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

August 31, 2017

Abstract

A central feature of economic development is the concentration of population in large metropolitan areas. We provide new evidence on the strength of the agglomeration forces underlying these population concentrations using spatially-disaggregated data for London from 1801-1921, the innovation in transport technology provided by steam railways, and a quantitative urban model. Steam railways dramatically reduced travel times and permitted the first large-scale separation of workplace and residence to achieve agglomeration economies in employment density. We show that our model is able to account both qualitatively and quantitatively for this observed separation and the resulting changes in land prices.

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

Work in Progress and Incomplete

*We are grateful to Bristol University, the European Research Council (ERC), the London School of Economics, and Princeton University for research support. 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

A central feature of economic development is the concentration of population in large metropolitan areas, including London (8.4 million) and New York (8.5 million). A major research question is determining the economic forces that explain the emergence of these population concentration, including agglomeration forces (e.g. economies of employment density), dispersion forces (e.g. inelastic land supply and commuting costs), and transport technology (e.g. costs of commuting). We combine spatially-disaggregated data for London from 1801-1921, the mid-nineteenth century transport revolution inaugurated by the invention of steam railways, and a quantitative urban model to provide new evidence on the strength of agglomeration forces. 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 exploitation of economies of employment density, as emphasized in standard theories of agglomeration. Using both reduced-form and structural approaches, we provide evidence of a causal effect of steam passenger railways on city size and structure. Using the structure of model, we estimate that these observed changes in city size and structure imply substantial agglomeration forces.

Methodologically, we develop a framework for decomposing the organization of economic activity within cities into the contributions of productivity, amenities and commuting costs. We show that this framework holds in an entire class of models consistent with a gravity equation for commuting flows and can be implemented using only data on bilateral commuting flows and land prices. Although we only observe bilateral commuting flows in 1921 at the end of our sample period, we show how this framework can be used to undertake counterfactuals for changes in transport technology, such as the development of the steam engine. Using observed historical data on population, land prices and the transport network back to the early-nineteenth century, we show how these counterfactuals can be used to structurally estimate the parameters of the model. Given the estimated parameters, we show how the model can be used to infer missing historical data on employment by workplace for each location over 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 show that incorporating agglomeration forces is central to matching the separation of workplace and residence in the City of London following the development of the steam railway.

Nineteenth-century London is arguably the poster child for the large metropolitan areas observed around the world today. In 1801, Greater London housed around 1 million people and spanned only 5 miles East to West. Therefore this was a walkable city of 60 squares and 8,000 streets that was not radically different from other large cities from history. By 1921, Greater London contained over 7 million people, measured more than 17 miles across, and was on a dramatically larger scale than any previous urban area. This was by some margin the largest city in the world (with New York and Paris having 1921 populations of 5.6 million and 4.9 million respectively), equivalent to the combined populations of Greece, Denmark and Bosnia-Herzegovina. 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).[1] In contrast to other cities such as Paris, London also developed through a largely haphazard and organic process. Until the creation of the Metropolitan Board of Works (MBW) in 1855, there no municipal authority that spanned the many different local jurisdictions that made up Greater London, and its 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 empirical literature on agglomeration forces, as reviewed in Rosenthal and Strange (2004) and Combes and Gobillon (2015). A key challenge in this literature 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 natural experiment from the invention of the steam locomotive, which was first developed to haul freight from mines, and only later used for passenger transport.

Second, our paper is related to a recent body of research on quantitative spatial models, including Allen and Arkolakis (2014), Redding and Sturm (2008), Redding (2016), Caliendo, Parro, Rossi-Hansberg, and Sarte (2017) and Klaus Desmet and Rossi-Hansberg (2017), as reviewed in Redding and Rossi-Hansberg (2017). In contrast to calibration-based approaches, we structurally estimate the parameters of our model, and provide overidentification checks in which we compare the model's counterfactual predictions to the observed impact of the steam railway. Our work is most closely related to Ahlfeldt, Redding, Sturm, and Wolf (2015), which uses a quantitative urban model to examine the impact of the division and reunification of Berlin. From an empirical perspective, the main difference is that we investigate a different question (the role of transport technology in shaping city development) and examine data from an entirely different empirical setting. From a theoretical point of view, we develop a new decomposition of economic activity (into productivity, amenities and commuting costs) and a new methodology for estimating the model's structural parameters. This decomposition and structural estimation both use the model's commuting gravity equation and complete census data on bilateral commuting flows.[2] These new methodologies is widely applicable in other contexts where bilateral commuting and land price data are available.

