

The Making of the Modern Metropolis: Evidence from London*

Stephan Heblich[†]
University of Bristol

Stephen J. Redding[‡]
Princeton University, NBER and CEPR

Daniel M. Sturm[§]
London School of Economics and CEPR

September 20, 2017

Abstract

Modern metropolitan areas involve large concentrations of economic activity and the transport of millions of people each day between their residence and workplace. However, relatively little is known about the role of these commuting flows in promoting agglomeration forces. We use the revolution in transport technology from the invention of steam railways, newly-constructed spatially-disaggregated data for London from 1801-1921, and a quantitative urban model to provide evidence on the determinants of the concentration of economic activity in metropolitan areas. Steam railways dramatically reduced travel times and hence permitted the first large-scale separation of workplace and residence to realize economies of scale. We show that our model is able to account both qualitatively and quantitatively for the observed changes in city size, structure and land prices.

KEYWORDS: agglomeration, urbanization, transportation,
JEL CLASSIFICATION: O18, R12, R40

Work in Progress

*We are grateful to Bristol University, the London School of Economics, and Princeton University for research support. We are grateful to colleagues for helpful comments. We would like to thank the Cambridge Group for the History of Population and Social Structure, the British Library and London Metropolitan Archives (LMA) for their help with data. We are grateful to Iain Bamford, Dennis Egger, Andreas Ferrara, Ben Glaeser and Florian Trouvain for excellent research assistance. The usual disclaimer applies.

[†]Dept. Economics, Priory Road Complex, Priory Road, Clifton, BS8 1TU. UK. Tel: 44 117 3317910. Email: stephan.heblich@bristol.ac.uk.

[‡]Dept. Economics and WWS, Fisher Hall, Princeton, NJ 08544. Tel: 1 609 258 4016. Email: reddings@princeton.edu.

[§]Dept. Economics, Houghton Street, London, WC2A 2AE. UK. Tel: 44 20 7955 752. Email: d.m.sturm@lse.ac.uk.

1 Introduction

Modern metropolitan areas include vast concentrations of economic activity, with Greater London and New York City today accounting for around 8.4 and 8.5 million people respectively. These immense population concentrations involve the transport of millions of people each day between their residence and workplace. Today, the London Underground alone handles around 3.5 million passenger journeys per day, and its trains travel around 76 million kilometers each year (about 200 times the distance between the earth and the moon). Yet relatively little is known about the role of these commuting flows in promoting the agglomeration forces that sustain metropolitan areas. On the one hand, these commuting flows impose substantial real resource costs, both in terms of time spent commuting and the construction of large networks of complex transportation infrastructure. On the other hand, they are also central to creating dense employment concentrations to realize economies of scale in business districts and foster the amenities available in distinctively residential neighborhoods.

In this paper, we use the mid-nineteenth century transport revolution from the invention of steam railways, a newly-created, spatially-disaggregated dataset for Greater London from 1801-1921, and a quantitative urban model to provide new evidence on the determinants of agglomeration. The key idea behind our approach is that the slow travel times achievable by human or horse power implied that most people lived close to where they worked when these were the main modes of transportation. In contrast, the invention of steam railways dramatically reduced the time taken to travel a given distance, thereby permitting the first large-scale separation of workplace and residence. This specialization by workplace and residence in turn enabled the realization of economies of scale in production and residential choices. Using both reduced-form and structural approaches, we find substantial effects of steam passenger railways on city size and structure. We show that our model is able to account both qualitatively and quantitatively for the observed changes in city size, structure and land prices.

Methodologically, we develop a new structural estimation procedure for the class of urban models characterized by a gravity equation for commuting flows. Although we only observe these bilateral commuting flows in 1921 at the end of our sample period, we show how this framework can be used to estimate the impact of the construction of the railway network. Combining our 1921 gravity equation data with historical information on population, land values and the transport network back to the early-nineteenth century, we use the model to infer missing employment workplace for each location going backwards in time. In overidentification checks, we compare these model predictions to the historical data on employment by workplace that do exist for the City of London. We find substantial direct effects of the railway through reduced commuting costs, but we also find substantial changes in the relative productivity and amenities of different locations within Greater London, which are consistent with agglomeration forces in production and residential choices.

Nineteenth-century London is arguably the poster child for the large metropolitan areas observed around the world today. In 1801, London's built-up area housed around 1 million people and spanned only 5 miles East to West. This was a walkable city of 60 squares and 8,000 streets that was not radically different from other large cities from history. In contrast, by 1901, Greater London contained over 6.5 million people, measured more than 17 miles across, and was on a dramatically larger scale than any previous urban area. This was the largest city in the world by some margin (with New York City and Greater Paris having populations of 3.4 million and 4 million respectively at the

turn of the twentieth century) and London’s population exceeded that of several European countries.¹ Therefore, nineteenth-century London provides a natural testing ground for assessing the empirical relevance of theoretical models of city size and structure.

Our empirical setting also has a number of other attractive features. During this period, there is a revolution in transport technology in the form of the steam locomotive, which was initially developed to haul freight at mines (at the Stockton to Darlington Railway in 1825), and only later applied to passenger transport (with the London and Greenwich Railway in 1836 the first to be built specifically for passengers).² In contrast to other cities such as Paris, London developed through a largely haphazard and organic process. Until the creation of the Metropolitan Board of Works (MBW) in 1855, there was no municipal authority that spanned the many different local jurisdictions that made up Greater London, and the MBW’s responsibilities were largely centered on infrastructure. Only in 1889 was the London County Council (LCC) created, and the first steps towards large-scale urban planning for Greater London were not taken until the Barlow Commission of 1940. Therefore, nineteenth-century London provides a setting in which we would expect both the size and structure of the city to respond to decentralized market forces.

We contribute to several strands of existing research. First, our paper connects with the theoretical and empirical literatures on agglomeration, including [Henderson \(1974\)](#), [Fujita, Krugman, and Venables \(1999\)](#), [Fujita and Thisse \(2002\)](#), [Davis and Weinstein \(2002\)](#), [Davis and Dingel \(2012\)](#) and [Kline and Moretti \(2014\)](#), as reviewed in [Rosenthal and Strange \(2004\)](#), [Duranton and Puga \(2004\)](#), [Moretti \(2011\)](#) and [Combes and Gobillon \(2015\)](#). A key challenge in empirical work on agglomeration is finding exogenous sources of variation to identify agglomeration forces. [Rosenthal and Strange \(2008\)](#) and [Combes, Duranton, Gobillon, and Roux \(2010\)](#) use geology as an instrument for population density, exploiting the idea that tall buildings are easier to construct where solid bedrock is accessible. [Greenstone, Hornbeck, and Moretti \(2010\)](#) provide evidence on agglomeration spillovers by comparing changes in total factor productivity (TFP) among incumbent plants in “winning” counties that attracted a large manufacturing plant and “losing” counties that were the new plant’s runner-up choice. In contrast, we exploit the transformation of the relationship between travel time and distance provided by the invention of the steam locomotive.

Second, our paper is related to a recent body of research on quantitative spatial models, including [Allen and Arkolakis \(2014\)](#), [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#), [Redding and Sturm \(2008\)](#), [Redding \(2016\)](#), [Monte \(2016\)](#), [Caliendo, Parro, Rossi-Hansberg, and Sarte \(2017\)](#), [Desmet, Nagy, and Rossi-Hansberg \(2017\)](#), [Allen, Arkolakis, and Li \(2017\)](#) and [Monte, Redding, and Rossi-Hansberg \(2017\)](#), as reviewed in [Redding and Rossi-Hansberg \(2017\)](#). All of these papers focus on time period for which modern transportation networks by rail and/or road existed, whereas we exploit the dramatic change in transport technology provided by the steam locomotive. We borrow our basic model structure from [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#), which introduces heterogeneity in worker commuting decisions following [McFadden \(1974\)](#) into the urban model of [Lucas and Rossi-Hansberg \(2002\)](#). Our main methodological contribution is to develop a new structural estimation procedure for quantitative urban models of this form that feature a gravity equation for commuting flows. This structural estimation procedure uses bilateral commuting flows for a baseline year (in our case 1921) and estimates the model’s parameters by undertaking comparative statics from this baseline year (in our case backwards in time).³ We show that this procedure can be used to recover unobserved

¹London overtook Beijing’s population in the 1820s, and remained the world’s largest city until the mid-1920s, when it was eclipsed by New York. By comparison, Greece’s 1907 population was 2.6 million, and Denmark’s 1901 population was 2.4 million.

²Stationary steam engines have a longer history, dating back at least to Thomas Newcomen in 1712, as discussed further below.

³In [Ahlfeldt, Redding, Sturm, and Wolf \(2015\)](#), only data on a random sample of bilateral commuting flows were available, which were not

historical employment by workplace (prior to 1921) from the bilateral commuting data for our baseline year and historical data on population and land rents. This procedure is applicable in other contexts, in which historical data are incomplete or missing, but bilateral commuting data are available for a baseline year.

Third, our paper is related to a growing empirical literature on the relationship between the spatial distribution of economic activity and transport infrastructure, as reviewed in [Redding and Turner \(2015\)](#). One strand of this literature has used variation across cities and regions, including [Duranton and Turner \(2011\)](#), [Chandra and Thompson \(2000\)](#), [Duranton and Turner \(2012\)](#), [Faber \(2014\)](#), [Michaels \(2008\)](#), [Donaldson \(2017\)](#), [Duranton, Morrow, and Turner \(2014\)](#), and [Baum-Snow, Henderson, Turner, Zhang, and Brandt \(2017\)](#). A second strand of this literature has looked within cities, including [Warner \(1978\)](#), [Jackson \(1987\)](#), [McDonald and Osuji \(1995\)](#), [Gibbons and Machin \(2005\)](#), [Baum-Snow and Kahn \(2005\)](#), [Billings \(2011\)](#), [Brooks and Lutz \(2013\)](#), and [Gonzalez-Navarro and Turner \(2016\)](#). Within this literature, our work is most closely related to research on suburbanization and decentralization, including [Baum-Snow \(2007\)](#), [Baum-Snow, Brandt, Henderson, Turner, and Zhang \(2017\)](#), and [Baum-Snow \(2017\)](#). Our contributions are again to use the large-scale variation from the transition from human/horse power to steam locomotion and to show that our model can account both qualitatively and quantitatively for the observed changes in city structure.

The remainder of the paper is structured as follows. Section 2 discusses the historical background. Section 3 summarizes the data sources and definitions. Section 4 presents reduced-form evidence on the role of transport infrastructure improvements in shaping patterns of urban development within Greater London. Section 5 introduces our theoretical model. Section 6 undertakes a quantitative analysis of the model centered around its gravity equation predictions for bilateral commuting flows. Section 7 concludes.

2 Historical Background

London has a long history of settlement that dates back to before the Roman Conquest of England in 43 CE. We distinguish four main definitions of its geographical boundaries, which we now list from largest to smallest, where each subsequent region is a subset of the previous one. First, we consider London together with the Home Counties that surround it, which contain a 1921 population of 9.61 million and an area of 12,829 kilometers squared, and encompass large parts of South-East England.⁴ Second, we examine Greater London, as defined by the modern boundaries of the Greater London Authority (GLA), which includes a 1921 population of 7.39 million and an area of 1,595 kilometers squared. Third, we consider the historical County of London, which has a 1921 population of 4.48 million and an area of 314 kilometers squared. Fourth, we examine the City of London, which has a 1921 population of 13,709 and an area of around 3 kilometers squared, and whose boundaries correspond approximately to the Roman city wall. From medieval times, the City of London acted as the main commercial and financial center of what became the United Kingdom, with the neighboring City of Westminster serving as the seat of Royal and Parliamentary government.⁵

Data are available for these four main geographical regions at two main levels of spatial aggregation: boroughs and parishes. The Home Counties including Greater London encompasses 257 boroughs and 1,161 parishes; Greater London contains 99 boroughs and 285 parishes; the County of London comprises 29 boroughs and 184 parishes; and the City of London includes 1 borough and 111 parishes.⁶ In Figure 1, we show the outer boundary of the Home

representative for each workplace and residence, and hence would have precluded the use of this estimation procedure.

⁴The Home Counties are the counties of Essex, Hertfordshire, Kent, Middlesex and Surrey.

⁵For historical discussions of London, see [Ball and Sunderland \(2001\)](#), [Kynaston \(2012\)](#), [Porter \(1995\)](#) and [White \(2012, 2007, 2008\)](#).

⁶Parish boundaries in the population census change over time. We construct constant definitions of parish boundaries using the classification

Counties with a thick black line; the boundary of Greater London with a thick red line; the boundary of the County of London with a thick purple line; and the boundary of the City of London with a thick green line (barely visible). Borough boundaries are shown with medium black lines; parish boundaries are indicated using medium gray lines; and the River Thames is denoted by the thick blue line. As apparent from the figure, our data permit a high-level of spatial resolution, where the median parish in Greater London has a 1901 population of 1,124 and an area of 7 kilometers squared, while the median borough in Greater London has a 1921 population of 35,639 and an area of 11 kilometers squared.

In the first half of the 19th-century, there was no municipal authority for the entire built-up area of Greater London, and public goods were largely provided by local parishes and vestries (centered around churches). As a result, in contrast to other cities such as Paris, London's growth was largely haphazard and organic.⁷ In response to the growing public health challenges created by an expanding population, the Metropolitan Board of Works (MBW) was founded in 1855, although its main responsibilities were for infrastructure, and many powers remained in the hands of the parishes and vestries.⁸ With the aim of creating a central municipal government with the powers required to deliver public services effectively, the London County Council (LCC) was formed in 1889. The new County of London was created from the Cities of London and Westminster and parts of the surrounding counties of Middlesex, Surrey and Kent.⁹ As the built-up area continued to expand, the concept of Greater London emerged, which was ultimately reflected in the replacement of the LCC by the Greater London Council (GLC) in 1965. Following the abolition of the GLC in 1985 by the government of Margaret Thatcher, Greater London again had no central municipal government, until the creation of the Greater London Authority (GLA) in 1999.

At the beginning of the 19th-century, the most commonly-used mode of transport was walking, with average travel speed in good road conditions of around 3 miles per hour (mph). The state of the art technology for long distance travel was the stage coach, but it was expensive because of the multiple changes in teams of horses required over long distances, and hence was relatively infrequently used. Even with this elite mode of transport, poor road conditions limited average long distance travel speeds to around 5 mph (see for example [Gerhold 2005](#)). Given these limited transport possibilities, most people lived close to where they worked, as discussed in the analysis of English 18th-century time use in [Voth \(2001\)](#). With the growth of urban populations, attempts to improve existing modes of transport led to the introduction of the horse omnibus from Paris to London in the 1820s. Its main innovation relative to the stage coach was increased passenger capacity for short-distance travel. However, the limitations of horse power and road conditions ensured that average travel speeds remained low at around 6 mph.¹⁰ A further innovation along the same lines was the horse tram (introduced in London in 1860), but average travel speeds again remained low, in part because of road congestion (again at around 6 mph).¹¹

Against this background, the steam passenger railway constituted a major transport innovation, although one with a long and uncertain gestation. The first successful commercial development of a stationary steam engine was

provided by [Shaw-Taylor, Davies, Kitson, Newton, Satchell, and Wrigley \(2010\)](#), as discussed further below.

⁷The main exceptions are occasional Royal interventions, such as the creation of Regent Street on the initiative of the future George IV in 1825.

⁸See for example [Owen \(1982\)](#). The main achievements of the MBW were the construction of London's Victorian sewage system and the Thames embankment, as discussed in [Halliday \(1999\)](#).

⁹The LCC continued the MBW's infrastructure improvements, including some new road construction through housing clearance (e.g. Kingsway close to the London School of Economics), and built some social housing. The first steps towards large-scale urban planning for Greater London were not taken until the Barlow Commission in 1940, as discussed in [Foley \(1963\)](#).

¹⁰See for example [Barker and Robbins \(1963\)](#) and [London County Council \(1907\)](#).

¹¹A later innovation was the replacement of the horse tram with the electric tram (with the first fully-operational services starting in 1901), but average travel speeds remained low at around 8 mph, again in part because of road congestion.

by Thomas Newcomen in 1712 to pump mine water. However, the development of the separate condenser and rotary motion by James Watt from 1763-75 substantially improved its efficiency and expanded its range of potential applications. The first commercial use of mobile steam locomotives was to haul freight from mines at the Stockton and Darlington railway in 1825. However, in part as a result of fears about the safety of steam locomotives and the dangers of asphyxiation from rapid travel, it was not until 1833 that carriages with passengers were hauled by steam locomotives at this railway. Only in 1836 did the London and Greenwich railway open as the first steam railway to be built specifically for passengers. The result was a dramatic transformation of the relationship between travel time and distance, with average travel speeds using this new technology of around 20 mph.¹²

Railway development in London, and the United Kingdom more broadly, was undertaken by private companies in a competitive and uncoordinated fashion.¹³ These companies submitted proposals for new railway lines for authorization by Acts of Parliament. In response to a large number of proposals to construct railway lines through Central London, a Royal Commission was established in 1846 to investigate these proposals. To preserve the built fabric of Central London, this Royal Commission recommended that railways be excluded from a central area delineated by the Euston Road to the North and the Borough and Lambeth Roads to the South.¹⁴ A legacy of this recommendation was the emergence of a series of railway terminals around the edge of this central area, which led to calls for an underground railway to connect these terminals. These calls culminated in the opening of the Metropolitan District Railway in 1863 and the subsequent development of the Circle and District underground lines. While these early underground railways were built using “cut and cover” methods, further penetration of Central London occurred with the development of the technology for boring deep-tube underground railways, as first used for the City and South London Railway, which opened in 1890, and is now part of the Northern Line.¹⁵

In Figures 2, 3 and 4, we display maps of the overground and underground railway network for 1841, 1881 and 1921 respectively. The parts of the Home Counties outside Greater London are shown with a white background; the areas of Greater London outside the County of London are displayed with a blue background; and the County of London is indicated with a gray background. Overground railway lines are shown in black and underground railway lines are displayed in red. In 1841, which is the first population census year in which any overground railways are present, there are only a few railway lines. These radiate outwards from the County of London, with a relatively low density of lines in the center of the County of London, which in part reflects the parliamentary exclusion zone. Four decades later in 1881, the County of London is criss-crossed by a dense network of railway lines, with greater penetration into the center of the County of London, in part because of the construction of the first underground railway lines. Another four decades later in 1921, there is a further increase in the density of railway lines, which is greatest for the parts of Greater London outside of the County of London.

