to nearest turnpike in km 1830	-0.0724	-0.0849	-0.0176	0.0084	0.0081
	(0.000)	(0.000)	(0.000)	(0.042)	(0.653)
3rd deg. poly. in log. 1851 pop. or emp.	Y	Y	Y	Y	Y
1851 pop. den. and pop. growth 21 to 51	Y	Y	Y	Y	Y
Registration district FE	Y	Y	Y	Y	Y
Geo. and transport controls	Y	Y	Y	Y	Y
Units with town nodes in LCP	Y	Y	Y	Y	Y
R-square	0.4835	0.3998	0.5216	0.5630	0.4777
N	9474	9046	9306	9389	3374

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	40.3
Counter-factual population growth 1851 to 1881 in % with 1841 network	31.8
Change in population growth in %	-8.5
Model predicted secondary emp. growth 1851 to 1881 in % with 1851 network	52.1
Counter-factual secondary emp. growth 1851 to 1881 in % with 1841 network	43.2
Change in secondary employment growth in %	-8.9
Model predicted tertiary emp. growth 1851 to 1881 in % with 1851 network	76.9
Counter-factual tertiary emp. growth 1851 to 1881 in % with 1841 network	72.9
Change in tertiary employment growth in %	-4.0
Model predicted agricultural growth 1851 to 1881 in % with 1851 network	-10.7
Counter-factual agricultural growth 1851 to 1881 in % with 1841 network	-9.2
Change in agricultural employment growth in %	1.5

Notes: for sources and details of calculations see text.

A Data Appendix

To be filled.