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 examined this relationship 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 (1982). Within this literature, our work is most closely related to research on suburbanization and decentralization, including

[1] Stationary steam engines have a longer history, dating back at least to Thomas Newcomen in 1712, as discussed further below.
[2] Only incomplete survey data on bilateral commuting flows were available for Berlin, which would have precluded this approach.

Baum-Snow (2007), Baum-Snow, Brandt, Henderson, Turner, and Zhang (2017), and Baum-Snow (2017). Our contributions relative to this research are to use the large-scale variation from the transition from human/horse power to steam locomotion and to combine both reduced-form and structural estimation approaches. We show that our structural 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 London. Section 5 introduces the theoretical models. Section 6 uses the model's gravity equation for commuting flows to develop a new decomposition of internal city structure. Section 7 structural estimates the parameters of the model. Section 8 concludes.

2 Historical Background

The settlement of London predates the Roman Conquest of England in 43 CE. The historical and commercial center of Greater London is what is known as the "City of London" (commonly referred to as simply the "City" or the "Square Mile"), whose boundaries correspond closely to the Roman wall constructed between 190 and 225 CE. 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.[3]

In the first half of the 19th-century, there was no central municipal authority responsible for Greater London as a whole, 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.[4] 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.[5] 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.[6] 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

[3]For historical discussions of London's growth, see Ball and Sunderland (2001), Kynaston (2012), Porter (1995) and White (2012, 2007, 2008).

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

[5]See 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).

[6]The 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).

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.[7]

Against this background, the steam 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 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 brought a major improvement in the efficiency and an expansion in the range of potential applications of steam engines. The first commercial use of steam locomotives was to hail 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.

Railway development in London and the United Kingdom more generally was undertaken by a large number of private companies and proceeded in a competitive and uncoordinated fashion. 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 River Thames to the South.[8] 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 1868 and the subsequent development of the Circle and District underground lines. While these early underground railways were built using "cut and cover" technology, further penetration of underground railways through Central London occurred with the development of the technology for boring deep-tube railways, as first used for the City and South London Railway, which opened in 1890, and is now part of the Northern Line. These deep-tube railways further reduced travel times into the most central parts of Greater London.

3 Data

We construct a new spatially-disaggregated dataset for London for the period 1801-1921. We distinguish four main geographical areas, which we now list from largest to smallest, where each subsequent region is a subset of the previous one. First, we consider the Home Counties including Greater London, which contain a 1921 population of 9.61 million and an area of 12,828.78 kilometers squared, and encompass large parts of South-East England.[9] 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 1,594.70 kilometers squared. Third, we consider the County

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

[8] This 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, 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).

[9] The Home Counties including Greater London are Essex, Hertfordshire, Kent, London, Middlesex and Surrey.

of London, which has a 1921 population of 4.48 million and an area of 313.52 kilometers squared. Fourth, we examine the City of London, with a 1921 population of 13,709 and an area of 2.98 kilometers squared.

Our main data source is the Population Census of England and Wales, which begins in 1801, and is enumerated every decade thereafter. Data are available at two main levels of spatial aggregation: boroughs and parishes, where the Home Counties including Greater London encompasses 257 boroughs and 2,325 parishes; Greater London contains 99 boroughs and 458 parishes; the County of London comprises 29 boroughs and 305 parishes; and the City of London includes 1 borough and 122 parishes.[10] In Figure 1, we show the outer boundary of the Home 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. In Figure 2, we zoom in to show the County of London in greater detail. As apparent from these figures, our data permit a high-level of spatial resolution, where the average parish in Greater London has a 1921 population of around 3,000 and an area of 3.46 kilometers squared, while the average borough in Greater London has a 1921 population of around 75,000 and an area of 16.12 kilometers squared.

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.[11] Using this matrix, we find that commuting flows between other parts of England and Wales and Greater London are small even in 1921.[12] Therefore, although our framework could accommodate these flows from other parts of England Wales, we abstract from them for simplicity and focus on Greater London, treating it as closed commuting market. Summing across rows in the matrix of bilateral commuting flows for Greater London, we obtain employment by workplace for each borough ("workplace employment"), while summing across columns, we obtain employment by residence ("residence employment") for each borough. 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 in 1921 with historical population data for each parish and borough from earlier population censuses from 1801-1911. Assuming that employment participation rates are stable over time, we use the 1921 employment participation rate for each borough and the historical population data to construct residence employment for each borough for each decade from 1801-1911.[13] Parish and borough boundaries are relatively stable throughout most of the nineteenth century, but experience substantial change in 1891 and later census years. For our reduced-form empirical analysis using the parish-level data, we constructed constant parish boundary data for the period 1801-1911. 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 for these years 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

[10] Parish boundaries in the population census change over time. We construct constant definitions of parish boundaries using the classification provided by Shaw-Taylor, Davies, Kitson, Newton, Satchell, and Wrigley (2010), as discussed further below.