¹²Consistent with this difference in travel speeds, railways were more frequently used for longer-distance travel, while omnibuses and trams were more important over shorter distances (including from railway terminals to final destinations), which tended to make these alternative modes of transport complements rather than substitutes. The share of railways in all passengers journeys by public transport was 49 percent in 1867 (the first year for which systematic data are available) and 32 percent in 1921 (see [London County Council 1907](#)). From 1860 onwards, Acts of Parliament authorizing railways typically included clauses requiring the provision of “workmen’s trains” with cheap fares for working-class passengers, as ultimately reflected in the 1883 Cheap Trains Act (see for example [Abernathy 2015](#)).

¹³For further historical discussion of railway development, see for example [Croome and Jackson \(1993\)](#), [Kellet \(1969\)](#), and [Wolmar \(2009, 2012\)](#).

¹⁴This parliamentary exclusion zone explains the location of Euston, King’s Cross and St. Pancras railway terminals all on the Northern side of the Euston Road. Exceptions were subsequently allowed, often in the form railway terminals over bridges coming from the south side of the Thames at Victoria (1858), Charring Cross (1864), Cannon Street (1866), and Ludgate Hill (1864), and also at Waterloo (1848).

¹⁵When it opened in 1863, the Metropolitan District Railway used steam locomotives. In contrast, the City and South London Railway was the first underground line to use electric traction from its opening in 1890 onwards.

3 Data

We construct a new spatially-disaggregated dataset for London for the period 1801-1921. Our main data source is the population census of England and Wales, which begins in 1801, and is enumerated every decade thereafter. A first key component for our quantitative analysis of the model is the complete matrix of bilateral commuting flows between the boroughs of England and Wales, which is reported for the first time in the 1921 population census.¹⁶ Using this matrix, we find that commuting flows between other parts of England and Wales and Greater London were small in 1921, such that Greater London was largely a closed commuting market.¹⁷ Summing across rows in the matrix of bilateral commuting flows for Greater London, we obtain employment by workplace for each borough (which we refer to as “workplace employment”). Summing across columns, we obtain employment by residence for each borough (which we refer to as “residence employment”). We also construct an employment participation rate for each borough in 1921 by dividing residence employment by population.

We combine these data on bilateral commuting flows for 1921 with historical population data for each parish and borough from earlier population censuses from 1801-1911. Assuming that the ratio of residence employment to population is stable for a given borough over time, we use the 1921 value of this ratio and the historical population data to construct residence employment for each borough for each decade from 1801-1911.¹⁸ Parish and borough boundaries are relatively stable throughout most of the nineteenth century, but experience substantial change in the early-twentieth century. For our reduced-form empirical analysis using the parish-level data, we constructed constant parish boundary data for the period 1801-1901. For our quantitative analysis of the model using the borough-level data, we use constant borough definitions through our sample period based on the 1921 boundaries. For years prior to 1921, we allocate the parish-level data to the 1921 boroughs by weighting the values for each parish by its share of the geographical area of the 1921 borough. Given that parishes have a much smaller geographical area than boroughs, most parishes lie within a single 1921 borough.

We measure the value of land for each parish and borough using rateable values, which correspond to the market valuation of land and buildings. In particular, the rateable value is defined as “The annual rent which a tenant might reasonably be expected, taking one year with one another, to pay for a hereditament, if the tenant undertook to pay all usual tenant’s rates and taxes ... after deducting the probable annual average cost of the repairs, insurance and other expenses” (see [London County Council 1907](#)). Valuations include all public services (such as railways, tramways, utilities etc), government property (such as courts, parliaments etc), and private property (such as factories, warehouses, wharves, offices, shops, theaters, music halls, clubs, hospitals, and all residential dwellings). These rateable values have a long history in England and Wales. They were originally used to raise revenue for local public goods, and different types of rateable values can be distinguished, depending on the use of the revenue raised. Where available, we use the rateable values from Schedule A of the national income tax (introduced in 1799), which is the schedule concerned with profits from the ownership of lands and property. As these rateable values are the basis for national income tax collection, they are widely regarded as corresponding most closely to market valuations.¹⁹ Re-valuations

¹⁶The population census for England and Wales is one of the first to report bilateral commuting data. In the United States, the 1960 population census is the first to report any commuting information, and the matrix of bilateral commuting flows between counties is not reported until 1990.

¹⁷In the 1921 population census, 96 percent of the workers employed in Greater London also lived in Greater London. Of the remaining 4 percent, approximately half live in the Home Counties, and the remainder live in other parts of England and Wales. As residence is measured based on where one slept on Census night, while workplace is one’s usual place of work, some of this 4 percent could be due to business trips or other travel.

¹⁸Empirically, we find relatively little variation in employment participation rates across boroughs in 1921.

¹⁹For example, [Stamp \(1922\)](#) argues that “It is generally acknowledged that the income tax, Schedule A, assessments are the best approach to the

were undertaken every five years. For the County of London, data are reported every five years from 1830 onwards and annually from 1871 onwards in the publications of London County Council. For the remainder of Greater London, data are reported for 1921 in the publications of London County Council and for 1815, 1843, 1848, 1852, 1860, 1881 and 1896 in the publications of the House of Commons. We use linear interpolation between these years to construct time-series on rateable values for each parish for each census year from 1831-1921.

Except for the City of London, data on workplace employment are not available prior to 1921. Our structural estimation of the model uses our bilateral commuting data for 1921, together with our data on residence employment and rateable values for earlier years, to generate model predictions for workplace employment for these earlier years. In overidentification checks, we compare these model predictions to the data on workplace employment for the City of London that are available from the Day Censuses of 1866, 1881, 1891 and 1911.²⁰ In the face of a declining residential population (“night population”), the City of London Corporation undertook these censuses of the “day population” to demonstrate its enduring commercial importance. The “day population” is defined as “every person ... residing, engaged, occupied, or employed in each and every house, warehouse, shop, manufactory, workshop, counting house, office, chambers, stable, wharf etc ... during the working hours of the day, whether they sleep or do not sleep there.”²¹ Given the high opportunity cost of unused space in the City of London, those present during the day are likely to be employed. Below, we compare these day population figures to our 1921 workplace employment data to provide evidence that the day population is indeed a good measure of workplace employment.

Finally, we combine these data with a variety of other Geographical Information Systems (GIS) data, including maps of the overground and underground railway network, and data on the residence and workplace addresses of barristers (a type of lawyer) from post office directories for the years 1841, 1852, 1882, 1899 and 1921, as discussed further in the data appendix.

4 Reduced-Form Evidence

In subsection 4.1, we present reduced-form evidence on the evolution of the organization of economic activity within Greater London over our historical time period. In subsection 4.2, we provide further evidence on the role of railways in shaping this reorganization of economic activity. In subsection 4.3, we report a non-parametric specification that estimates a separate railway treatment for each parish in Greater London.

4.1 City Size and Structure over Time

We first illustrate the dramatic changes in the spatial organization of economic activity within Greater London over our sample period. In Figure 5, we display total population over time for the City of London (left panel) and Greater London (right panel). In each case, population is expressed as an index relative to its value in 1801 (such that 1801=1). In the first half of the nineteenth century, population in the City of London was relatively constant (at around 130,000), while population in Greater London grew substantially (from 1.14 million to 2.69 million). From 1851 onwards (shown by the red vertical line), there is a sharp drop in population in the City of London, which falls by around 90 percent to

true values” (page 25). For years where these data are not available, we use the County or Poor Law rateable values. For years where both sets of data are reported, we find that they are highly correlated with one another. After the Metropolis Act of 1869, all rateable values for the County of London are computed on the basis of Schedule A income taxation.

²⁰See [Corporation of London \(1866\)](#), [Corporation of London \(1881\)](#), [Salmon \(1891\)](#), and [Monckton \(1911\)](#) respectively.

²¹[Salmon \(1891\)](#), page 97.

13,709 by 1921, with this rate of decline slowing over time towards the end of the nineteenth century and beginning of the twentieth century. Over the same period, the population of Greater London as a whole continues to grow rapidly from 2.69 million in 1851 to 7.39 million in 1921.²² Therefore, the rapid expansion in population for Greater London throughout the nineteenth century goes hand in hand with a precipitous drop in population in its most important commercial center from the mid-nineteenth century onwards.

In Figure 6, we contrast the evolution of “night” and “day” population for the City of London. The “night” population data are the same as those shown in the left panel of Figure 5 and are taken from the population census (based on residence on census night). In contrast, the “day” population data are from the City of London Day Censuses, except for the 1921 figure, which equals workplace employment from the population census.²³ Figure 6 shows that the sharp decline in night population in the City of London from 1851 onwards coincides with a sharp rise in its day population. This pattern suggests that the combination of population decline in the City of London and population expansion for Greater London as a whole is explained by the City of London increasingly specializing as a workplace rather than as a residence. Extrapolating the day population series further back in time to the 1850s (not shown in the figure) would suggest that night and day population were approximately equal to one another at this time, which is consistent with most people living close to where they worked in the early decades of the nineteenth century. At the end of the sample period, the workplace employment number for 1921 from the population census is in line with a continuation in the trend in the day population from the City of London day census, which supports the idea that most of the people recorded in the City of London day census were indeed employed.

We now show that the population decline in the City of London from the mid-nineteenth century onwards is not part of an economic decline in this location. Instead, in the period in which the City of London increasingly specializes as a workplace rather than as a residence, it becomes a relatively more valuable location. In Figure 7, we display the City of London’s share of rateable value in the County of London. In the early-nineteenth century, this share declines from 11-9 percent, which is consistent with a geographical expansion in the built-up area of Greater London. As this geographical expansion occurs, and undeveloped land becomes developed, the share of already-developed land in overall land values tends to fall. However, in contrast to this pattern, in the years after 1851 when the City of London experiences the largest declines in residential population, its rateable value share *increases* sharply from 9-14 percent. In the decades at the end of the nineteenth century, the pattern of a decline in this rateable value share again reasserts itself, consistent with the continuing geographical expansion in the built-up area of Greater London. Therefore, the steep population decline in the City of London in the decades immediately after 1851 involves an *increase* rather than a decrease in the relative value of this location.

A comparison of Figures 5-6 to the evolution of the railway network over time in Figures 2-4 above already suggests that the observed changes in night population, day population and land value are likely to be related to the change in transport technology. The delayed response in the City of London’s population in Figure 5 (from 1851 onwards) relative to the first opening of a steam railway (in 1836) is consistent with the railway network becoming increasingly valuable as more locations are connected to it and with it taking time for firms and workers to adjust to the new transport technology. The sharpest declines in the population of the City of London from 1851-1881 in Figure 5 correspond closely to the greatest increases in the penetration of overground and underground railways into the

²²Although the second decade of the twentieth century spans the First World War from 1914-18, the primitive nature of aircraft and airship technology at that time ensured that Greater London experienced little bombing and destruction (see for example [White 2008](#)).

²³The City of London Day Censuses were taken in 1866, 1881, 1891 and 1911. We interpolate between neighboring years for 1871 and 1901.

central areas of the County of London in Figure 3. We provide further evidence below on this timing of the population response to the new transport technology using a difference-in-differences specification.

4.2 Difference-in-Differences Specification

We now use our spatially-disaggregated parish-level data for Greater London from 1801-1901 to provide evidence on the role of railways in driving this reorganization of economic activity. The main identification challenge is that railways are unlikely to be randomly assigned, because they were constructed by private-sector companies, whose stated objective was to maximize shareholder value. As a result, parishes in which economic activity would have grown for other reasons could be more likely to be assigned railways. We address this identification challenge by considering specifications that include both a parish fixed effect and a parish time trend, and examining the relationship between the timing of deviations from these parish trends and the arrival of the railway. We start by estimating an average treatment effect for Greater London as a whole and later explore heterogeneity within Greater London. Before including our full set of controls, we consider the following baseline specification:

$$\ln H_{\ell t} = \alpha_{\ell} + \sum_{\tau=0}^{\tau=T} \beta_{\tau} (R_{\ell} \times \mathbb{I}_{\tau}) + d_t + u_{\ell t} \quad (1)$$

where ℓ indexes parishes; t indicates the census year; $H_{\ell t}$ is parish population; α_{ℓ} is a parish fixed effect; d_t is a census year dummy; R_{ℓ} is an indicator variable that equals one if a parish has an overground or underground railway station in at least one census year during our sample period; τ captures time relative to the treatment year, which is defined as the census year minus the last census year in which the parish has no railway (such that $\tau = 0$ is the treatment year),²⁴ and \mathbb{I}_{τ} is an indicator variable that equals one for parishes that are treated with a railway in treatment year τ and zero otherwise. We cluster the standard errors on boroughs, which allows the error term to be serially correlated within parishes over time, and to be correlated across parishes within boroughs.²⁵

The inclusion of the parish fixed effects (α_{ℓ}) controls for the non-random assignment of railways based on the level of log parish population or other time-invariant factors. Therefore, we allow for the fact that parishes treated with a railway could have had higher population levels in all years (both before and after the railway). The census year dummies (d_t) control for secular changes in population across all parishes. The key coefficients of interest $\{\beta_{\tau}\}$ are those on the interaction terms between the railway indicator (R_{ℓ}) and the treatment year indicator (\mathbb{I}_{τ}), which capture the treatment effect of a railway on parish population in treatment year τ . They have a “difference-in-differences” interpretation, where the first difference compares treated to untreated parishes, and the second difference undertakes this comparison before and after the arrival of the railway. The main effect of R_{ℓ} is captured in the parish fixed effects and the main effect of \mathbb{I}_{τ} is captured in the census year dummies. We include six interaction terms for decades from 10 to 60+ years after a parish receives a railway station. We aggregate treatment years greater than 60 into the 60+ category to ensure that this final category has a sufficient number of observations.²⁶

²⁴We define the treatment census year as the last census year in which the parish has no railway (rather than the first census year in which it has a railway), because census years only occur every decade, and hence our definition ensures that the parish cannot be affected by the railway before the treatment year. For example, if the railway arrives in a parish in 1836, the treatment census year ($\tau = 0$) is 1831, and the first census year after the treatment year is 1841 ($\tau = 10$).

²⁵We also experimented with Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors following [Conley \(1999\)](#), and found these to be typically smaller than the standard errors clustered on borough that are reported below.

²⁶The earliest treatment year is 1831, which implies a maximum of 70 years after the treatment in our parish sample. Of our 3,133 parish-year observations, 1,408 involve parishes that have a railway station in at least one census year during the sample period. The distribution of these 1,408 observations across the treatment years is: $\tau \leq -60$ (134); $\tau = -50$ (83); $\tau = -40$ (106); $\tau = -30$ (128); $\tau = -20$ (128); $\tau = -10$ (128); $\tau = 0$ (128); $\tau = 10$ (128); $\tau = 20$ (119); $\tau = 30$ (104); $\tau = 40$ (96); $\tau = 50$ (59); and $\tau \geq 60$ (67).

In Column (1) of Table 1, we report the results of estimating this baseline specification from equation (1). We find positive and statistically significant treatment effects of the railway on parish population, which range from around 60-270 percent (up to the log approximation). In Column (2), we augment this specification with a parish time trends, which allows for the non-random assignment of railways based on trends in log parish population over time.²⁷ In this specification, we allow for the fact that parishes treated with a railway could have had higher trend population growth in all years (both before and after the railway). We now identify the treatment effect of the railway solely from deviations from these parish time trends after the arrival of the railway. Again we find positive and statistically significant treatment effects, which are now somewhat smaller but still substantial, ranging from 11-135 percent. In Column (3), we present a robustness test, in which we drop the parish time trends and census year dummies, and replace them with a full set of interactions between census year dummies and twenty-five dummies for quantiles of distance from the Guildhall (as a measure of the center of Greater London).²⁸ This robustness test allows parish population growth rates to vary non-parametrically across the census decades depending on distance from the center of Greater London. Even though we abstract from any variation in population growth across the distance grid cells, we continue to find positive and statistically significant treatment effects of the railway on parish population.²⁹

In Column (4), we return to the specification with parish time trends, and check whether or not the timing of the deviation from these parish time trends coincides with the arrival of the railway. In particular, we augment specification from Column (2) with interaction terms between the railway dummy (R_ℓ) and dummies for treatment years before the arrival of the railway (\mathbb{I}_τ for $\tau < 0$). The excluded category is the treatment year ($\tau = 0$). We consider a symmetric time window, in which we include six interaction terms for decades from 10 to 60+ years before and after a parish receives a railway station. As apparent from Column (4), we find no evidence of statistically significant deviations from the parish time trends before the arrival of the railway. But we continue to find large and statistically significant deviations from these parish time trends in the years after the arrival of the railway. Therefore, this specification supports an interpretation of our estimates in Column (2) with parish fixed effects and parish time trends as capturing a causal effect of the railway on parish population.

We now use our “difference-in-differences” specification to provide further evidence connecting the decline in population in the City of London shown in Figure 5 above to the arrival of the railway. In particular, we allow the railway treatment effect to differ between the City of London and other parts of Greater London by augmenting our baseline specification from equation (1) (as reported in Column (1) of Table 1) with a three-way interaction term between the railway dummy, the treatment year dummy, and a dummy for the City of London:

$$\ln H_{\ell t} = \alpha_\ell + \sum_{\tau=0}^{\tau=T} \beta_\tau (R_\ell \times \mathbb{I}_\tau) + \sum_{\tau=0}^{\tau=T} \gamma_\tau \left(R_\ell \times \mathbb{I}_\tau \times \mathbb{I}_\ell^{\text{City}} \right) + d_t + u_{\ell t}, \quad (2)$$

where $\mathbb{I}_\ell^{\text{City}}$ is an indicator variable that equals one for parishes in the City of London and zero otherwise; all other variables are defined as above; the railway treatment effect for parishes in the City of London is now given by $(\beta_\tau + \gamma_\tau)$; and the railway treatment effect for other parts of Greater London remains equal to β_τ . A legacy of the parliamentary exclusion zone discussed above is that relatively few parishes within the City of London are treated by the railway, and these treatments occur relatively late in the sample period. As a result, there is a relatively short interval after

²⁷One of the parish time trends is colinear with the year dummies and hence is omitted without loss of generality.