[11] The population census for England and Wales is one of the earliest 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.

[12] In 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 based on where one slept on Census night, while workplace is based on one's usual place of work, some of this 4 percent correspond to business trips.

[13] We allow in our quantitative analysis of the model for a common change in the employment participation rate across all boroughs, which enters the model like a change in the units in which productivity, amenities and the reservation level of utility in the wider economy are measured. Empirically, we find relatively little variation in employment participation rates across boroughs in 1921.

boroughs, most parishes lie within a single 1921 borough.

We measure land rents 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). 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. We use the rateable values from Schedule A of the national income tax, which is the schedule concerned with profits from the ownership of lands and property, where the national income tax was first introduced in 1799. As these rateable values are the basis for national income tax collection, they are widely regarded as corresponding most closely to market valuations.[14] We construct these data for each parish and borough in Greater London during our sample period from various publications of London County Council and the House of Commons, as discussed further in the data appendix.

We use our bilateral commuting data for 1921, together with our data on residence employment and land rents, to structurally estimate the model and generate predictions for workplace employment for years before 1921. 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.[15] In the face of a declining residential population ("night population"), the City of London undertook these censuses of the "day population" to demonstrate its continuing 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."[16] Given the high opportunity cost of unused space in the City of London, those present during the day are likely to be employed, and we provide evidence below that the day population is a reasonable 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 historical data on residence and workplace addresses from years before 1921 from business directories, as discussed further in the data appendix.

4 Reduced-Form Evidence

In subsection 4.1, we present reduced-form evidence of the change in the organization of economic activity within Greater London over our historical time period. In subsection 4.2, we use an event-study specification to provide causal evidence of the role of the development of steam railways in shaping this reorganization of economic activity. In subsection 4.3, we generalize this event-study specification to explore further the heterogeneity in the treatment effect of steam railways within Greater London.

[14] For example, Stamp (1922) argues that "It is generally acknowledged that the income tax, Schedule A, assessments are the best approach to the true values" (page 25). In general, these Schedule A rateable values can differ from those used for the Poor Law or County rates, except in the County of London after the Metropolis Act of 1869, which required all rateable values to be computed on the basis of Schedule A income taxation.

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

[16] Salmon (1891), page 97.

4.1 City Size and Structure over Time

In Figure 3, we display the evolution of population from 1801-1921 for the City of London (left panel) and Greater London (right panel). Population is expressed as 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 (shown by the red vertical line) onwards, there is a sharp drop in population in the City of London, which falls by around 90 percent to 13,709 by 1921, with the 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. Therefore, the rapid expansion in population for Greater London as a whole throughout the nineteenth century goes hand in hand with a precipitous drop in population in its most central commercial area from mid-century onwards.

The population series displayed in Figure 3 are from the population census for England and Wales and are based on residence ("night population"). In Figure 4, we contrast the evolution of this "night population" for the City of London with the time path of its "day population," as measured by the Day Censuses of 1866, 1881, 1891 and 1911 (where we extrapolate back to 1861). For 1921, the day population numbers equal workplace employment from the population census for England and Wales. As apparent from the figure, the sharp decline in night population in the City of London from 1851 onwards coincides with a sharp rise in its day population. This contrast between night and day population in Figure 4 suggests that the differences in the evolution of night population in Figure 3 are explained by the City of London becoming increasingly specialized 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) suggests 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 England and Wales 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.

In Figure 5, we display the City of London's share of rateable value in the County of London, which includes the City of Westminster and the most densely-populated inner suburbs of Greater London. In the early nineteenth and late-twentieth centuries, the City of London's rateable value share declines from 11-9 and 14-12 percent respectively. These declines have a natural explanation in terms of an expansion in the geographical boundaries of developed land with urban growth. As these geographical boundaries expand outwards, and undeveloped land becomes developed, this tends to reduce the share of already-developed land in overall land values. In contrast, in the decades immediately after 1851 when the City of London experiences the greatest declines in residential population, its rateable value share increases sharply from 9-14 percent. Therefore, as the City of London increasingly specialized as a workplace rather than as a residence in Figures 3-4, it became a relatively more valuable location in Figure 5.

In Figures 6, 7 and 8, we show 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, the first 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. A comparison of Figures 3-4 and Figures 6-8 already suggests that the evolution of population and employment in the City of London is likely to be related to changes in the transport technology. The delayed response in the City of London's population in Figure 3 (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 3 correspond closely to the greatest increases in the penetration of overground and underground railways into the central areas of the County of London in Figure 7. We provide further evidence on this timing of the population response to the new transport technology using an event study specification below.

4.2 Event Study Specification

We now use our spatially-disaggregated parish-level data from 1801-1911 and an event-study specification to provide causal evidence of the role of steam railways in driving the observed reorganization of economic activity. The main identification challenge is that steam railways are unlikely to randomly assigned, because railways 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 using an event study specification, in which we include both a parish fixed effect and a parish time trend, and examine non-parametrically the relationship between deviations from these parish trends and the timing of the arrival of the railway. In particular, we consider the following regression specification:

$$\begin{aligned} \ln L_{pt} \quad &= \sum_{p=1}^{P} (\alpha_p \times \mathbb{I}_p) + \sum_{p=1}^{P} (\delta_p \times \mathbb{I}_p \times \text{Year}_t) \\ &+ \sum_{\tau=-T}^{\tau=T} \beta_\tau (R_p \times \mathbb{I}_\tau) + \sum_{\tau=-T}^{\tau=T} \gamma_\tau \left(R_p \times \mathbb{I}_\tau \times \mathbb{I}_{\text{City}}\right) + u_{pt}, \end{aligned} \tag{1}$$

where p indexes parishes; t indicates the census year; τ references the treatment census year, which we define as the census year minus the last census year in which the parish has no railway (such that $\tau = 0$ is the treatment year); [17] L_{pt} is parish population; $\mathbb{I}_p$ is an indicator variable that is one for parish p and zero otherwise; α_p is the estimated parish fixed effect; Year_t is a time trend for census year; δ_p is the estimated parish time trend; R_p is an indicator variable that equals one if a parish is ever intersected by a railway in any year of the sample; $\mathbb{I}_\tau$ is an indicator variable for treatment year τ; $\mathbb{I}_{\text{City}}$ is an indicator variable that equals one if a parish is in the City of London and zero otherwise; u_{pt} is a stochastic error.

The inclusion of the fixed effect (α_p) and time trend (δ_p) ensures that this specification has a "difference-in-differences" interpretation, where the first difference is across parishes and the second difference is across time.

[17] We 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 ($\tau = 10$).

To address the concern that parishes that are never intersected by a railway could differ systematically from other parishes, we drop them from our baseline estimation sample, although including these parishes yields a similar pattern of results, both qualitatively and quantitatively. As our sample period using these parish-level data is 110 years long, and the first railways appear less than through the sample in the census year 1841, we drop observations with treatment years less than -50 and greater than 50. This ensures that the observations in the sample are distributed relatively evenly around the treatment year, although again we find a similar pattern of results, both qualitatively and quantitatively, if we include observations with treatment years outside this interval.

The inclusion of the parish fixed effect (α_p) controls for the non-random assignment of railways based on the level of log parish population. The inclusion of the parish time trend (δ_p) controls for the non-random assignment of railways based on trends in log parish population. The first key coefficients of interest $\{\beta_\tau\}$ capture average deviations from these parish time trends for all parishes, for years leading up to and following the treatment year, where the treatment year ($\tau = 0$) is the excluded category. The second key coefficients of interest $\{\gamma_\tau\}$ capture the difference in the deviation from parish time trends for the City of London relative to other parts of Greater London, for years leading up to and following the treatment year, where the treatment year ($\tau = 0$) is again the excluded category. If there is a causal effect of railways in increasing population in the parts of Greater London outside the City of London, we should expect the $\{\beta_\tau\}$ to be flat and statistically insignificant in the years leading up to the arrival of the railway, and positive and statistically significant thereafter. Similarly, if the causal effect of railways on population is smaller in the City of London than in other parts of Greater London, we should expect $\{\gamma_\tau\}$ to be flat and statistically insignificant in the year leading up to the arrival of the railway, and negative and statistically significant thereafter. Statistically significant deviations of $\{\beta_\tau\}$ or $\{\gamma_\tau\}$ from zero in the years before the arrival of the railway would be evidence of non-random assignment that is not controlled for by either the parish fixed effect (α_p) or time trend (δ_p).

In Figure 9, we display the estimated coefficients $\{\beta_\tau\}$ and 95 percent confidence intervals for parishes outside the City of London. Consistent with railways causing increases in population outside the City of London, the estimated $\{\beta_\tau\}$ are statistically indistinguishable from zero in the years before the arrival of the railway, but become positive and statistically significant after the arrival of the railway. The estimated magnitudes are large, with a railway estimated to raise population by over 30 percent above its pre-existing trend after 50 years. In Figure 10, we display the estimated coefficients $\{\beta_\tau\}$ and 95 percent confidence intervals for parishes outside the City of London. Consistent with railways causing redistributions of population away from the City of London to other parts of Greater London, we find that the estimated $\{\gamma_\tau\}$ are again statistically indistinguishable from zero in the years before the arrival of the railway, but become negative and statistically significant after the arrival of the railway. The estimated magnitudes are now smaller and less precisely estimated, which is consistent with only a small number of parishes inside the City of London being intersected by railways, because of the parliamentary exclusion zone.