²⁸Distance from the Guildhall varies from less than 1 kilometer to just over 110 kilometers across parishes within Greater London, so that each of the 25 distance grid cells includes just over 4 kilometers of distance.

²⁹As another robustness test, we experimented with defining the treatment based on only overground railways (instead of both underground and overground railways), and found a similar pattern of results.

the arrival of the railway for these parishes. Therefore, we only include three City-of-London interaction terms $\{\gamma_\tau\}$ for decades 10 to 30+ years after a parish receives a railway station.³⁰

In Column (5) of Table 1, we estimate this specification from equation (2). Again we find positive and statistically significant treatment effects of the railway on parish population for other parts of Greater London (as captured by β_τ), which remain of around the same magnitude as in Column (1). However, we find substantially and statistically significantly smaller treatment effects of the railway on parish population for the City of London (as reflected in large negative and statistically significant estimates of γ_τ). Furthermore, the estimated γ_τ are larger in absolute magnitude than the estimated β_τ , implying an overall negative treatment effect of the railway on the population of parishes in the City of London ($\beta_\tau + \gamma_\tau$), which is statistically significant at conventional critical values.

In Column (6), we augment this specification with parish time trends to allow parishes that are treated with a railway to have different trend rates of population growth in all years (both before and after the railway). Even in this specification, where we identify the railway treatment effect solely from deviations from parish time trends, we find the same pattern of negative and statistically significant treatment effects for parishes in the City of London and positive and statistically significant treatment effects for parishes in other parts of Greater London.

In Column (7), we again check whether the timing of these deviations from parish time trends coincides with the arrival of the railway. We augment the specification in Column (6) by including interactions with treatment years before the arrival of the railway for both sets of coefficients (β_τ and γ_τ). The excluded category is again the treatment year ($\tau = 0$), and we again consider symmetric time windows before and after the arrival of the railway. We find no evidence of statistically significant deviations from the parish time trends before the arrival of the railway, whether for the City of London ($\beta_\tau + \gamma_\tau$) or for other parts of Greater London (β_τ). However, we continue to find large and statistically significant deviations from the parish time trends in the years after the arrival of the railway, which are negative for the City of London ($\beta_\tau + \gamma_\tau < 0$) and positive for other parts of Greater London ($\beta_\tau > 0$). This specification provides strong support for a causal interpretation of the effects of the railway in reducing population in the City of London and raising it in other parts of Greater London. Indeed, it is hard to think of confounding factors that are timed to coincide precisely with the arrival of the railway, and are structured to have exactly the same pattern of opposite effects on population in the City of London and other parts of Greater London.

4.3 Non-parametric Specification

To provide further evidence on the heterogeneity in railway treatment effects, we now report a non-parametric specification, in which we estimate a separate railway treatment for each parish. We show that the difference in estimated treatment effects between the City of London and the rest of Greater London in the previous subsection reflects a more general pattern, in which the estimated railway treatment varies systematically with distance from the center of Greater London. As in the previous subsection, we consider a “difference-in-differences” specification, in which the first difference is across parishes, and the second difference is across time.

In a first step, we compute the relative population of parishes, by differencing the log population for each parish

³⁰Of the 1,221 parish-year observations for the City of London, only 154 of these observations involve parishes that have a railway station in at one census year during the sample period. The distribution of these 154 observations across the treatment years is $\tau \leq -30$ (73); $\tau \leq -20$ (14); $\tau \leq -10$ (14); $\tau = 0$ (14); $\tau = 10$ (14); $\tau = 20$ (11) and $\tau \geq 30$ (14).

in each year from the mean across parishes in that year:

$$\ln \tilde{H}_{\ell t} = \ln H_{\ell t} - \frac{1}{N} \sum_{\ell=1}^N \ln H_{\ell t}, \quad (3)$$

where $\tilde{H}_{\ell t}$ denotes relative population and N is the number of parishes. By differencing from mean population in each year, we remove any secular trend in population across all Greater London parishes over time, which allows us to control for the fact that different parishes are treated with the railway in different census years.

In a second step, we compute the growth in the relative population of each parish over the thirty-year period before the arrival of the railway (from $\tau = -30$ to $\tau = 0$):

$$\Delta \ln \tilde{H}_{\ell}^{\text{pre}} = \ln \tilde{H}_{\ell, \tau=0} - \ln \tilde{H}_{\ell, \tau=-30}, \quad (4)$$

where the difference over time differences out any fixed effect in the level of log relative parish population. We focus on a narrow thirty-year window to ensure a similar time interval over which population growth is computed for all parishes. We cannot compute this difference in equation (4) for parishes that are never treated with the railway, and hence drop these parishes. All other parishes have at least thirty years before the arrival of the railway, because our sample begins in 1801, and the first railway in Greater London is built in 1836.

In a third step, we compute the growth in the relative population of each parish over the thirty-year period after the arrival of the railway (from $\tau = 0$ to $\tau = 30$):

$$\Delta \ln \tilde{H}_{\ell}^{\text{post}} = \ln \tilde{H}_{\ell, \tau=30} - \ln \tilde{H}_{\ell, \tau=0}, \quad (5)$$

where the difference over time again differences out any fixed effect in the level of log relative parish population. We again focus on a narrow thirty-year window. We drop any parish with less than thirty years between its treatment year and the end of our parish-level sample in 1901.

In a fourth and final step, we compute the “difference-in-difference,” namely the change in each parish’s growth in relative population between the thirty-year periods before and after the arrival of the railway.

$$\Delta \ln \tilde{\tilde{H}}_{\ell} = \Delta \ln \tilde{H}_{\ell}^{\text{post}} - \Delta \ln \tilde{H}_{\ell}^{\text{pre}}, \quad (6)$$

where the double tilde indicates that this is a “difference-in-difference.” By taking the difference between the growth rates before and after the arrival of the railway, we difference out any parish time trend that is common to these two periods. Therefore, we again focus on deviations from parish trends, as in the previous subsection.

In Figure 11, we display these double differences in relative population growth for each parish against the straight-line distance from its centroid to the Guildhall in the City of London. We indicate parishes in the City of London by hollow red circles, while parishes in the other parts of Greater London are denoted by solid blue circles. We also show the locally-weighted linear least squares regression relationship between the two variables as the solid black line. We find a sharp non-linear relationship between the railway treatment and distance from the Guildhall. Given that relative log population is measured as a difference from its average value, the average treatment effect across all parishes is equal to zero. However, for parishes within five kilometers of the Guildhall, we find negative average estimated treatment effects (an average of -0.56 log points), particularly for those parishes inside the City of London. In contrast, for parishes beyond five kilometers from the Guildhall, we find positive average estimated treatment effects

(an average of 0.19 log points), where these substantial differences between the two groups are statistically significant at conventional critical values.

Therefore, in this non-parametric specification that allows for heterogeneous treatment effects across parishes, we again find evidence of a systematic reorganization of economic activity. We find that the arrival of the railway reduces relative population growth in parishes close to the commercial center of Greater London, and increases relative population growth in parishes further from the commercial center of Greater London.

4.4 Mechanisms

We now provide further direct evidence that the changes in night and day population in the previous subsections involve changes in commuting behavior. Although the 1921 population census is the first to report systematic information on bilateral commuting patterns, we can track historical residence and workplace addresses for selected professions from post office directories. We focus on barristers (a type of lawyer) for which data on a large number of individuals were available for the years 1841, 1852, 1882, 1899 and 1921. As discussed further in the data appendix, we randomly sampled up to 4 barristers from each surname letter A-Z, which yielded a sample of 73-85 barristers in each year. We geocoded each individual's residence and workplace address (taking account of street name changes) and computed the straight-line distance between these addresses. In Figure 8, we display kernel density estimates of the distribution of commuting distances across barristers for each year. As apparent from the figure, we find a marked shift in the distribution of commuting distances between 1841-52 and 1882-1921. Whereas the median commuting distance is less than 3 kilometers for 1841-52, it rises to more than 5 kilometers for 1882-1921. This timing of the marked shift in commuting distances lines up well with the timing of the expansion of the railway network in the County of London in Figures 2- 4 and the sharp drop in population in the City of London in Figure 5. In Figure 9, we provide further evidence on a change in transport use by graphing passenger journeys using public transport per head of population in the County of London over time (see also Barker 1980). Public transport includes underground rail, overground rail, short-stage coach, omnibus and tram. As shown in the figure, the increasing specialization of locations as workplace or residence is reflected in an increase in the intensity of public transport use.

Finally, in Figure 10, we provide evidence on the specialization of boroughs as workplace or residence locations at the end of our time period in 1921. We display each borough's share in total employment in the County of London, for both workplace employment and residence employment separately. We find that workplace employment is substantially more spatially concentrated than residence employment. The City of London stands out as the borough that is most specialized as a workplace, accounting for more than 15 percent of total employment in the County of London, and having by far the largest ratio of employment to residents. The City of Westminster is the next most specialized workplace. Boroughs that are the most specialized residences include Islington, Lambeth and Wandsworth, which are part of an inner ring of suburbs surrounding the Cities of London and Westminster.

5 Theoretical Model

We now develop our theoretical framework to explain the above changes in the spatial organization of economic activity. We consider a city (Greater London) embedded within a wider economy (the United Kingdom). The city consists of a discrete set of locations R (the boroughs observed in our data). Workers are geographically mobile and

choose between the city and the wider economy. Population mobility implies that the expected utility from living and working in the city equals the reservation level of utility in the wider economy \bar{U} . If a worker chooses the city, she chooses a residence n and a workplace i from the set of locations $n, i \in R$ to maximize her utility.³¹ We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, as discussed further below.³²

5.1 Preferences

Worker preferences are defined over consumption of a composite final good and residential floor space. The indirect utility function is assumed to take the Cobb-Douglas form such that utility for a worker ω residing in n and working in i is given by:³³

$$U_{ni}(\omega) = \frac{z_{ni}(\omega)w_i}{\kappa_{ni}P_n^\alpha Q_n^{1-\alpha}}, \quad 0 < \alpha < 1, \quad (7)$$

where P_n is the price of the composite final good, Q_n is the price of floor space, w_i is the wage, κ_{ni} is an iceberg commuting cost, and $z_{ni}(\omega)$ is an idiosyncratic amenity draw that captures all the idiosyncratic factors that can cause an individual to live and work in particular locations within the city.³⁴

We observe positive residents, positive employment and a single rateable value for each borough in our data. Therefore, we assume that all boroughs are incompletely specialized in commercial and residential activity, and that no-arbitrage ensures a common price of floor space for residential and commercial use (Q_n). We assume that the composite final good is costlessly tradeable and choose it as our numeraire ($P_n = 1$ for all $n \in R$). As discussed further below, this composite final good is produced using labor, non-traded services and floor space. All floor space is owned by absentee landlords, who receive payments from the residential and commercial use of floor space, and consume only the composite final good.

We assume that idiosyncratic amenities ($z_{ni}(\omega)$) are drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and each worker:

$$G_{ni}(z) = e^{-B_n z^{-\epsilon}}, \quad B_n > 0, \quad \epsilon > 1, \quad (8)$$

where B_n determines average residential amenities in location n . Therefore, we allow some locations to be more attractive in terms of their residential amenities than others (e.g. leafy streets and scenic views). In principle, these differences in average amenities (as determined by B_n) could be either exogenous or endogenously determined by agglomeration forces. We explore both these cases in our quantitative analysis of the model below. The Fréchet shape parameter ϵ determines the dispersion of idiosyncratic amenities, which controls the sensitivity of worker location decisions to economic variables (e.g. wages and the cost of living). The smaller is ϵ , the greater is the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables.

Conditional on choosing to live in Greater London, equations (7) and (8) imply that the probability a worker

³¹Motivated by our empirical finding above that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs between Greater London and the wider economy. Therefore, a worker cannot live in the city and work in the wider economy or vice versa.

³²To ease the exposition, we typically use n for residence and i for workplace, except where otherwise indicated.

³³For empirical evidence using U.S. data in support of the constant housing expenditure share implied by the Cobb-Douglas functional form, see [Davis and Ortalo-Magné \(2011\)](#).

³⁴Although we model commuting costs in terms of utility, they enter the indirect utility function (7) multiplicatively with the wage, which implies that there is a closely-related formulation in terms of the opportunity cost of time spent commuting.

chooses to reside in n and work in i is

$$\pi_{ni} \equiv \frac{H_{ni}}{H} = \frac{B_n w_i^\epsilon (\kappa_{ni} Q_n^{1-\alpha})^{-\epsilon}}{\sum_{r \in R} \sum_{s \in R} B_r w_s^\epsilon (\kappa_{rs} Q_r^{1-\alpha})^{-\epsilon}}, \quad (9)$$

where H_{ni} is the measure of commuters from n to i and H is total city employment (which equals total city residents). Summing across workplaces, we obtain the probability that an individual lives in each location (π_n^R), while summing across residences, we arrive at the probability that an individual works in each location (π_n^M):

$$\pi_n^R = \frac{H_n^R}{H} = \frac{\sum_{s \in R} B_n w_s^\epsilon (\kappa_{ns} Q_n^{1-\alpha})^{-\epsilon}}{\sum_{r \in R} \sum_{s \in R} B_r w_s^\epsilon (\kappa_{rs} Q_r^{1-\alpha})^{-\epsilon}}, \quad \pi_n^M = \frac{H_n^M}{H} = \frac{\sum_{r \in R} B_r w_i^\epsilon (\kappa_{ri} Q_r^{1-\alpha})^{-\epsilon}}{\sum_{r \in R} \sum_{s \in R} B_r w_s^\epsilon (\kappa_{rs} Q_r^{1-\alpha})^{-\epsilon}}, \quad (10)$$

where H_n^R is the measure of residents and H_n^M is the measure of employment. The Fréchet distribution for idiosyncratic amenities implies that expected utility is equalized across pairs of residence and workplace within Greater London and equal to the reservation level of utility in the wider economy

$$\bar{U} = \delta \left[\sum_{r \in R} \sum_{s \in R} B_r w_s^\epsilon (\kappa_{rs} Q_r^{1-\alpha})^{-\epsilon} \right]^{\frac{1}{\epsilon}}, \quad (11)$$

where $\delta = \Gamma((\epsilon - 1)/\epsilon)$; $\Gamma(\cdot)$ is the Gamma function; and we have used our choice of numeraire ($P_n = 1$).

5.2 Production Technology

The composite final good is produced under conditions of perfect competition using labor, non-traded services and floor space.³⁵ The final goods production technology is assumed to take the Cobb-Douglas form with unit cost:

$$1 = \frac{1}{A_i^F} w_i^\beta p_i^\gamma Q_i^{1-\beta-\gamma}, \quad 0 < \beta, \gamma < 1, \quad \beta + \gamma = 1, \quad (12)$$

where A_i^F is final goods productivity; p_i is the price of non-traded services in location i ; and we have used our choice of numeraire ($P_n = 1$).

We assume that non-traded services are produced using labor and floor space under conditions of perfect competition. Again we assume that the production technology takes the Cobb-Douglas form with unit cost:

$$p_i = \frac{1}{A_i^I} w_i^\mu Q_i^{1-\mu}, \quad 0 < \mu < 1, \quad (13)$$

where A_i^I is non-traded services productivity in location i .

Using the non-traded services production technology (13), the unit cost function for the final good can be rewritten in the following form:

$$1 = \frac{1}{A_i} w_i^{\tilde{\beta}} Q_i^{1-\tilde{\beta}}, \quad A_i \equiv A_i^F (A_i^I)^\gamma, \quad (14)$$

$$\tilde{\beta} \equiv \beta + \gamma\mu, \quad 1 - \tilde{\beta} = (1 - \beta - \gamma) + \gamma(1 - \mu) = 1 - (\beta + \gamma\mu), \quad 0 < \tilde{\beta} < 1,$$

where $\tilde{\beta}$ is a composite measure of labor intensity for the final goods and non-traded services sectors as a whole and A_i is a composite measure of productivity. Again, this composite productivity measure (A_i) could be either exogenous or endogenously determined by agglomeration forces. We explore both these possibilities in our quantitative analysis of the model below.

³⁵London had substantial employment in both industry and services during our sample period. It was one of the main industrial centers in the United Kingdom, with manufacturing accounting for over 25 percent of employment in Greater London in the population census of 1911. In the model, we interpret employment in services as a non-traded input into the production of the final consumption good.

Re-arranging equation (14), we obtain the following key implication of profit maximization and zero profits for each location with positive production:

$$w_i = A_i^{1/\tilde{\beta}} Q_i^{-(1-\tilde{\beta})/\tilde{\beta}}. \quad (15)$$

Intuitively, the maximum wage (w_i) that a location can afford to pay workers is increasing in the location's composite productivity (A_i) and decreasing in the price of floor space (Q_i). We use this relationship in our empirical analysis to solve out for the equilibrium wage as a function of composite productivity and the price of floor space.

5.3 Market Clearing

Commuter market clearing implies that total employment in each location (H_i^M) equals the number of workers choosing to commute to that location:

$$H_i^M = \sum_{n \in R} \pi_{ni|n} H_n^R, \quad (16)$$

where total employment (H_n^M) is the sum of final goods employment and non-traded services employment; $\pi_{ni|n}$ is the probability of commuting to workplace i conditional on living in residence n :

$$\pi_{ni|n}^R = \frac{\pi_{ni}}{\pi_n^R} = \frac{(w_i/\kappa_{ni})^\epsilon}{\sum_{s \in R} (w_s/\kappa_{ns})^\epsilon}. \quad (17)$$

Therefore, each location faces an upward-sloping supply curve for workers that increases with its wage relative to that in other locations, and decreases with its commuting costs relative to those in other locations. With a continuous measure of workers and residents, there is no uncertainty in the supply of workers to each location.