Together, the absence of pre-trends in Figures 9-10, and the clear juxtaposition of positive effects outside the City of London and negative effects inside the City of London, provides empirical support for our interpretation that the steam railway played a causal role in the reorganization of economic activity within Greater London.

4.3 Treatment Heterogeneity

In the event study specification in the previous section, we assumed that the railway treatment effect is heterogeneous inside versus outside the City of London. Although this assumption is consistent with our characterization

of the reorganization of economic activity within Greater London in subsection 4.1 above, we now relax it using an alternative non-parametric specification, in which we estimate separate railway treatment effect for each parish. 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. Again we drop parishes that are never intersected by a railway. Additionally, to ensure that all parishes are treated with the railway for a similar interval of time, we now further restrict the sample to observations with treatment years in between -30 and 30.

In a first step, we difference log population for each parish in each year from the mean across parishes in that year, which ensures that we focus on relative population across parishes:

$$\ln \tilde{L}_{pt} = \ln L_{pt} - \frac{1}{P} \sum_{p=1}^{P} \ln L_{pt}, \tag{2}$$

where $\tilde{L}_{pt}$ denotes relative population and P is the number of parishes. This first difference from average log population in each year allows us to control for the fact that the railway arrives in different years in different parishes and there is a secular increase in population over time.

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 $\tau = 0$):

$$\Delta \ln \tilde{L}_{p}^{\text{pre}} = \ln \tilde{L}_{p,\tau=0} - \ln \tilde{L}_{p,\tau=-30}, \tag{3}$$

where the difference over time differences out any fixed effect in the level of log relative parish population.

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 $\tau = 30$):

$$\Delta \ln \tilde{L}_{p}^{\text{post}} = \ln \tilde{L}_{p,\tau=30} - \ln \tilde{L}_{p,\tau=0}, \tag{4}$$

where the difference over time again differences out any fixed effect in the level of log relative parish population.

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.

$$\Delta \ln \tilde{\tilde{L}}_{p} = \Delta \ln \tilde{L}_{p}^{\text{post}} - \Delta \ln \tilde{L}_{p}^{\text{pre}}, \tag{5}$$

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-specific time trend that is common to these two periods. Therefore, we again focus on deviations from parish-specific trends, as in the previous subsection.

In Figure 11, we display these double differences in relative population growth for each parish against Euclidean distance from the center of the City of London, as defined by the location of the Guildhall. We indicate parishes in the City of London by hollow red circles, while parishes in the other parts of the County of London are denoted by solid blue circles. We show mean population growth within and beyond five kilometers from the Guildhall by the horizontal red and blues lines respectively. As apparent from the figure, we find a sharp non-linear relationship between the treatment effect from the arrival of the railway on population growth and distance from the Guildhall. Given that relative log population is measured as a deviation from its average value across parishes, the average treatment effect of the railway across all parishes is necessarily equal to zero. However, for parishes within five kilometers of the

Guildhall, we find negative estimated treatment effects, particularly for those parishes inside the City of London. In contrast, for parishes beyond five kilometers from the Guildhall, we find positive estimated treatment effects, where these 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 reorganization of economic activity in response to the arrival of the railway. Even without imposing the distinction between the City of London and other parts of the County of London, we find that the railway reduces relative population growth close to the center of the City of London, and increases relative population growth further from the center of the City of London.

4.4 Mechanisms

We now provide further direct evidence that the changes in night and day population in the previous sections involve changes in commuting behavior. Although the 1921 population census is the first to report systematic information on bilateral commuting patterns, business directories report the residence and workplace addresses for a number of professions for 1841, 1852, 1882, 1899 and 1921. We focus on barristers (a type of lawyer) for which data on a large number of individuals were available for each year. 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 Euclidean distance between these addresses. In Figure 12, we display kernel density estimates of the distribution of commuting distances across barristers for each year. As apparent from the figure, we find that there is 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 6- 8 and the sharp drop in population in the City of London in Figure 3. In Figure 13, 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, and horse and motor omnibus. 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 14, we provide cross-section evidence on the specialization of boroughs within the County of London as workplace or residence centers at the end of our time period in 1921. As evident from the figure, 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 Model

We now develop an equilibrium model of internal city structure to use the observed changes in the organization of economic activity to structurally estimate the strength of agglomeration and dispersion forces. We consider a city

(Greater London) embedded within a wider economy (the United Kingdom). The city consists of a discrete set of locations (the boroughs that are observed in our data). Workers are geographically mobile and choose whether to live and work in the city or the wider economy.[18] Population mobility implies that the expected utility from choosing to live and work in the city equals the reservation level of utility in the wider economy $\bar{U}$. If a worker chooses the city, she choose a residence n and a workplace i from the set of locations $n, i \in R$ to maximize her utility.[19] 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 land, and transport connections.

5.1 Preferences

Worker preferences are defined over consumption of a composite final good and residential land use. 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:[20]

$$U_{ni}(\omega) = \frac{z_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \qquad 0 < \alpha < 1, \tag{6}$$

where P_n is the price of the composite final good, Q_n is the price of land, 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.[21]

We observe positive residents and employment for all boroughs in our data. Therefore, no-arbitrage between alternative land uses ensures that Q_n is the common price of land for residential and commercial use. 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 land. All land is owned by absentee landlords, who receive payments for land use, and consume only the composite final good.

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

$$G_{ni}(z) = e^{-B_n z^{-\epsilon}}, \qquad B_n > 0, \ \epsilon > 1, \tag{7}$$

where B_n in the Fréchet scale parameter, which determines average residential amenities in location n. Therefore, this parameter allows some locations to more attractive in terms of their residential amenities than others (e.g. leafy streets and scenic views). 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 ϵ, the greater the heterogeneity in idiosyncratic amenities, and the less sensitive worker location decisions to economic variables. Conditional on choosing to live in Greater London, equations (6) and (7) imply that the probability a worker chooses to reside in n and work in i is

$$\pi_{ni} \equiv \frac{H_{ni}}{H} = \frac{B_n w_i^{\epsilon} \left(\kappa_{ni} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{r \in R} \sum_{s \in R} B_r w_s^{\epsilon} \left(\kappa_{rs} Q_r^{1-\alpha}\right)^{-\epsilon}}, \tag{8}$$

[18] Motivated 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.

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

[20] For 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).

[21] Although we model commuting costs in terms of utility, they enter the indirect utility function (6) multiplicatively with the wage, which implies that there is a closely-related formulation in terms of the opportunity cost of time spent commuting.

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):

$$\pi_n^R = \frac{H_n^R}{H} = \frac{\sum_{s\in R} B_n w_s^\epsilon \left(\kappa_{ns} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{r\in R}\sum_{s\in R} B_r w_s^\epsilon \left(\kappa_{rs} Q_r^{1-\alpha}\right)^{-\epsilon}}, \qquad \pi_i^M = \frac{H_i^M}{H} = \frac{\sum_{r\in R} B_r w_i^\epsilon \left(\kappa_{ri} Q_r^{1-\alpha}\right)^{-\epsilon}}{\sum_{r\in R}\sum_{s\in R} B_r w_s^\epsilon \left(\kappa_{rs} Q_r^{1-\alpha}\right)^{-\epsilon}}, \tag{9}$$

where H_n^R is the measure of residents and H_i^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

$$\bar{U} = \delta \left[\sum_{r\in R}\sum_{s\in R} B_r w_s^\epsilon \left(\kappa_{rs} Q_r^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{10}$$

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 land.[22] The final goods production technology is assumed to take the Cobb-Douglas form with unit cost:

$$1 = \frac{1}{A_i^F} w_i^\beta p_i^\gamma Q_i^{1-\beta-\gamma}, \qquad 0<\beta,\gamma<1, \qquad \beta+\gamma=1, \tag{11}$$

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 land under conditions of perfect competition. Again we assume that the production technology takes the Cobb-Douglas form with unit cost:

$$p_i = \frac{1}{A_i^I} w_i^\mu Q_i^{1-\mu}, \qquad 0<\mu<1, \tag{12}$$

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

Using the non-traded services production technology (12), the unit cost function for the final good can be re-written in the following form:

$$1 = \frac{1}{A_i} w_i^{\tilde{\beta}} Q_i^{1-\tilde{\beta}}, \qquad A_i \equiv A_i^F \left(A_i^I\right)^\gamma, \tag{13}$$
$$\tilde{\beta} \equiv \beta+\gamma\mu, \qquad 1-\tilde{\beta} = (1-\beta-\gamma)+\gamma(1-\mu) = 1-(\beta+\gamma\mu), \qquad 0<\tilde{\beta}<1,$$

where A_i is a composite measure of productivity and $\tilde{\beta}$ is a composite measure of labor intensity for the final goods and non-traded services sectors as a whole.

Re-arranging equation (13), we obtain the following key implication of profit maximization and zero profits for each location with positive production:

$$w_i = A_i^{1/\tilde{\beta}} Q_i^{-(1-\tilde{\beta})/\tilde{\beta}}. \tag{14}$$

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 land (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 land.