Land market clearing implies that total income from floor space equals the sum of payments for floor space for residential and commercial use:

$$\mathbb{Q}_n = Q_n L_n = (1 - \alpha) v_n H_n^R + \left(\frac{1 - \tilde{\beta}}{\tilde{\beta}} \right) w_n H_n^M, \quad (18)$$

where $\mathbb{Q}_n = Q_n L_n$ is the total value of floor space (which corresponds to rateable value in our data); L_n is the supply of floor space; and v_n is average residential income. The supply of floor space in each location is determined by geographical land area (K_n) and density of development (φ_n):

$$L_n = \varphi_n K_n. \quad (19)$$

We consider both the case in which the density of development (φ_n) is exogenous and the case in which it is endogenous to the surrounding concentration of economic activity. Average residential income (v_n) is a weighted average of the wages in all locations where the weights are given by the conditional commuting probabilities:

$$v_n = \sum_{i \in R} \pi_{ni|n}^R w_i. \quad (20)$$

5.4 General Equilibrium

We begin by characterizing the properties of a benchmark version of the model in which productivity, amenities, commuting costs and the supply of floor space are exogenous. Given the model's parameters $\{\alpha, \tilde{\beta}, \epsilon\}$, the reservation level of utility in the wider economy \bar{U} , and vectors of exogenous location characteristics $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\kappa}, \mathbf{L}\}$, the general equilibrium of the model is referenced by four vectors $\{\boldsymbol{\pi}^M, \boldsymbol{\pi}^R, \mathbf{Q}, \mathbf{w}\}$ and total city population (H), where we

indicate vectors or matrices using bold math font. These five components of the equilibrium vector are determined by the following system of five equations: the workplace choice probabilities (π^M in equation (10)), the residential choice probabilities (π^R in equation (10)), land market clearing (18), profit maximization and zero profits (15), and population mobility (11).

Proposition 1 *Assuming exogenous, finite and strictly positive location characteristics ($A_n \in (0, \infty)$, $B_n \in (0, \infty)$, $\kappa_{ni} \in (0, \infty) \times (0, \infty)$, $L_n \in (0, \infty)$), there exists a unique general equilibrium vector $\{\pi^M, \pi^R, \mathbf{Q}, \mathbf{w}, H\}$.*

Proof. See the web appendix. ■

In this case of exogenous location characteristics, there are no agglomeration forces, and hence the model's congestion forces of commuting costs and an inelastic supply of land ensure the existence of a unique equilibrium. The assumption that productivity (A_n), amenities (B_n) and commuting costs (κ_{ni}) are finite and strictly positive ensures that all locations have positive employment and residents, because the support of the Fréchet distribution for idiosyncratic amenities is unbounded from above. Therefore, there is always a positive measure of workers that choose to live and work in each pair of residence and employment locations for positive and finite values of productivity, amenities and commuting costs. To the extent that we observe zero commuting flows in the data for some pairs of locations, we interpret them as corresponding in the model to the case in which commuting costs become arbitrary large and the measure of commuters becomes arbitrarily small.

In contrast, if productivity, amenities, commuting costs and the supply of floor space $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\kappa}, \mathbf{L}\}$ are endogenous, this creates the possibility of multiple equilibria, depending the strength of agglomeration forces relative to the exogenous differences in characteristics across locations. As we show below, an important feature of our quantitative analysis of the model is that there is a one-to-one mapping from the observed data and model parameters to the unobserved location characteristics $\{\mathbf{A}, \mathbf{B}, \boldsymbol{\kappa}, \mathbf{L}\}$. This invertibility property of the model holds regardless of whether these location characteristics are exogenous or endogenous, and regardless of whether the model has a single equilibrium or multiple equilibrium. Intuitively, we observe an equilibrium in the data, and the observed values of the endogenous variables in this equilibrium, together with the structure of the model, contain enough information to recover the unobserved location characteristics that support this observed equilibrium (regardless of whether or not there could have been another equilibrium for the same parameter values).

6 Quantitative Analysis

We now show how the model can be used to generate predictions for the removal of the railway network, starting from our baseline year of 1921, when we observe bilateral commuting flows between each pair of boroughs. Beginning from this initial equilibrium and using changes in rateable values and residence employment going backwards in time, we use the structure of the model to infer missing data for earlier years on workplace employment. In overidentification checks, we show that the model provides a good approximation to the historical data on workplace employment that are available for these earlier years for the City of London. We also show how the model can be used to decompose the observed changes in the spatial organization of economic activity within Greater London into the contributions of changes in commuting costs, floor space, productivity, and amenities. We use the recursive structure of the model to

undertake this quantitative analysis in a number of steps. Each step involves the minimal set of assumptions, before making additional assumptions to move to the next step.

6.1 Commuting and Employment (Step 1)

In our first step, we simply use the observed data on bilateral commuting flows (H_{nit}) from the population census in our baseline year $t = 1921$ to directly compute the following variables in that baseline year: total city employment,

$$H_t = \sum_{n \in R} \sum_{i \in R} H_{nit}, \quad (21)$$

the unconditional commuting probability (π_{nit}),

$$\pi_{nit} = \frac{H_{nit}}{H_t}, \quad (22)$$

workplace employment (H_{it}^M) and residence employment (H_{nt}^R),

$$H_{it}^M = \sum_{n \in R} H_{nit} \quad H_{nt}^R = \sum_{i \in R} H_{nit}, \quad (23)$$

and the conditional probability of commuting to workplace i conditional on living in residence n ($\pi_{nit|n}$),

$$\pi_{nit|n} = \frac{H_{nit}}{H_{nt}^R}. \quad (24)$$

6.2 Wages and Expected Income in the Initial Equilibrium (Step 2)

In our second step, we solve for wages (w_{nt}) and expected residential income (v_{nt}) in the initial equilibrium in $t = 1921$ using the observed workplace employment (H_{nt}^M), residence employment (H_{nt}^R) and rateable values ($\mathbb{Q}_{nt} = Q_{nt}L_{nt}$). We assume central values for the utility and production function parameters. In particular, we set the share of consumer expenditure on residential land ($1 - \alpha$) equal to 0.25, which is consistent with [Davis and Ortalo-Magné \(2011\)](#). We assume that the share of expenditure on commercial land for the composite production sector composed of the final good and non-traded services ($1 - \tilde{\beta}$) equal to 0.20, which is line with [Valentinyi and Herrendorf \(2008\)](#).

Given these parameters, we use equation (20) to substitute for expected residential income (v_{nt}) in the land market clearing condition (18), and obtain the following system of equations:

$$\mathbb{Q}_{nt} = (1 - \alpha) \left[\sum_{i \in R} \pi_{nit|n} w_{it} \right] H_{nt}^R + \left(\frac{1 - \tilde{\beta}}{\tilde{\beta}} \right) w_{nt} H_{nt}^M. \quad (25)$$

which determines a unique wage in each location (w_{nt}) in the initial equilibrium, given the observed data on workplace employment (H_{nt}^M), residence employment (H_{nt}^R) and rateable values (\mathbb{Q}_{nt}). Intuitively, there is a unique wage vector that rationalizes the observed rateable values given the observed workplace and residence employment and the assumption that factor payments are a constant share of consumer expenditure and firm revenue.

Using these solutions for wages (w_{nt}), and the observed conditional commuting probabilities ($\pi_{nit|n}$), we immediately recover expected residential income (v_{nt}) from equation (20):

$$v_{nt} = \sum_{i \in R} \pi_{nit|n} w_{it}. \quad (26)$$

Note that changes in the units in which rateable values are measured lead to proportionate changes in wages and expected residential income. From equations (25) and (26), the solutions for wages and expected residential income

are homogeneous of degree one in the units in which rateable values are measured, given the observed employment and commuting data $\{\pi_{nit|n}, H_{nt}^R, H_{nt}^M\}$. Therefore, changes in these units over time (e.g. with inflation as our rateable values are measured in current prices) lead to proportionate changes in wages and expected residential income.

This second step involves making assumptions about the parameters of the utility and production technologies, but does not impose any restrictions on commuting costs or the determinants of productivity, amenities or land supplies. Commuting costs are implicitly captured in observed commuting flows, while productivity, amenities and land suppliers are implicitly captured in observed rateable values. Together, commuting flows and rateable values are sufficient statistics to infer wages given the assumed parameters of the utility and production functions.

6.3 Changes in Commuting Costs (Step 3)

In our third step, we construct measures of the change in commuting costs implied by the evolution of the railway network over time. We use a measure of commuting costs based on least-cost-paths following [Donaldson \(2017\)](#) and [Allen and Arkolakis \(2014\)](#). In particular, we discretize Greater London into a raster of grid points and assign a cost or weight to traveling across each grid point that depends on whether or not it is on a railway.³⁶ We normalize the cost of traveling over points on a railway to one ($\delta^{\text{Rail}} = 1$). We assume a cost of traveling over points not on a railway of $\delta^{\text{Land}} = 7$, which is based on assumed average travel speeds by foot and rail of 3 and 21 mph respectively.³⁷ Given these assumed parameters, we measure the travel cost between the centroids of a pair of boroughs within Greater London as the sum of the costs of traveling across the intermediate points along the least-cost path between those centroids. This travel cost changes over time with the set of points connected to the railway network. Denoting the vector of travel costs by $\delta = [1 \quad \delta^{\text{Land}}]$ and the set of points connected to the railway network at time t by $\mathfrak{F}_t^{\text{Rail}}$, we can write the bilateral travel cost between the centroids of boroughs n and i at time t as $\kappa_{nit}(\mathfrak{F}_t^{\text{Rail}}, \delta)$.

We compute this measure of bilateral commuting costs between each pair of boroughs for each census year, using the observed railway network in that year. As a check on these measures, we compare them to our data on bilateral commuting flows for 1921. In particular, we examine the conditional correlation between bilateral commuting flows and commuting costs, after controlling for workplace and residence fixed effects. First, we regress the log of bilateral commuting flows for 1921 on these fixed effects, and generate the residuals. Next, we regress the log of bilateral commuting costs for that year on the same fixed effects, and generate the residuals. In [Figure 12](#), we display these two residuals against one another, as well as the linear regression fit between them, for all pairs $n \neq i$ with positive commuting flows. Inevitably, our parsimonious commuting cost measure abstracts from the many idiosyncratic factors that can affect these costs for individual pairs. Furthermore, the slope of this regression relationship does not have a structural interpretation, because bilateral commuting costs depend on the railway network, which is in general endogenous. Nonetheless, we find that our measures of bilateral commuting costs using our assumed parameters exhibit the log linear relationship with bilateral commuting flows predicted by our model. This relationship is strong and statistically significant, with the regression R-squared for the conditional correlation equal to 0.77.

Given these bilateral commuting cost measures for each pair of boroughs for each census year, we can compute the proportional change in bilateral commuting costs between our baseline year of $t = 1921$ and any previous year

³⁶The density of railway stations in London is high relative to the size of the boroughs used in our structural estimation of the model. Therefore, for simplicity, we abstract from the role of railway stations as points of interchange with the railway network.

³⁷These assumed costs based on average travel speeds are close to the estimates for rail and land transport in [Donaldson \(2017\)](#).

$\tau < t$ from taking the railway network backwards in time:

$$\hat{\kappa}_{nit,\tau} = \frac{\kappa_{nit}(\mathfrak{F}_\tau^{\text{Rail}}, \delta)}{\kappa_{nit}(\mathfrak{F}_{1921}^{\text{Rail}}, \delta)}, \quad (27)$$

where we use a hat above a variable to denote a relative value, such that $\hat{x}_t = x_\tau/x_t$. The only reason that bilateral commuting costs change over time is the railway network (in general $\mathfrak{F}_\tau^{\text{Rail}} \neq \mathfrak{F}_{1921}^{\text{Rail}}$ for $\tau \neq 1921$). As this railway network is strictly smaller in any earlier year $\tau < t$, bilateral commuting costs weakly increase going backwards in time, such that $\hat{\kappa}_{nit,\tau} \geq 1$. However, the magnitude of this increase varies across pairs of boroughs, depending on the different patterns of railway lines in the two years. We compute these relative changes in bilateral commuting costs for each census year back to 1831 (before the construction of the first railway in Greater London).

Given the observed evolution of the railway network over time, the key additional assumptions used in this third step are that commuting costs can be modeled as the solution to a least-cost path problem, as well as the relative costs of travel by rail and land for each point. The tightness of the relationship between our bilateral commuting cost measures and bilateral commuting flows for 1921 in Figure 12 suggests that these assumptions provide a reasonable approximation to bilateral commuting costs.

6.4 Wages and Workplace Employment in Each Year (Step 4)

In our fourth step, we use the structure of our model to generate predictions for wages and workplace employment in earlier years $\tau < t$ before our baseline year of $t = 1921$, given our bilateral commuting data for our baseline year, our residence employment and rateable values data for earlier years, and our measures of changes in commuting costs. We follow an “exact hat algebra” approach similar to that used in the quantitative international trade literature following [Dekle, Eaton, and Kortum \(2007\)](#). We use the observed values of variables in an initial equilibrium, and the general equilibrium conditions of the model, to generate predictions for the impact of taking the railway network backwards in time, conditional on the observed changes in residence employment and rateable values.

The key parameter determining the elasticity of commuting flows with respect to commuting costs is the Fréchet shape parameter determining the heterogeneity in workers idiosyncratic preferences. We set this Fréchet shape parameter $\epsilon = 3$, which is in line with the estimates using commuting and migration data in [Monte, Redding, and Rossi-Hansberg \(2017\)](#), [Bryan and Morten \(2015\)](#) and [Galle, Yi, and Rodriguez-Clare \(2015\)](#).³⁸ Given this parameter, the land market clearing condition (18) for any earlier year $\tau < t$ can be re-written in terms of the observed variables or model solutions for our baseline year of $t = 1921$ and the relative changes in the endogenous variables of the model between those two years:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1 - \alpha)\hat{v}_{nt}v_{nt}\hat{H}_{nt}^R H_{nt}^R + \left(\frac{1 - \tilde{\beta}}{\tilde{\beta}}\right)\hat{w}_{nt}w_{nt}\hat{H}_{nt}^M H_{nt}^M, \quad (28)$$

where a hat above a variable again denotes a relative change, such that $\hat{x}_t = x_\tau/x_t$.

We observe workplace employment, residence employment and rateable values in our baseline year $\{H_{nt}^M, H_{nt}^R, \mathbb{Q}_{nt}\}$; we have solved for wages and expected residential income in our baseline line $\{w_t, v_t\}$ in step 2; we also observe the relative changes in rateable values $\{\hat{\mathbb{Q}}_n\}$; and we have constructed relative changes in residence employment $\{\hat{H}_{nt}^R\}$ from our population census, as discussed in the data section above. Hence, the land market clearing condition (28)

³⁸Given an instrument for the railway network, the relationship between bilateral commuting flows and travel costs in 1921 shown in Figure 12 could be used to estimate the Fréchet shape parameter and the relative cost of land/rail transport $\{\epsilon, \delta^{\text{land}}\}$, instead of using assumed values of these parameters. In future work, we plan to compare such instrumental variables estimates with our assumed parameter values here.

provides a system of equations for each of the N boroughs in terms of the $3 \times N$ unknown relative changes in wages, expected residential income and workplace employment $\{\hat{w}_n, \hat{v}_n, \hat{H}_{nt}^M\}$. We now show that the relative changes in expected residential income and workplace employment $\{\hat{v}_{nt}, \hat{H}_{nt}^M\}$ can be written in terms of the unknown relative changes in wages $\{\hat{w}_{nt}\}$ and the values of known variables. Substituting these relationships into the land market clearing condition (28), we obtain a system of N equations that determines the equilibrium values of the N unknown relative changes in wages $\{\hat{w}_{nt}\}$, given the measured relative changes in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$.

We start with the relative change in expected residential income (\hat{v}_{nt}), which from equation (20) for any year $\tau < t$ can be re-written as follows:

$$\hat{v}_{nt}v_{nt} = \sum_{i \in R} \frac{\pi_{nit|n} \hat{w}_{it}^\epsilon \hat{\kappa}_{nit,\tau}^{-\epsilon}}{\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^\epsilon \hat{\kappa}_{nst,\tau}^{-\epsilon}} \hat{w}_{it} w_{it}. \quad (29)$$

We observe the conditional commuting probabilities in the initial equilibrium $\{\pi_{nit|n}\}$; we have solved for initial wages and expected residential income $\{w_{nt}, v_{nt}\}$ from step 2; and we have measured the change in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$ from step 3. Therefore the only unknown in equation (29) is the change in wages $\{\hat{w}_{nt}\}$.

We next turn to the relative change in workplace employment (\hat{H}_{nt}^M), which from the commuter market clearing condition (16) for any year $\tau < t$ can be re-written as follows:

$$\hat{H}_{it}^M H_{it}^M = \sum_{n \in R} \frac{\pi_{nit|n} \hat{w}_{it}^\epsilon \hat{\kappa}_{nit,\tau}^{-\epsilon}}{\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^\epsilon \hat{\kappa}_{nst,\tau}^{-\epsilon}} \hat{H}_{nt}^R H_{nt}^R. \quad (30)$$

Again we observe the conditional commuting probabilities in the initial equilibrium $\{\pi_{nit|n}\}$; we observe the initial values and changes of residence employment $\{H_{nt}^R, \hat{H}_{nt}^R\}$; we have solved for initial wages $\{w_{nt}\}$ from step 2; and we have measured the change in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$ from step 3. Therefore the only unknown in equation (30) is the change in wages $\{\hat{w}_{nt}\}$.

Substituting equations (29) and (30) in the land market clearing condition (28), we obtain our system of N equations for each borough that determines the N changes in wages $\{\hat{w}_{nt}\}$ consistent with the equilibrium conditions of the model, given the measured changes in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$ and the observed changes in residence employment and rateable values $\{\hat{H}_{nt}^R, \hat{Q}_{nt}\}$. Using these solutions for changes in wages $\{\hat{w}_{nt}\}$, we can immediately recover the implied changes in workplace employment $\{\hat{H}_{it}^M\}$ from the commuter market clearing condition (30). As this equation is homogeneous of degree zero in wages, the model's predictions for changes in workplace employment are invariant to inflation in wages over time. Using these model predictions and the observed data, we can also recover all of the other unobserved endogenous variables of the model $\{\pi_{nit|n}, \pi_{ni\tau}, H_{n\tau}^M, w_{n\tau}, v_{n\tau}\}$ in the earlier year from the equilibrium conditions of the model, including the implied historical bilateral commuting flows between boroughs.