[22] This formulation models non-traded services as an input into the production of goods. In the 19th and early-20th centuries, London was one of the main centers of manufacturing activity in the United Kingdom, with manufacturing accounting for over 25 percent of employment in Greater London in the population census of 1911.

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:

$$H_i^M = \sum_{n \in R} \pi_{ni|n} H_n^R, \tag{15}$$

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:

$$\pi_{ni|n}^R = \frac{\pi_{ni}}{\pi_n^R} = \frac{(w_i/\kappa_{ni})^\epsilon}{\sum_{s \in R} (w_s/\kappa_{ns})^\epsilon}. \tag{16}$$

Therefore, each location faces an upward-sloping supply curve for workers that increases with its wage relative to the wage in all other locations, and 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 land equals payments to land used residentially plus payments to land used commercially:

$$\mathbb{Q}_n = Q_n L_n = (1-\alpha) v_n H_n^R + \left(\frac{1-\tilde{\beta}}{\tilde{\beta}} \right) w_n H_n^M, \tag{17}$$

where $\mathbb{Q}_n = Q_n L_n$ is the total value of land rents (rateable value); L_n is the supply of land; v_n is average residential income; and we have used the Cobb-Douglas production technologies for the final good and non-traded services. 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:

$$v_n = \sum_{i \in R} \pi_{ni|n}^R w_i. \tag{18}$$

5.4 General Equilibrium

We begin by characterizing the properties of a benchmark version of the model in which location properties are exogenous, before relaxing this assumption to introduce endogenous agglomeration forces. 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). These five components of the equilibrium vector are determined by the following system of five equations: the workplace choice probabilities ($\boldsymbol{\pi}^M$ in equation (9)), the residential choice probabilities ($\boldsymbol{\pi}^R$ in equation (9)), land market clearing (17), profit maximization and zero profits (14), and population mobility (10).

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

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 of exogenous and strictly positive values of productivity (A_n) and amenities (B_n) ensures that all location

are 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 values of productivity and amenities.

5.5 Production and Residential Externalities

Having established the properties of the model with exogenous location characteristics, we now introduce endogenous agglomeration forces. We allow productivity to depend on production fundamentals (a_i) and externalities (E_i^L). Production fundamentals capture features of physical geography that make a location more or less productive independently of the surrounding density of economic activity (e.g. access to natural water). Production externalities capture agglomeration forces in production and are model according to the standard approach in urban economics as the distance weighted sum of surrounding employment. We assume that production externalities depend on total employment (the sum of employment for the final good and non-traded services) and have the same functional form for both of these economic activities. Under these assumptions, our composite measure of productivity can be expressed as the following function of production fundamentals (a_i) and externalities (E_i^M):

$$A_i = a_i \left(E_i^M\right)^{\lambda}, \qquad E_i^M \equiv \left[\sum_{s \in R} \frac{H_s^M}{d_{is}^M}\right], \tag{19}$$

where d_{is}^M is an iceberg cost that increases with distance and λ parameterizes the relative importance of externalities for productivity.

We model the externalities in workers' residential choices analogously to those in firms' production choices. We allow residential amenities to depend on residential fundamentals (b_n) and residential externalities (E_n^R). Residential fundamentals capture features of physical geography that make a location a more or less attractive place to live independently of the surrounding density of economic activity (e.g. scenic views). Residential externalities capture other agglomeration forces in residential decisions and are modeled analogously to those in production. In particular, we assume that residential externalities are the distance-weighted sum of surrounding residents, so that residential amenities can be expressed as follows:

$$B_n = b_n \left(E_n^R\right)^{\eta}, \qquad E_n^R \equiv \left[\sum_{r \in R} \frac{H_r^M}{d_{nr}^R}\right], \tag{20}$$

where d_{nr}^R is an iceberg cost that increases with distance and η parameterizes the relative importance of externalities for amenities.

The introduction of these agglomeration forces generates the potential for multiple equilibria if they are sufficiently strong relative to the exogenous differences in characteristics across locations. An important feature of our empirical approach is that it explicitly addresses the potential for multiple equilibria, as discussed further in the structural estimation of the model below.

6 Gravity Decomposition

We now use the gravity equation for commuting flows together with the remaining structure of the model to develop our new decomposition of internal city structure into the contributions of productivity, amenities and commuting costs. We implement this gravity decomposition in our empirical analysis below for 1921, when we have data on

bilateral commuting flows between locations. In particular, given known values of the model's parameters $\{\alpha, \tilde{\beta}, \epsilon\}$, and observed data on bilateral commuting flows, rateable values and land area $\{H_{ni}, \mathbb{Q}_n, L_n\}$, we show that the model can be inverted to recover unique values of commuting costs, productivity and amenities $\{\kappa_{ni}, A_n, B_n\}$ (up to a normalization) that exactly rationalize the observed data as an equilibrium of the model.