Remarkably, these model predictions for reversing the construction of the railway network require no assumptions about the strength of agglomeration economies in productivity $\{A_{nt}\}$ or amenities $\{B_{nt}\}$ or about the extent to which the density of development $\{\varphi_{nt}\}$ is endogenous to changes in the spatial organization of economic activity. The reason is that these model predictions condition on the observed changes in rateable values $\{\hat{Q}_{nt}\}$ and residence employment $\{\hat{H}_{nt}^R\}$ going backwards in time. Given the equilibrium conditions of the model, these observed changes backwards in time $\{\hat{Q}_{nt}, \hat{H}_{nt}^R\}$, when combined with the observed values of variables in the initial equilibrium $\{\pi_{nit|n}, H_{nt}^M, H_{nt}^R, Q_{nt}, w_{nt}, v_{nt}\}$, are sufficient statistics to recover the changes in wages $\{\hat{w}_{nt}\}$ and workplace employment $\{\hat{H}_{nt}^M\}$ implied by the measured changes in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$.

Implicitly, the observed bilateral conditional commuting probabilities in the initial equilibrium $\{\pi_{nit|n}\}$ in equa-

tions (29) and (30) capture the bilateral pattern of commuting costs in this initial equilibrium. Similarly, the known rateable values, workplace employment, residence employment, wages and expected residential income in the initial equilibrium $\{H_{nt}^M, H_{nt}^R, Q_{nt}, w_{nt}, v_{nt}\}$ reflect productivity, amenities, the supply of floor space in this equilibrium $\{A_{nt}, B_{nt}, L_{nt}\}$. Finally, the observed changes in rateable values and residence employment $\{\hat{Q}_{nt}, \hat{H}_{nt}^R\}$ contain information about changes in productivity, amenities, the supply of floor space $\{\hat{A}_{nt}, \hat{B}_{nt}, \hat{L}_{nt}\}$ going backwards in time. However, we do not require knowledge of these unobserved changes in productivity, amenities, and the supply of floor space $\{\hat{A}_{nt}, \hat{B}_{nt}, \hat{L}_{nt}\}$. Given the measured changes in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$ and the equilibrium conditions of the model, the observed variables $\{\pi_{nit|n}, H_{nt}^M, H_{nt}^R, Q_{nt}, w_{nt}, v_{nt}, \hat{w}_{nt}, \hat{H}_{nt}^M\}$ are enough by themselves to determine the implied changes in wages and workplace employment $\{\hat{w}_{nt}, \hat{H}_{nt}^M\}$.

In generating these model predictions, we do not use any information about workplace employment for years prior to our baseline year of 1921. Therefore, as an overidentification check, we can compare the model's predictions for the change in workplace employment $\{\hat{H}_{nt}^M\}$ to the historical data on workplace employment in the City of London that are available from the Day Censuses. In Figure 13, we display the results of this overidentification check. We use the solid blue line (without markers) to show the measures of residence employment that we construct from the population census data, as discussed in the data section above. As our quantitative analysis conditions on the observed changes in residence employment in each borough, the model's predictions and the data necessarily coincide in each year for residence employment. Furthermore, as residence employment is constructed from the population data, we find the same pattern in Figure 13 as in Figure 6, with residence employment in the City of London declining sharply from the mid-19th century onwards. We use the solid black line (with triangle markers) to indicate the observed data on workplace employment (from the population census for 1921 and from the Day Census for earlier years). Finally, we use the solid green line (with circle markers) to display the model's predictions for workplace employment from the change in commuting costs implied by moving from the railway network in our baseline year of 1921 to that in each census year back to 1831 (before the first railway in Greater London).

As our quantitative analysis also conditions on the observed variables in the initial equilibrium, the model's predictions for workplace employment in 1921 necessarily coincide with the data. However, for years prior to 1921, these two time-series can diverge from one another. Although the model does not perfectly capture the fluctuations in workplace employment in the City of London from one decade to the next, we find that it is strikingly successful in capturing the sharp trend increase in workplace employment from 1866 onwards, with the model's predictions for the change from 1861-1921 close to the observed data. By 1831 before the construction of any railways, the model predicts a workplace employment in the City of London only just above its residence employment in the data (86,689 compared to 79,524), consistent with the idea that most people lived close to where they worked in an era when the main modes of transport were by human or horse power. Therefore, this overidentification check provides strong support for the model's predictions. Despite the parsimony of our model of commuting costs, with just two parameters for the relative cost of rail/land travel (δ^{Land}) and the elasticity of commuting flows to commuting costs (ϵ), the predictions of our model for the removal of the railway network successfully capture the large-scale changes in the concentration of workplace employment in the City of London over this period.

Having validated the model's predictions using our historical data for the City of London, we now examine its predictions for workplace employment for the other boroughs for which such historical data are not available. In Figure 14, we display each borough's share of workplace and residence employment in the County of London in 1831.

Residence employment is constructed from our population census data. Workplace employment is obtained from the model's predictions for taking the railway network back in time, as outlined above. Comparing Figure 14 to our earlier Figure 10, we find that workplace and residence employment are much closer together in 1831 than in 1921. Therefore, boroughs are much less specialized by workplace and residence in the era before the railway, consistent with short commuting distances at that time. Even in 1831, the City of London was by far the largest net importer of workers. However, its net imports of around 7,000 workers at that time are a pale shadow of those of over 350,000 workers in 1921. Whereas the City of Westminster and Holborn are both small net exporters of residents in 1831 in Figure 14, they are both net importers of workers in 1921 in Figure 10. Therefore, as the geographical boundaries of Greater London and its commercial center expanded outwards, the specialization of individual locations within Greater London evolved over time

In summary, although our model is necessarily an abstraction, this fourth step of our quantitative analysis provides strong evidence that it successfully captures the dramatic reorganization of economic activity in Greater London following the invention of the steam railway.

6.5 Adjusted Productivity and Adjusted Residential Amenities (Step 5)

In our fifth step, we use the equilibrium conditions of the model to recover the implied changes in adjusted productivities and amenities $\{\hat{A}_{nt}, \hat{B}_{nt}\}$ that rationalize the observed changes in rateable values and residence employment $\{\hat{Q}_{nt}, \hat{H}_{nt}^R\}$, given our measured changes in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$ from steps 1-4 above. In the model, adjusted productivity $\{\hat{A}_{nt}\}$ is a summary statistic for the economic environment facing firms, which captures both productivity $\{A_{nt}\}$ and the density of development $\{\varphi_{nt}\}$, as defined formally below. Similarly, adjusted amenities $\{\hat{B}_{nt}\}$ is a summary statistic for the residential setting experienced by workers, which captures amenities $\{B_{nt}\}$, the reservation level of utility in the wider economy $\{\bar{U}_t\}$, and the density of development $\{\varphi_{nt}\}$, as also defined formally below. Together, the changes in commuting costs and adjusted productivities and amenities $\{\hat{\kappa}_{nit,\tau}, \hat{A}_{nt}, \hat{B}_{nt}\}$ fully summarize the change in the location characteristics that are required to rationalize the observed changes in the data $\{\hat{Q}_{nt}, \hat{H}_{nt}^R\}$ as an equilibrium of the model.

Using the zero profit condition (15), the equality between rateable values and the product of the price and quantity of floor space ($\mathbb{Q}_{nt} = Q_{nt}L_{nt}$), and the relationship between floor space and the density of development (19), the change in adjusted productivity $\{\hat{A}_{nt}\}$ can be recovered from the changes in wages and rateable values $\{\hat{w}_{nt}, \hat{Q}_{nt}\}$:

$$\hat{A}_{nt} = \hat{w}_{nt}^\beta \hat{Q}_{nt}^{1-\beta}, \quad (31)$$

where the change in adjusted productivity $\{\hat{A}_{nt}\}$ is defined as:

$$\hat{A}_{nt} \equiv \hat{A}_{nt} \hat{\varphi}_{nt}^{1-\beta}. \quad (32)$$

Intuitively, given the observed changes in wages $\{\hat{w}_{nt}\}$ and rateable values $\{\hat{Q}_{nt}\}$, there is a unique change in adjusted productivity $\{\hat{A}_{nt}\}$ that is consistent with firms making zero profits in equilibrium.

Using the residential choice probability (equation (10) for π_{nt}^R), expected utility (11), the equality between rateable values and the product of the price and quantity of floor space ($\mathbb{Q}_{nt} = Q_{nt}L_{nt}$), and the relationship between floor space and the density of development (19), the change in adjusted amenities $\{\hat{B}_{nt}\}$ can be recovered from the changes

in residence employment, rateable values, wages and commuting costs $\{\hat{H}_{nt}^R, \hat{Q}_{nt}, \hat{w}_{nt}, \hat{k}_{nit,\tau}\}$ as follows:

$$\hat{\pi}_{nt}^R = \hat{B}_{nt} \hat{Q}_{nt}^{-\epsilon(1-\alpha)} \left(\hat{M}A_{nt}^R \right)^\epsilon, \quad (33)$$

where the change in adjusted amenities $\{\hat{B}_{nt}\}$ is defined as,

$$\hat{B} \equiv \hat{U}_t^\epsilon \hat{B}_{nt} \hat{\varphi}_{nt}^{\epsilon(1-\alpha)}, \quad (34)$$

and the change in resident market access $(\hat{M}A_{nt}^R)$ summarizes the change in commuting opportunities,

$$\hat{M}A_{nt}^R = \left[\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^\epsilon \hat{k}_{nst,\tau}^{-\epsilon} \right]^{\frac{1}{\epsilon}}. \quad (35)$$

Intuitively, if a location experiences an increase in residents ($\hat{\pi}_{nt}^R > 1$) and/or rateable values ($\hat{Q}_{nt} > 1$) in equation (33), it must be either because it became more attractive in terms of its amenities ($\hat{B}_{nt} > 1$) or experienced an improvement in its resident market access to workplaces ($\hat{M}A_{nt}^R > 1$).

Notably, we are able to solve for the change in adjusted productivity and amenities $\{\hat{A}_{nt}, \hat{B}_{nt}\}$ without making assumptions about what determines productivity $\{A_{nt}\}$, amenities $\{B_{nt}\}$ or floor space $\{\varphi_{nt}\}$. We have thus solved for these variables without having to specify the functional form or strength of agglomeration economies or the elasticity of the supply of floor space. Intuitively, the requirements of utility maximization, profit maximization and market clearing and the structure of the model enable us to determine these sufficient statistics for the economic environment facing firms and workers regardless of how they are determined.

As discussed above, wages scale proportionately with inflation in rateable values, which is relevant for interpreting changes in adjusted productivity and amenities $\{\hat{A}_{nt}\}$, amenities $\{\hat{B}_{nt}\}$. From the zero-profit condition (31), adjusted productivity is a power function of wages and rateable values with exponents that sum to one, which implies that adjusted productivity also scales proportionately with inflation in rateable values. Therefore, the relative levels of adjusted productivity, wages and rateable values are invariant to inflation. From the residential choice probability (33), changes in adjusted amenities also scale with a constant elasticity with inflation in rateable values, and with changes in the reservation level of utility in the wider economy. Nevertheless, relative levels of adjusted amenities across locations within Greater London are invariant to both inflation and changes in the reservation level of utility.

In Figure 15, we show the changes in adjusted productivity in the City of London from equation (31) and their components. The blue line (with no markers) denotes the relative change in rateable values observed in the data; the green line (with circle markers) indicates the relative change in wages in the model; and the black line (with triangle markers) represents the relative change in adjusted productivity in the model. Both rateable values and wages increase between 1831 and 1921, which requires an increase in adjusted productivity for zero profits to be maintained. The increases in rateable values and adjusted productivity track one another closely. This increase in adjusted productivity is consistent with the realization of economies of scale from higher employment density.

In Figure 16, we display the analogous changes in adjusted amenities in the City of London from equation (33) and their components. The blue line (with no markers) again denotes the relative change in rateable values observed in the data; the green line (with circle markers) indicates the relative change in amenities observed in the data; the gray line (with cross markers) represents the relative change in resident market access in the model; and the black line (with triangle markers) corresponds to the relative change in adjusted amenities in the model. We find that the

decline in residents in the City in London from the mid-nineteenth century onwards is largely explained in the model by the increase in rateable values. Resident market access rose over this period, because of the fall in commuting costs and increase in wages, which worked in the opposite direction by making the City of London more attractive as a residential location. Adjusted amenities were relatively flat over the period as a whole, rising somewhat up to 1871, before declining thereafter.

Together, Figures 15 and 16 support the idea that the innovation in transport technology from the invention of the steam railway allowed the City of London to specialize according to its comparative advantage as a workplace rather than as a residence. The resulting competition for space led to higher land values, which induced residents to relocate to other less expensive locations, from where it was now easier to commute to work in the City of London.

6.6 Counterfactuals (Step 6)

We now use our quantitative model to undertake counterfactuals to further explore the role of the steam railway in reshaping the organization of economic activity within Greater London. As workers are geographically mobile in the model, their welfare is pinned down by the reservation level of utility in the wider economy, and hence is unaffected by the construction of the railway network. Therefore, the benefits from the railway network are accrued by landlords through changes in the value of land, which we can compare to estimates of the construction cost of the railway network. We begin by undertaking two different counterfactuals for the effects of the invention of the steam railway, one holding productivity, amenities and the density of development constant at their starting values in 1831 before the first railway, and the other holding them constant at their final values in 1921 at the end of our sample period. These two counterfactuals tell us the extent to which the value of constructing the railway network depends on whether one uses the productivity, amenities and density of development before or after its construction. Making assumptions about the functional form of agglomeration economies and the supply of floor space, we can also undertake a third counterfactual, in which we allow these location characteristics to respond endogenously to the construction of the railway network and the resulting reorganization of economic activity. This third counterfactual tells us about the relative importance of the direct effect of the railway network (though commuting costs) and its indirect effects (though agglomeration forces and the supply of floor space).

In our first two counterfactuals, we start from an initial equilibrium in either 1921 or 1831, and consider a change in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$, holding constant adjusted productivity and amenities $\{\tilde{A}_{nt}, \tilde{B}_{nt}\}$. Therefore, the key difference between these counterfactuals and our quantitative analysis in steps 1-5 above is that adjusted productivity and amenities are held constant here, instead of being allowed to change over time to match the observed changes in rateable values and residence employment. The counterfactual relative change in the endogenous variables of the model $\{\hat{Q}_{nt}, \hat{w}_{nt}, \hat{H}_{nt}^M, \hat{H}_{nt}^R, \hat{H}_t\}$ solves the following system of five equations:

$$\hat{Q}_{nt} Q_{nt} = (1 - \alpha) \left[\sum_{i \in R} \frac{\pi_{nit|n} \hat{w}_{it}^{\epsilon} \hat{\kappa}_{nit,\tau}^{-\epsilon}}{\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^{\epsilon} \hat{\kappa}_{nst,\tau}^{-\epsilon}} \hat{w}_{it} w_{it} \right] \hat{H}_{nt}^R H_{nt}^R + \left(\frac{1 - \tilde{\beta}}{\tilde{\beta}} \right) \hat{w}_{nt} w_{nt} \hat{H}_{nt}^M H_{nt}^M, \quad (36)$$

$$\hat{Q}_{nt} = \hat{w}_{nt}^{-\beta/(1-\beta)}, \quad (37)$$

$$\hat{H}_{it}^M H_{it}^M = \sum_{n \in R} \frac{\pi_{nit|n} \hat{w}_{it}^{\epsilon} \hat{\kappa}_{nit,\tau}^{-\epsilon}}{\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^{\epsilon} \hat{\kappa}_{nst,\tau}^{-\epsilon}} \hat{H}_{nt}^R H_{nt}^R. \quad (38)$$

$$\hat{H}_{nt}^R = \hat{Q}_{nt}^{-\epsilon(1-\alpha)} \left[\sum_{s \in R} \pi_{nst|n} \hat{w}_{st}^{\epsilon} \hat{\kappa}_{nst,\tau}^{-\epsilon} \right] \hat{H}, \quad (39)$$

$$1 = \left[\sum_{r \in R} \sum_{s \in R} \pi_{rst} \hat{w}_{st}^\epsilon \hat{\kappa}_{rst,\tau}^{-\epsilon} \hat{Q}_{rt}^{-\epsilon(1-\alpha)} \right]^{\frac{1}{\epsilon}}. \quad (40)$$

The first of these equations is the land market clearing condition from equation (28), into which we have substituted the relative change in expected residential income from equation (29); the second of these equations is the zero-profit condition (31), in which we have held adjusted productivity constant ($\hat{A}_{nt} = 0$); the third of these equations is the commuter market clearing condition (30); the fourth equation is the residential choice probability (33), in which we have held adjusted amenities constant ($\hat{B}_{nt} = 0$), and into which we have substituted the relative change in resident market access (35); the fifth equation is the population mobility condition (11), in which we have used the fact that the reservation level of utility is constant ($U = \bar{U}$ and hence $\hat{U} = 1$). We have expressed each of these equations in terms of the relative change in rateable values (\hat{Q}_{nt}), which equals the change in relative land values (\hat{Q}_{nt}), because we have held constant the density of development ($\hat{\varphi}_{nt} = 1$ and hence $\hat{L}_{nt} = 1$).

The only unknowns in the system of equations (36)-(40) are the counterfactual relative changes in rateable values, wages, workplace employment, residence employment and total city employment $\{\hat{Q}_{nt}, \hat{w}_{nt}, \hat{H}_{nt}^M, \hat{H}_{nt}^R, \hat{H}_t\}$. For the initial equilibrium in 1921, we observe the commuting probabilities, rateable values, workplace employment and residence employment $\{\pi_{nit|n}, \pi_{ni}, Q_{nt}, H_{nt}^M, H_{nt}^R\}$ in that year; we solve for the wages and expected residential income $\{w_{nt}, v_{nt}\}$ in that year in steps 1-5 above; and we measure the change in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$. For the initial equilibrium in 1831, we observe rateable values and residence employment $\{Q_{nt}, H_{nt}^R\}$ in that year; we solve for the remaining endogenous variables of the model $\{\pi_{nit|n}, \pi_{ni}, H_{nt}^M, w_{nt}, v_{nt}\}$ in that year in steps 1-5 above; and we measure the change in commuting costs $\{\hat{\kappa}_{nit,\tau}\}$.

[XXX Results from these counterfactuals to be completed XXX]

7 Conclusions

We provide evidence on the role of the separation of workplace and residence in understanding the emergence of large metropolitan areas. We use the empirical setting of London’s rapid nineteenth-century population growth, which spans a major change in transport technology, with the invention of steam passenger railways from the late-1830s onwards. The key idea behind our approach is that the slow travel times achievable by human or horse power implied that most people lived near where they worked when these were the main modes of transportation. In contrast, steam passenger railways dramatically reduced the time taken to travel a given distance, thereby permitting the first large-scale separation of workplace and residence.

In the opening decades of the nineteenth-century, rapid population growth for the metropolitan region of Greater London went together with a stagnant population and a declining land value share for its most important commercial center in the City of London. In contrast, following the expansion of the steam passenger railway network during the middle decades of the nineteenth-century, we observe an acceleration of population growth for Greater London as a whole, but a sharp decline in population in the City of London. This precipitous population decline goes hand-in-hand with a steep increase in employment and relative land values in the City of London. To better connect these changes in location specialization to the innovation in transport technology, we provide direct evidence of changes in commuting distances and report the results of reduced-form “difference-in-differences” regression specifications. After allowing for the non-random assignment of railways to locations based on levels and trend rates of growth of

population, we observe substantial increases in population above trend in the years immediately following a railway connection, consistent with a causal effect of railways in raising population. We find that this increase above trend growth is smaller for more central locations than for outlying locations, providing support for the view that railways played a causal role in reshaping the organization of economic activity by workplace and residence.

To further interpret these empirical findings, we develop a quantitative urban model that allows locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections. Our key methodological contribution is to develop a new structural estimation procedure for this class of urban models that feature a gravity equation for commuting flows. Although we only observe these bilateral commuting flows in 1921 at the end of our sample period, we show how this framework can be used to estimate the impact of the construction of the railway network. Combining our 1921 gravity equation data with historical information on population, land values and the transport network back to the early-nineteenth century, we use the model to infer missing employment workplace for each location going backwards in time. Although we do not use any information on employment by workplace before our baseline year in this estimation of the model, we show in overidentification checks that the model provides a good approximation to the evolution of the historical employment by workplace data that are available for the City of London. We find substantial direct effects of the railway through reduced commuting costs, but we also find substantial changes in the relative productivity and amenities of different locations within Greater London, which are consistent with agglomeration forces in production and residential choices.

References

- ABERNATHY, S. T. (2015): “Opening up the suburbs: workmen’s trains in London 1860-1914,” *Urban History*, 42(1), 70–88.
- AHLFELDT, G., S. REDDING, D. STURM, AND N. WOLF (2015): “The Economics of Density: Evidence from the Berlin Wall,” *Econometrica*, 83(6), 2127–2189.
- ALLEN, T., AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129(3), 1085–1140.
- ALLEN, T., C. ARKOLAKIS, AND X. LI (2017): “Optimal City Structure,” Yale University, mimeograph.
- BALL, M., AND D. SUNDERLAND (2001): *An Economic History of London 1800-1914*. Routledge, Abingdon.
- BARKER, T. (1980): “Towards an Historical Classification of Urban Transport Development Since the Later Eighteenth Century,” *Journal of Transport History*, 1(1), 75–90.
- BARKER, T. C., AND M. ROBBINS (1963): *A History of London Transport: The Nineteenth Century*, vol. 1. George Allen & Unwin, London.
- BAUM-SNOW, N. (2007): “Did Highways Cause Suburbanization?,” *Quarterly Journal of Economics*, 122(2), 775–780.
- BAUM-SNOW, N. (2017): “Urban Transport Expansions, Employment Decentralization and the Spatial Scope of Agglomeration Economies,” University of Toronto.

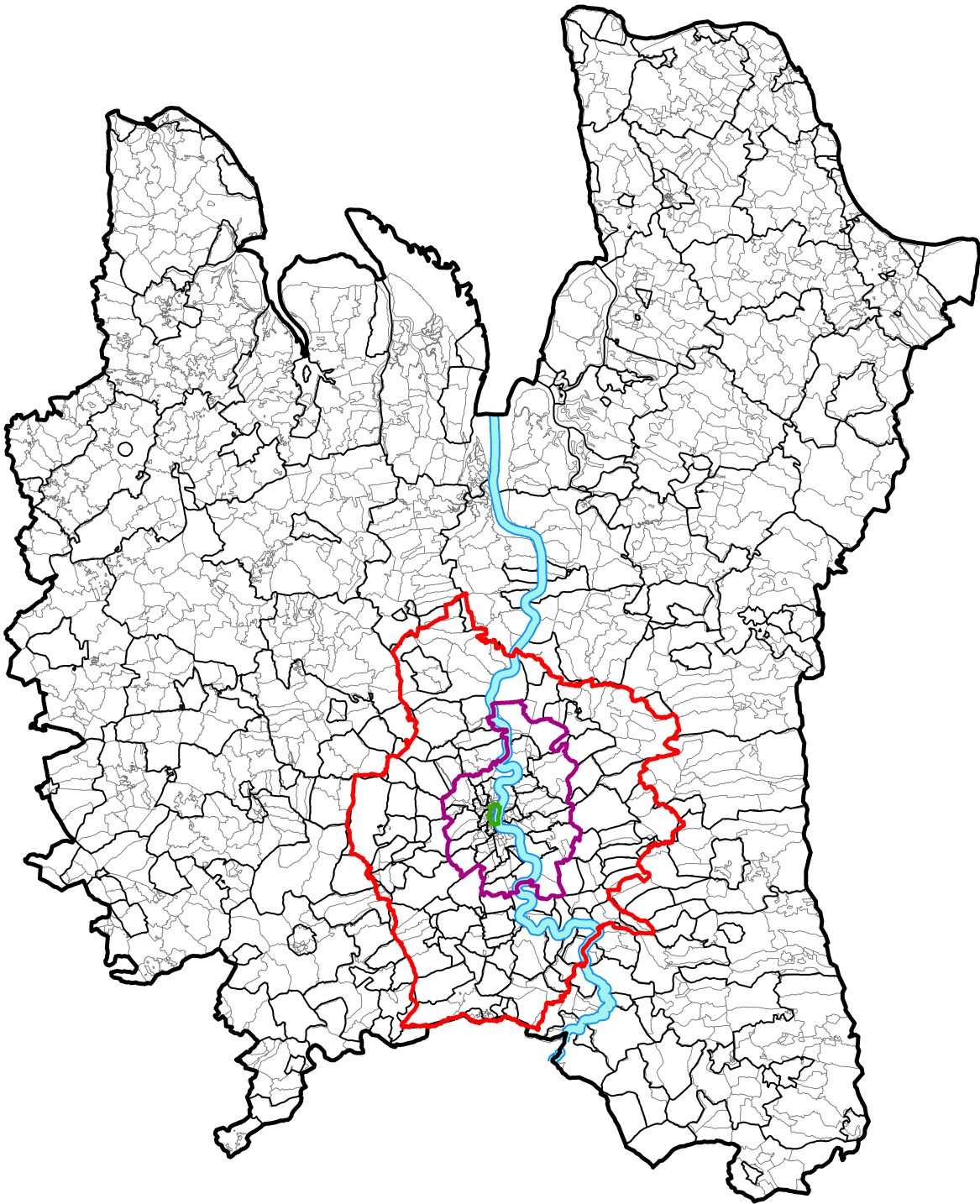
- BAUM-SNOW, N., L. BRANDT, J. HENDERSON, M. A. TURNER, AND Q. ZHANG (2017): “Roads, Railroads and Decentralization of Chinese Cities,” *Review of Economics and Statistics*, forthcoming.
- BAUM-SNOW, N., V. HENDERSON, M. A. TURNER, Q. ZHANG, AND L. BRANDT (2017): “Highways, Market Access and Urban Growth in China,” University of Toronto.
- BAUM-SNOW, N., AND M. E. KAHN (2005): “Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000,” *Brookings-Wharton Papers on Urban Affairs: 2005*, 1(4), 147–197.
- BILLINGS, S. B. (2011): “Estimating the value of a new transit option,” *Regional Science and Urban Economics*, 41(6), 525–536.
- BROOKS, L., AND B. LUTZ (2013): “Vestiges of Transit: Urban Persistence at a Micro Scale,” *Federal Reserve Board of Governors Working Paper*.
- BRYAN, G., AND M. MORTEN (2015): “Economic Development and the Spatial Allocation of Labor: Evidence From Indonesia,” Stanford University, mimeograph.
- CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2017): “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” Princeton University, mimeograph.
- CHANDRA, A., AND E. THOMPSON (2000): “Does public infrastructure affect economic activity? Evidence from the rural interstate highway system,” *Regional Science and Urban Economics*, 30(4), 457–490.
- COMBES, P.-P., G. DURANTON, L. GOBILLON, AND S. ROUX (2010): “Estimating Agglomeration Economies with History, Geology, and Worker Effects,” in *Agglomeration Economies*, ed. by E. L. Glaeser. Chicago University Press, Chicago.
- COMBES, P.-P., AND L. GOBILLON (2015): “The Empirics of Agglomeration Economies,” in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. C. Strange, vol. 5A, chap. 5, pp. 247–348. Elsevier North Holland, Amsterdam.
- CONLEY, T. G. (1999): “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, 92(1), 1–45.
- CORPORATION OF LONDON (1866): *Day Census, City of London, Report, Local Government and Taxation Committee*. Corporation of London, London.
- (1881): *Report on the City Day Census*. Longmann, Green and Company, London.
- CROOME, D. F., AND A. A. JACKSON (1993): *Rails Through the Clay: a History of London’s Tube Railways*. Capital Transport Publishing, London.
- DAVIS, D. R., AND J. DINGEL (2012): “A Spatial Knowledge Economy,” NBER Working Paper, 18188.
- DAVIS, D. R., AND D. E. WEINSTEIN (2002): “Bones, Bombs, and Break Points: The Geography of Economic Activity,” *American Economic Review*, 92(5), 1269–1289.
- DAVIS, M. A., AND F. ORTALO-MAGNÉ (2011): “Housing Expenditures, Wages, Rents,” *Review of Economic Dynamics*, 14(2), 248–261.

- DEKLE, R., J. EATON, AND S. KORTUM (2007): “Unbalanced Trade,” *American Economic Review*, 97(2), 351–355.
- DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2017): “The Geography of Development,” *Journal of Political Economy*, forthcoming.
- DONALDSON, D. (2017): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, forthcoming.
- DURANTON, G., P. MORROW, AND M. A. TURNER (2014): “Roads and Trade: Evidence from the US,” *Review of Economic Studies*, 81(2), 681–724.
- DURANTON, G., AND D. PUGA (2004): “Micro-Foundations of Urban Agglomeration Economies,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J.-F. Thisse, vol. 4, pp. 2063–2117. Elsevier, Amsterdam.
- DURANTON, G., AND M. A. TURNER (2011): “The Fundamental Law of Road Congestion: Evidence from US Cities,” *American Economic Review*, 101(6), 2616–52.
- (2012): “Urban Growth and Transportation,” *Review of Economic Studies*, 79(4), 1407–1440.
- FABER, B. (2014): “Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country,” *Review of Economic Studies*, forthcoming.
- FOLEY, D. L. (1963): *Controlling London’s Growth: Planning the Great Wen*. University of California Press, Berkeley.
- FUJITA, M., P. KRUGMAN, AND A. J. VENABLES (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge MA.
- FUJITA, M., AND J.-F. THISSE (2002): *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*. Cambridge University Press, Cambridge.
- GALLE, S., M. YI, AND A. RODRIGUEZ-CLARE (2015): “Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade,” University of California, Berkeley, mimeograph.
- GERHOLD, D. (2005): *Carriers and Coachmasters: Trade and Travel Before the Turnpikes*. Phillimore, Bodmin, Cornwall.
- GIBBONS, S., AND S. MACHIN (2005): “Valuing Rail Access Using Transport Innovations,” *Journal of Urban Economics*, 57(1), 148–169.
- GONZALEZ-NAVARRO, M., AND M. A. TURNER (2016): “Subways and Urban Growth: Evidence from Earth,” Brown University, mimeograph.
- GREENSTONE, M., R. HORNBECK, AND E. MORETTI (2010): “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 118(3), 536–598.
- HALLIDAY, S. (1999): *The Great Stink of London: Sir Joseph Bazalgette and the Cleansing of the Victorian Metropolis*. Sutton, London.
- HENDERSON, J. V. (1974): “The Sizes and Types of Cities,” *American Economic Review*, 64(4), 640–56.

- JACKSON, K. T. (1987): *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press, Oxford.
- KELLET, J. R. (1969): *The Impact of Railways on Victorian Cities*. Routledge and Kegan Paul, London.
- KLINE, P., AND E. MORETTI (2014): “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *Quarterly Journal of Economics*, 129(1), 275–331.
- KYNASTON, D. (2012): *City of London: The History*. Vintage, London.
- LONDON COUNTY COUNCIL (1907): *London Statistics*. London County Council, London.
- LUCAS, R. E., AND E. ROSSI-HANSBERG (2002): “On the Internal Structure of Cities,” *Econometrica*, 70(4), 1445–1476.
- MCDONALD, J. F., AND C. I. OSUJI (1995): “The Effect of Anticipated Transportation Improvement on Residential Land Values,” *Regional Science and Urban Economics*, 25(3), 261–278.
- McFADDEN, D. (1974): “The Measurement of Urban Travel Demand,” *Journal of Public Economics*, 3(4), 303–328.
- MICHAELS, G. (2008): “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System,” *Review of Economics and Statistics*, 90(4), 683–701.
- MONCKTON, H. P. (1911): *City of London Day Census*. Simkins, Marshall, Hamilton, Kent & Company, London.
- MONTE, F. (2016): “The Local Incidence of Trade Shocks,” Georgetown University, mimeograph.
- MONTE, F., S. REDDING, AND E. ROSSI-HANSBERG (2017): “Commuting, Migration and Local Employment Elasticities,” Princeton University, mimeograph.
- MORETTI, E. (2011): “Local Labor Markets,” in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 4b, pp. 1238–1303. Elsevier North Holland, Amsterdam.
- OWEN, D. (1982): *The Government of Victorian London: The Metropolitan Board of Works, the Vestries and the City Corporation*. Belknap Press, Cambridge MA.
- PORTER, R. (1995): *London: A Social History*. Harvard University Press, Cambridge MA.
- REDDING, S. J. (2016): “Goods Trade, Factor Mobility and Welfare,” *Journal of International Economics*, 101, 148–167.
- REDDING, S. J., AND E. ROSSI-HANSBERG (2017): “Quantitative Spatial Models,” *Annual Review of Economics*, p. forthcoming.
- REDDING, S. J., AND D. M. STURM (2008): “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98(5), 1766–1797.
- REDDING, S. J., AND M. A. TURNER (2015): “Transportation Costs and the Spatial Organization of Economic Activity,” in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. Strange, chap. 20, pp. 1339–1398. Elsevier, Amsterdam.

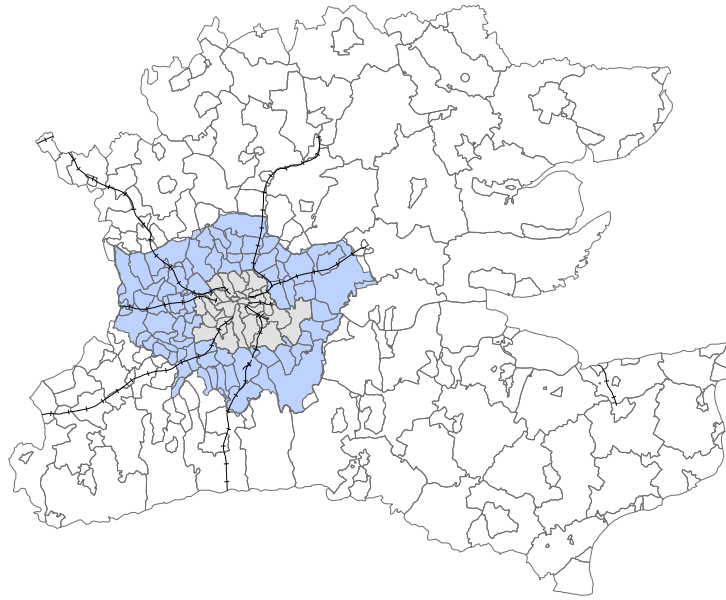
- ROSENTHAL, S. S., AND W. C. STRANGE (2004): "Evidence on the Nature and Sources of Agglomeration Economics," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. Thisse, vol. 4. Elsevier North Holland, Amsterdam.
- (2008): "The Attenuation of Human Capital Spillovers," *Journal of Urban Economics*, 64(2), 373–389.
- SALMON, J. (1891): *Ten Years' Growth of the City of London*. Simkin, Marshall, Hamilton Kent & Company, London.
- SHAW-TAYLOR, L., R. S. DAVIES, P. M. KITSON, G. NEWTON, A. E. M. SATCHELL, AND E. A. WRIGLEY (2010): "The occupational structure of England and Wales c.1817-1881," Work in Progress. For details of the wider project see <http://www.geog.cam.ac.uk/research/projects/occupations/>.
- STAMP, S. J. (1922): *British Incomes and Property: The Application of Official Statistics to Economic Problems*. P. S. King & Son, London.
- VALENTINYI, A., AND B. HERRENDORF (2008): "Measuring Factor Income Shares at the Sectoral Level," *Review of Economic Dynamics*, 11(4), 820–835.
- VOTH, H.-J. (2001): *Time and Work in England 1750-1830*. Oxford University Press, Oxford.
- WARNER, S. B. (1978): *Streetcar Suburbs: The Process of Growth in Boston, 1870-1900*. Harvard University Press, Cambridge MA.
- WHITE, J. (2007): *London in the 19th Century: A Human Awful Wonder of God*. Vintage, Reading.
- (2008): *London in the 20th Century: A City in the 20th Century*. Vintage, Reading.
- (2012): *London in the Eighteenth Century: A Great and Monstrous Thing*. Bodley Head, London.
- WOLMAR, C. (2009): *Fire and Steam: How the Railways Transformed Britain*. Atlantic, London Books.
- (2012): *The Subterranean Railway: How the London Underground was Built and How it Changed the City Forever*. Atlantic, London Books.

Figure 1: Administrative Boundaries



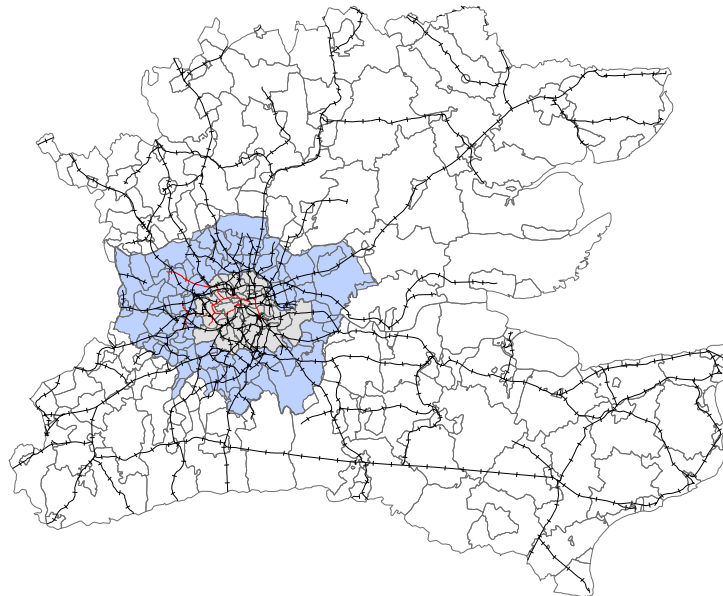
Note: Home counties (thick black outer boundary); Greater London area (red outer boundary); London County Council (LCC) (purple outer boundary); City of London (green outer boundary); boroughs (medium black lines); and parishes (medium gray lines)

Figure 2: Overground Railway Network 1841



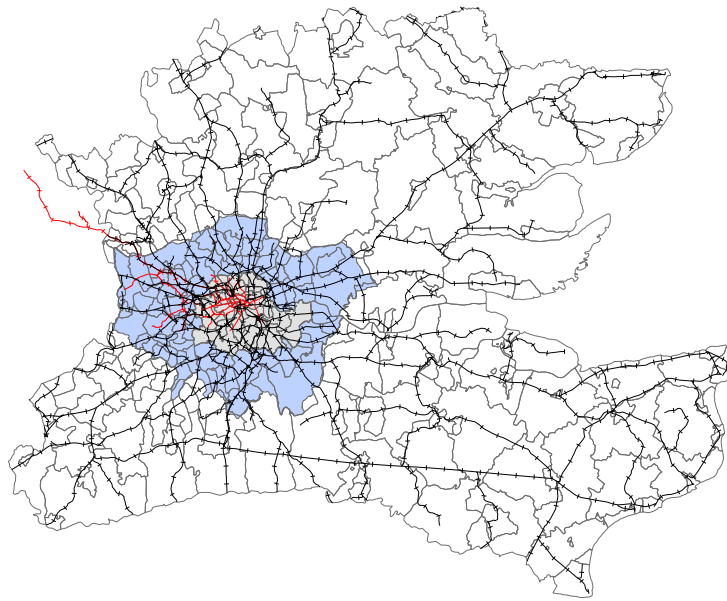
Note: Home counties outside Greater London (white background); Greater London outside London County Council (blue background); London County Council (gray background); overground railway lines shown in black

Figure 3: Overground and Underground Railway Network 1881



Note: Home counties outside Greater London (white background); Greater London outside London County Council (blue background); London County Council (gray background); overground railway lines shown in black; underground railway lines shown in red

Figure 4: Overground and Underground Railway Network 1921



Note: Home counties outside Greater London (white background); Greater London outside London County Council (blue background); London County Council (gray background); overground railway lines shown in black; underground railway lines shown in red

Figure 5: Population Index Over Time (City of London and Greater London, 1801 equals 1)

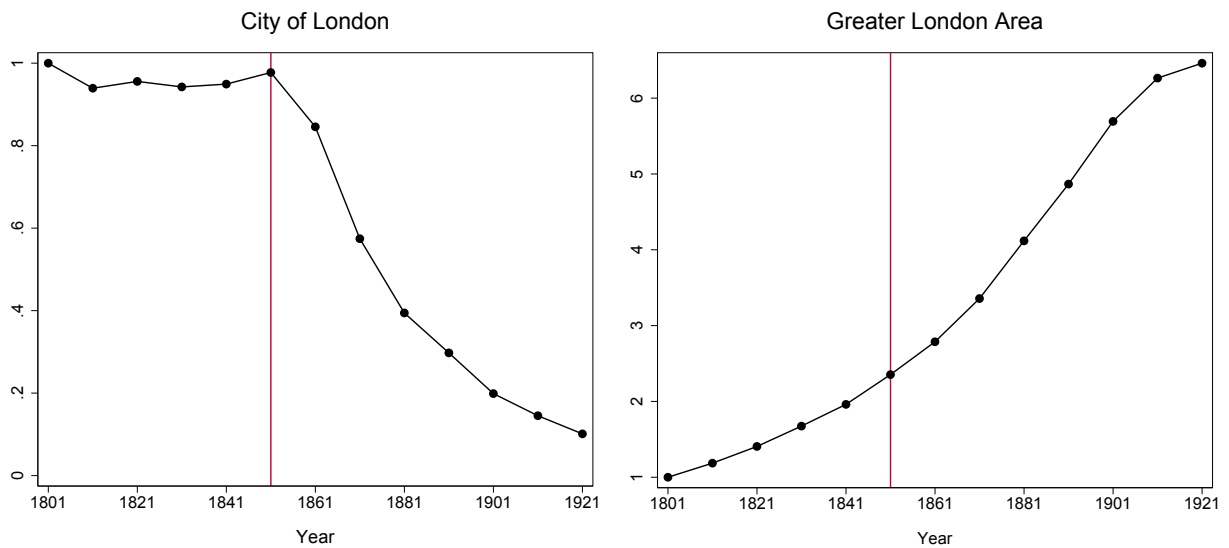


Figure 6: Day and Night Population Over Time (City of London)

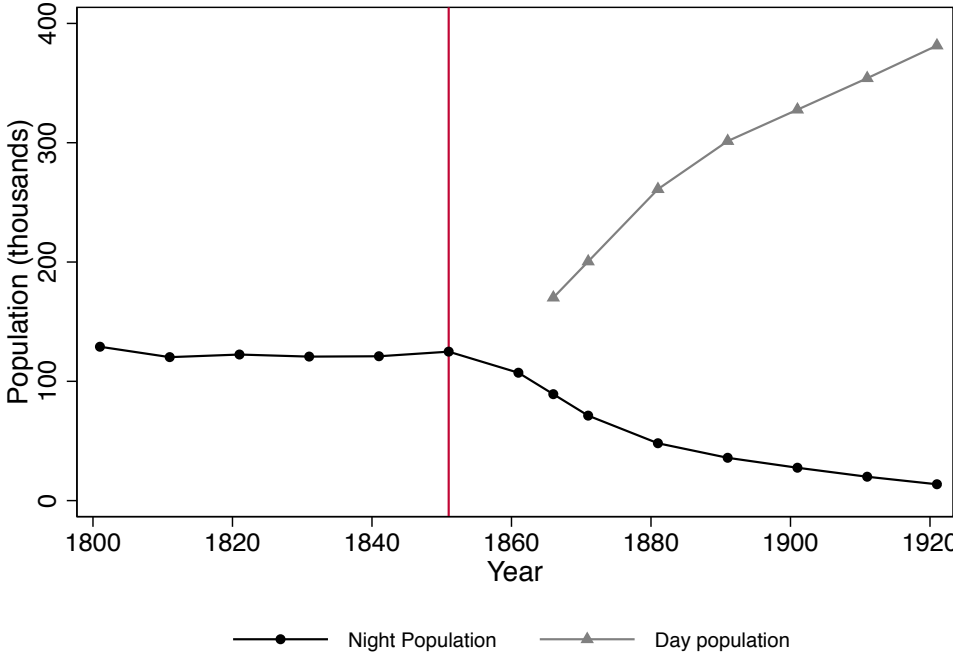


Figure 7: City of London Share of Rateable Value in the County of London



Figure 8: Commuting Distances for Barristers (a Type of Lawyer) Over Time

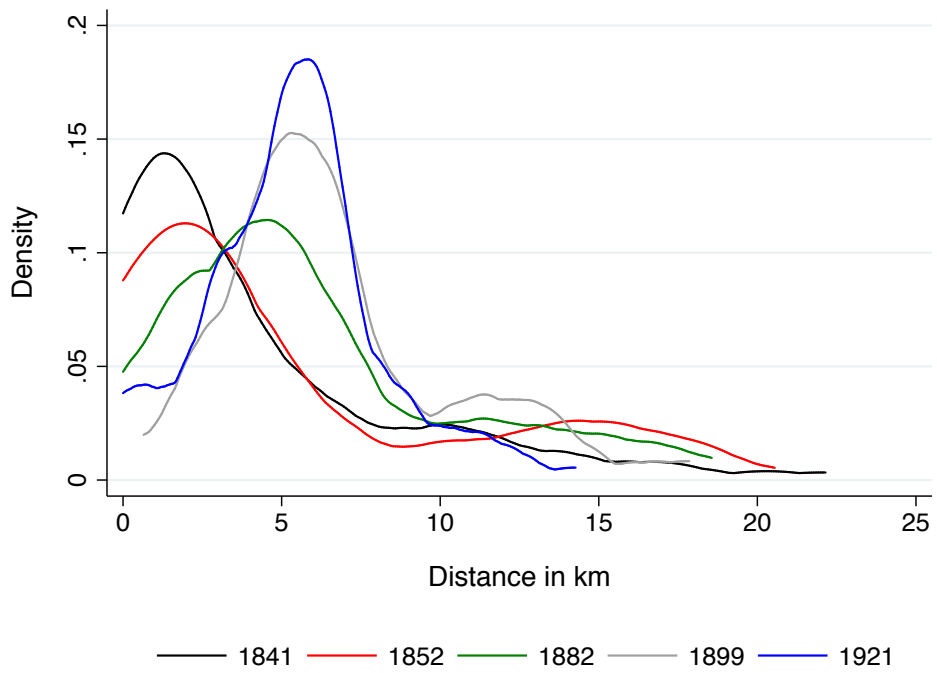
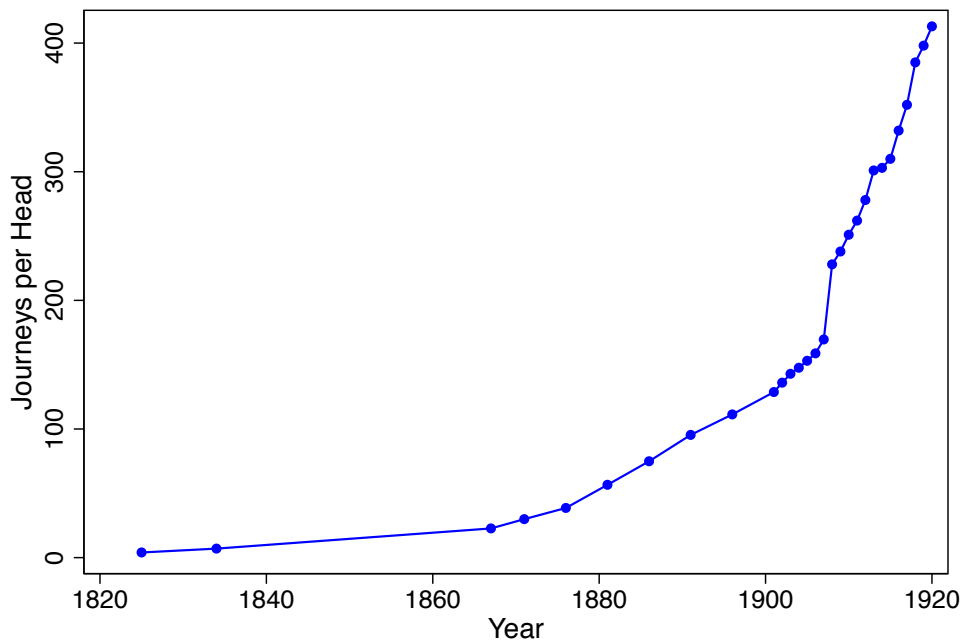


Figure 9: Public Transport Passenger Journeys per Head in the County of London



Note: Public transport includes underground, suburban and mainline rail, tram, short-stage coach, and horse and motor omnibus. Source: London Statistics (London County Council) and Barker (1980).

Figure 10: Borough Shares of Employment by Workplace and Residence in the County of London 1921

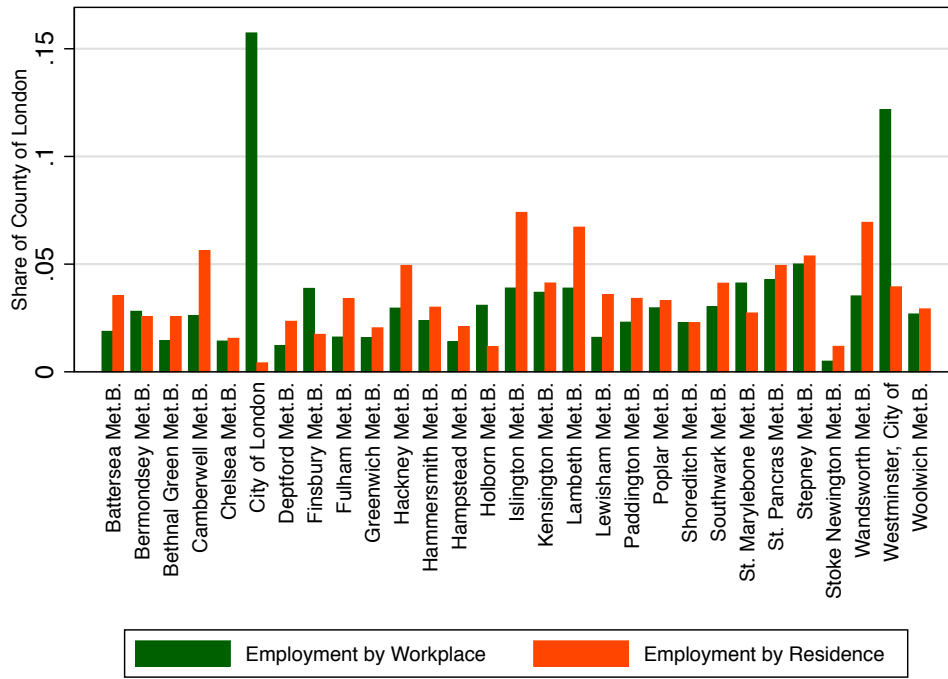
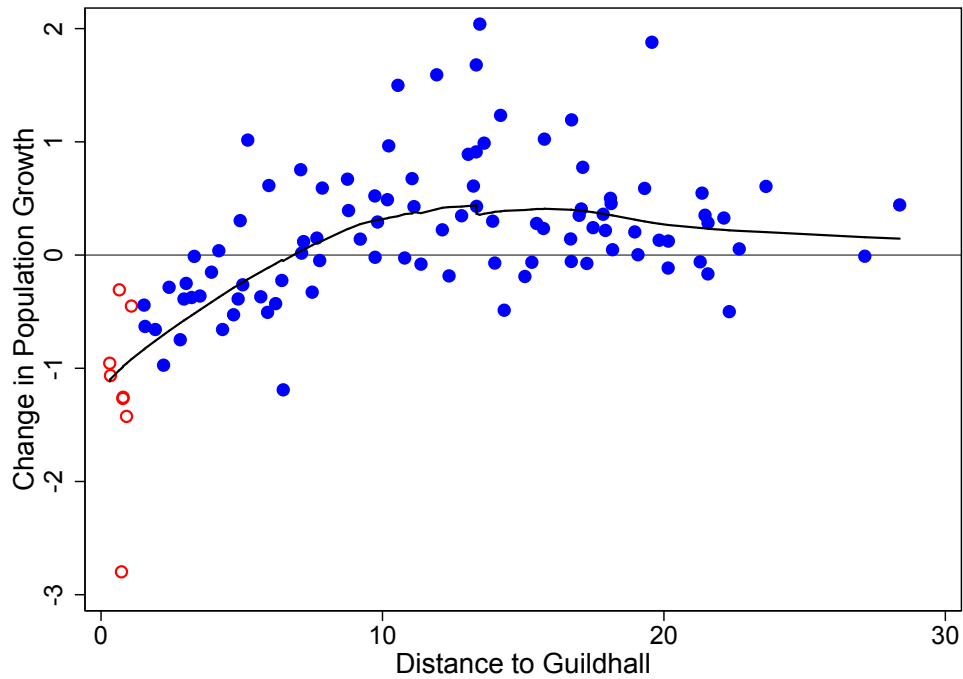


Figure 11: Non-parametric Railway Treatment Estimates for each Parish in Greater London



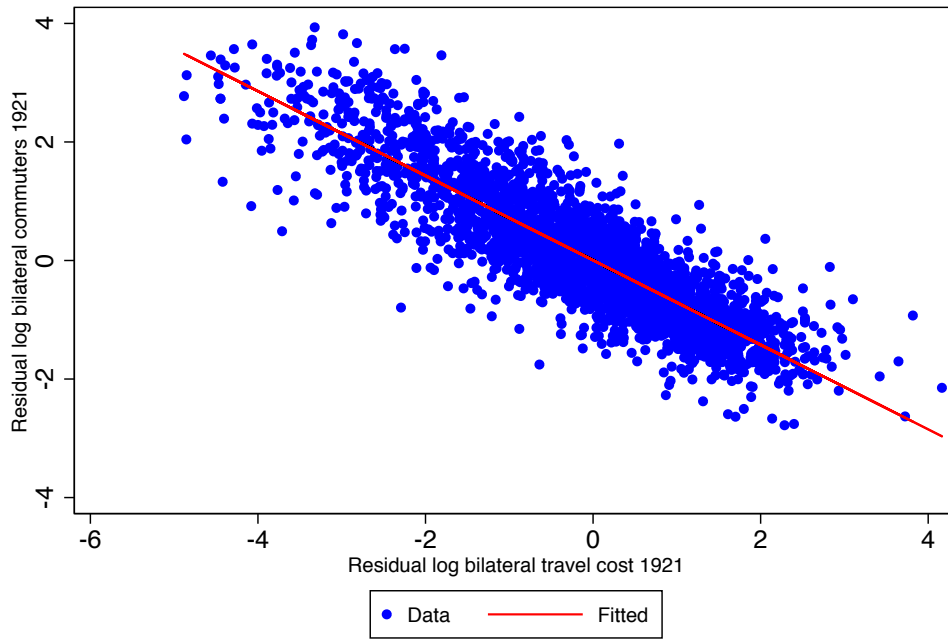
Note: Hollow red circles denote parishes in the City of London; solid blue circles denote parishes in other parts of Greater London.

Table 1: Treatment Effects from Overground and Underground Railways for Greater London Parishes from 1801-1901

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\log L_{pt}$	$\log L_{pt}$	$\log L_{pt}$	$\log L_{pt}$	$\log L_{pt}$	$\log L_{pt}$	$\log L_{pt}$
$\beta_{\tau=10}$	0.610*** (0.137)	0.112*** (0.042)	0.074* (0.040)	0.100 (0.061)	0.763*** (0.251)	0.163** (0.073)	0.146 (0.093)
$\beta_{\tau=20}$	0.946*** (0.188)	0.267*** (0.066)	0.135** (0.059)	0.273** (0.123)	1.129*** (0.335)	0.369*** (0.133)	0.380* (0.200)
$\beta_{\tau=30}$	1.320*** (0.259)	0.479*** (0.110)	0.230*** (0.079)	0.506** (0.214)	1.540*** (0.450)	0.633*** (0.228)	0.675* (0.345)
$\beta_{\tau=40}$	1.734*** (0.336)	0.707*** (0.146)	0.319*** (0.109)	0.753** (0.297)	1.904*** (0.495)	0.858*** (0.265)	0.927** (0.429)
$\beta_{\tau=50}$	2.386*** (0.535)	1.093*** (0.236)	0.535*** (0.148)	1.166*** (0.440)	2.436*** (0.594)	1.246*** (0.368)	1.353** (0.582)
$\beta_{\tau=60}$	2.733*** (0.633)	1.347*** (0.307)	0.560** (0.217)	1.448*** (0.547)	2.756*** (0.682)	1.521*** (0.458)	1.670** (0.717)
$\beta_{\tau=0}$	—	—	—	—	—	—	—
$\beta_{\tau=-10}$				-0.047 (0.068)			-0.066 (0.077)
$\beta_{\tau=-20}$				-0.082 (0.128)			-0.108 (0.137)
$\beta_{\tau=-30}$				-0.088 (0.169)			-0.129 (0.186)
$\beta_{\tau=-40}$				-0.102 (0.228)			-0.146 (0.245)
$\beta_{\tau=-50}$				-0.076 (0.270)			-0.111 (0.278)
$\beta_{\tau=-60}$				-0.083 (0.323)			-0.097 (0.318)
$\gamma_{\tau=10}$					-1.207*** (0.141)	-0.291*** (0.066)	-0.303*** (0.040)
$\gamma_{\tau=20}$					-1.795*** (0.146)	-0.813*** (0.090)	-0.838*** (0.075)
$\gamma_{\tau=30}$					-2.661*** (0.139)	-1.459*** (0.106)	-1.504*** (0.118)
$\gamma_{\tau=0}$	—	—	—	—	—	—	—
$\gamma_{\tau=-10}$							0.046 (0.039)
$\gamma_{\tau=-20}$							-0.004 (0.069)
$\gamma_{\tau=-30}$							0.038 (0.084)
Parish time trends		Yes		Yes		Yes	Yes
Distance Grid \times Year Dummies			Yes				
Year Dummies	Yes	Yes		Yes	Yes	Yes	Yes
Parish fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3133	3133	3133	3133	3133	3133	3133
R-squared	0.917	0.981	0.963	0.981	0.924	0.982	0.982

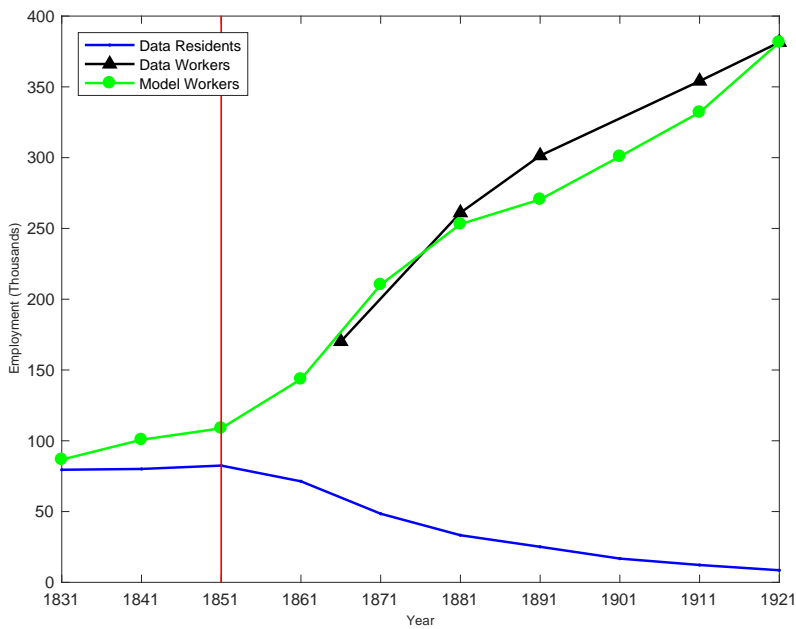
Note: Railway treatment is defined based on whether a parish has an overground or underground railway station; β_{τ} is the rail treatment for treatment year τ ; γ_{τ} allows the rail treatment for treatment year τ to differ between the City of London and other parts of Greater London; $\tau = 0$ corresponds to the treatment year; $\tau = 10$ corresponds to ten years after the treatment year (and so on); $\tau = -10$ corresponds to ten years before the treatment year (and so on); observations are parishes and years; standard errors are clustered on boroughs; * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

Figure 12: Conditional Correlation Between Bilateral Commuting Flows and Bilateral Travel Costs in 1921



Note: Residuals from separate regressions on residence and workplace fixed effects. Sample includes all positive bilateral commuting flows to other locations.

Figure 13: Employment by Workplace and Residence in the City of London (Model and Data) 1831-1921



Note: “Data Residents” is employment by residence as observed in the data; “Data Workers” is employment by workplace as observed in the data; “Model Workers” is the model’s prediction for employment by workplace; the model is calibrated to the 1921 values of rateable values, employment by workplace and employment by residence and used to generate predictions for the removal of the railway network back to 1831.

Figure 14: Borough Shares of Employment by Workplace and Residence in the County of London 1831

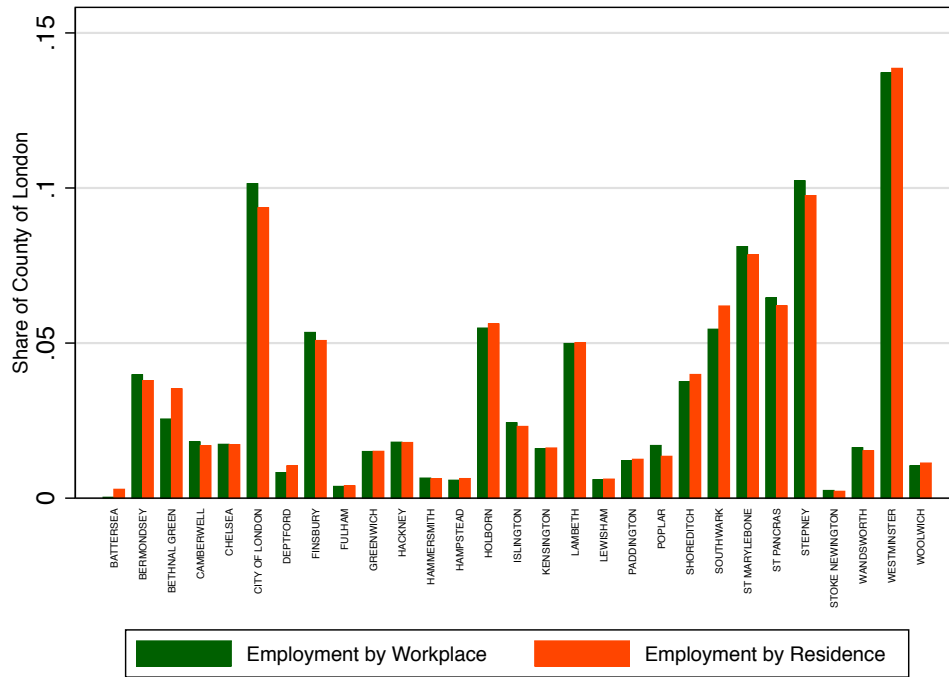
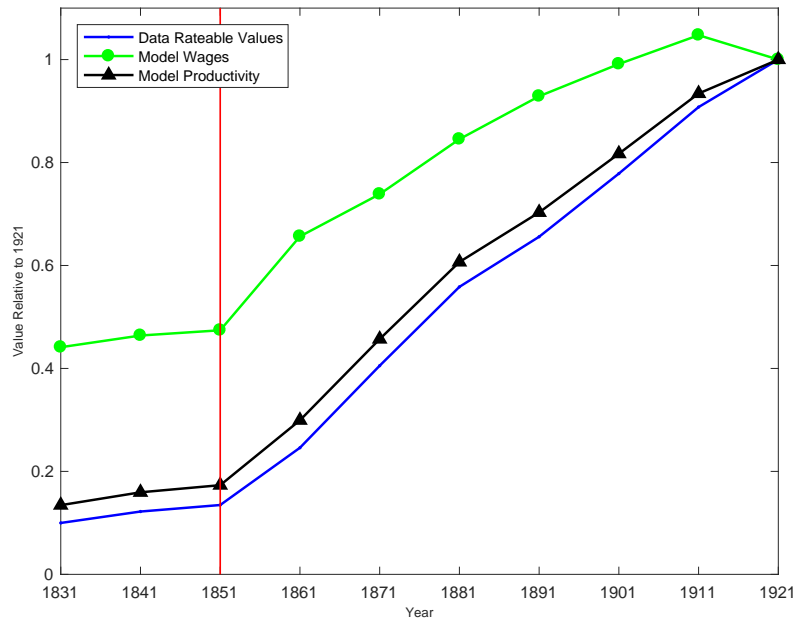
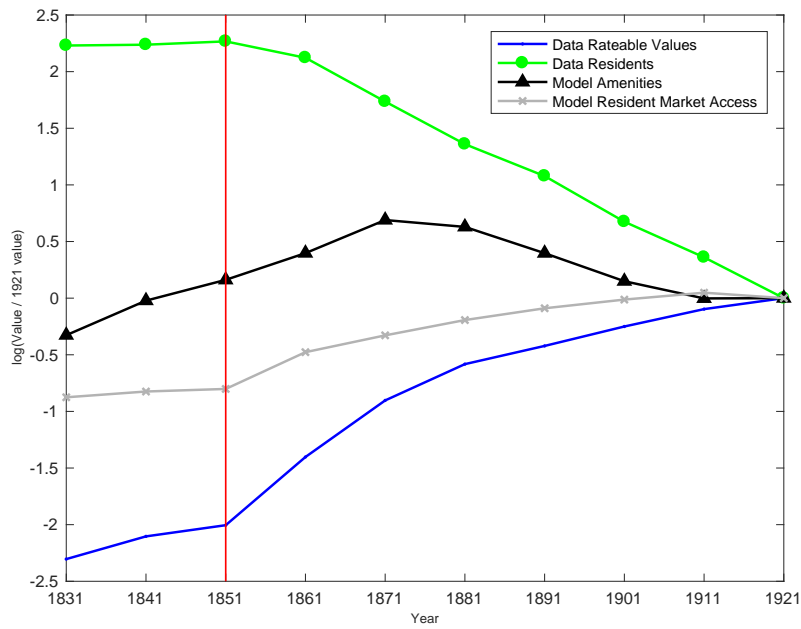


Figure 15: Rateable Values, Wages and Productivity in the City of London 1831-1921 (1921=1)



Note: Rateable Values are observed in the data; the model is calibrated to the 1921 values of rateable values, employment by workplace and employment by residence and used to generate predictions for the removal of the railway network back to 1831. Wages and productivity are predictions from the model. Variables expressed as an index where 1921 equals one.

Figure 16: Rateable Values, Employment by Residence, Amenities and Residents Market Access in the City of London 1852-1921 (1921=1)



Note: Rateable Values and employment by residence are observed in the data; the model is calibrated to the 1921 values of rateable values, employment by workplace and employment by residence and used to generate predictions for the removal of the railway network back to 1831. Amenities and resident market access are predictions from the model. Variables expressed as a log index such that 1921 equals zero.

Web Appendix for “The Making of the Modern Metropolis: Evidence from London (Not for Publication)”

Stephan Heblich

University of Bristol

Stephen J. Redding

Princeton University, NBER and CEPR

Daniel M. Sturm

London School of Economics and CEPR

A Proof of Proposition 1

Proof. With exogenous and strictly positive location characteristics ($A_n \in (0, \infty)$, $B_n \in (0, \infty)$, $\kappa_{ni} \in (0, \infty) \times (0, \infty)$, $L_n \in (0, \infty)$), all locations are incompletely specialized as both workplaces and residences, because the support of the Fréchet distribution for idiosyncratic amenities is unbounded from above. Using the probability of residing in a location (equation (10) for π_n^R), the probability of working in a location (equation (10) for π_n^M), the zero-profit condition (15), and the indifference condition between the city and the larger economy (11), the fraction of workers residing in location n can be written as:

$$\pi_n^R = \frac{H_{Rn}}{H} = \left(\frac{\gamma}{\bar{U}}\right)^\epsilon \sum_{s \in R} B_n A_s^{\epsilon/\tilde{\beta}} \kappa_{ns}^{-\epsilon} Q_s^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}} Q_n^{-\epsilon(1-\alpha)},$$

while the fraction of workers employed in location n can be written as:

$$\pi_n^M = \frac{H_{Mn}}{H} = \left(\frac{\gamma}{\bar{U}}\right)^\epsilon \sum_{r \in R} B_r A_n^{\epsilon/\tilde{\beta}} \kappa_{ri}^{-\epsilon} Q_i^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}} Q_r^{-\epsilon(1-\alpha)},$$

and expected worker income conditional on residing in block i from equation (20) can be written as:

$$v_n = \sum_{i \in R} \frac{A_i^{\epsilon/\tilde{\beta}} \kappa_{ni}^{-\epsilon} Q_i^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}}}{\sum_{s \in R} A_s^{\epsilon/\tilde{\beta}} \kappa_{ns}^{-\epsilon} Q_s^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}}} \left[A_i^{1/\tilde{\beta}} Q_i^{-(1-\tilde{\beta})/\tilde{\beta}} \right],$$

and the land market clearing condition from equation (18) can be written as:

$$\left(\frac{1-\tilde{\beta}}{\tilde{\beta}}\right) \frac{w_n \pi_n^M}{Q_n} + (1-\alpha) \frac{v_n \pi_n^R}{Q_n} = \frac{L_n}{H_n}.$$

Combining the above relationships, this land market clearing condition can be re-expressed as:

$$D_n(\mathbf{Q}) = \frac{1-\tilde{\beta}}{\tilde{\beta}} \left[\frac{A_n^{1/\tilde{\beta}}}{Q_n^{1+(1-\tilde{\beta})/\tilde{\beta}}} \right] \left[\sum_{r \in R} \frac{B_r A_i^{\epsilon/\tilde{\beta}} \kappa_{ri}^{-\epsilon}}{Q_i^{\epsilon(1-\tilde{\beta})/\tilde{\beta}} Q_r^{\epsilon(1-\alpha)}} \right] + \frac{1-\alpha}{Q_n} \left[\sum_{i \in R} \left(\frac{A_i^{\epsilon/\tilde{\beta}} \kappa_{ni}^{-\epsilon} Q_i^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}}}{\sum_{s \in R} A_s^{\epsilon/\tilde{\beta}} \kappa_{ns}^{-\epsilon} Q_s^{-\epsilon(1-\tilde{\beta})/\tilde{\beta}}} \right) \frac{A_i^{1/\tilde{\beta}}}{Q_i^{(1-\tilde{\beta})/\tilde{\beta}}} \right] \left[\sum_{s \in R} \frac{B_n A_s^{\epsilon/\tilde{\beta}} \kappa_{ns}^{-\epsilon}}{Q_s^{\epsilon(1-\tilde{\beta})/\tilde{\beta}} Q_n^{\epsilon(1-\alpha)}} \right] - L_n = 0,$$

for all $n \in R$, where we have chosen units in which to measure utility such that $(\bar{U}/\gamma)^\epsilon/H = 1$. The above land market clearing condition provides a system of equations for the N boroughs in terms of the N unknown floor prices Q_n , which has the following properties:

$$\lim_{Q_n \rightarrow 0} D_n(\mathbf{Q}) = \infty > L_n, \quad \lim_{Q_n \rightarrow \infty} D_n(\mathbf{Q}) = 0 < L_n,$$

$$\frac{dD_n(\mathbf{Q})}{dQ_n} < 0, \quad \frac{dD_n(\mathbf{Q})}{dQ_i} < 0, \quad \left| \frac{dD_n(\mathbf{Q})}{dQ_n} \right| > \left| \frac{dD_n(\mathbf{Q})}{dQ_i} \right|.$$

It follows that there exists a unique vector of residential floor prices \mathbf{Q} that solves this system of land market clearing conditions. Having solved for the vectors of floor prices (\mathbf{Q}), the vector of wages \mathbf{w} follows immediately from the zero-profit condition for production (15). Given floor prices (\mathbf{Q}) and wages (\mathbf{w}), the probability of residing in a location (π^R) follows immediately from (10), and the probability of working in a location (π^M) follows immediately from (10). Having solved for $\{\pi^M, \pi^R, \mathbf{Q}, \mathbf{w}\}$, the total measure of workers residing in the city can be recovered from our choice of units in which to measure utility ($(\bar{U}/\gamma)^\epsilon/H = 1$), which together with population mobility (11) implies:

$$H = \left[\sum_{r \in R} \sum_{s \in R} B_r w_s^\epsilon (\kappa_{rs} Q_r^{1-\alpha})^{-\epsilon} \right].$$

We therefore obtain $\mathbf{H}^M = \pi^M H$ and $\mathbf{H}^R = \pi^R H$. This completes the determination of the equilibrium vector $\{\pi^M, \pi^R, H, \mathbf{Q}, \mathbf{w}\}$. ■