To implement this decomposition, we assume central values for the model's parameters from the existing empirical literature. 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 the findings of Valentinyi and Herrendorf (2008). We set the Fréchet shape parameter determining the heterogeneity in workers idiosyncratic preferences across locations equal to 3, which is consistent with Bryan and Morten (2015) and Galle, Yi, and Rodriguez-Clare (2015). We derive our decomposition in terms of the overall values of productivity and amenities (including both location fundamentals and externalities). Therefore, we can undertake these decompositions without specifying the parameters determining the strength of production and residential externalities.

In step one, we use observed bilateral commuting flows (H_{ni}) to solve for the unconditional commuting probability (π_{ni}), employment (H_i^M), residents H_n^R, and the conditional commuting probability $(\pi_{ni|n})$:

$$\pi_{ni} = \frac{H_{ni}}{H}, \qquad H_i^M = \sum_{n \in R} H_{ni}, \qquad H_n^R = \sum_{i \in R} H_{ni}, \qquad \pi_{ni|n} = \frac{H_{ni}}{H_n^R}. \tag{21}$$

In step two, we use observed employment (H_n^M), residents (H_n^R) and rateable values $(\mathbb{Q}_n)$, together with the equations for land market clearing (17) and expected residential income (18), to solve for the unique vector of wages in each location that solves the following system of equations:

$$\mathbb{Q}_n = (1 - \alpha) \left[\sum_{i \in R} \pi_{ni|n} w_i \right] H_n^R + \left(\frac{1 - \tilde{\beta}}{\tilde{\beta}} \right) w_n H_n^M. \tag{22}$$

In step three, we use these solutions for wages (w_n) and the conditional commuting probabilities (16) to recover implied commuting costs. In particular, dividing the conditional commuting probability for residence n and workplace i $(\pi_{ni|n})$ by residence n's conditional commuting probability with itself $(\pi_{nn|n})$, we obtain:

$$\left(\frac{\kappa_{ni}}{\kappa_{nn}} \right)^{-\epsilon} = \left(\frac{w_i}{w_n} \right)^{\epsilon} \left(\frac{\pi_{ni|n}}{\pi_{nn|n}} \right), \tag{23}$$

where the left-hand side of equation (23) is inversely related to commuting costs and can be interpreted as a measure of each location n's bilateral ease of commuting to other locations $i \neq n$ relative to its ease of commuting to itself.

In the data, all locations have positive commuting flows to themselves $(\pi_{nn|n} > 0)$, but may have zero commuting flows to some other locations $i \neq n$ $(\pi_{ni|n} = 0$ for some $i \neq n)$. These zero commuting flows are rationalized in the model by infinite commuting costs to these locations $(\kappa_{ni} \to \infty$ for some $i \neq n$ and $\kappa_{nn} < \infty)$, which implies a zero bilateral ease of commuting $((\kappa_{ni}/\kappa_{nn})^{-\epsilon} = 0)$ in equation (23). Under the additional assumption that each location faces no commuting costs with itself $(\kappa_{nn} = 1)$, equation (23) can be used to recover the level of bilateral commuting costs (κ_{ni}) between each pair of residence n and workplace i, where we allow for the possibility that commuting costs are not symmetric $(\kappa_{ni} \neq \kappa_{in})$. We use this additional assumption of no own commuting costs $(\kappa_{nn} = 1)$ in step five below, but relax it in our later structural estimation of the model.

In step four, we use the observed rateable values and land area $\{\mathbb{Q}_n, L_n\}$ to solve for the implied land rents:

$$Q_n = \frac{\mathbb{Q}_n}{L_n}. \tag{24}$$

In step five, we use the observed residents (H_n^R) and land rents (Q_n), and our solutions for wages (w_n) and commuting costs (κ_{ni}) together with the residents probability (π_n^R) in equation (9), to solve for relative amenities (B_n):

$$\frac{\pi_n^R}{\pi^{R*}} = \left(\frac{B_n}{B^*}\right)\left(\frac{Q_n}{Q^*}\right)^{-\epsilon(1-\alpha)}\left(\frac{MA_n^R}{MA^{R*}}\right)^{\epsilon}, \qquad MA_n^R \equiv \left[\sum_{s\in R}\left(w_s/\kappa_{ns}\right)^{\epsilon}\right]^{\frac{1}{\epsilon}} \tag{25}$$

where a superscript asterisk denotes a geometric mean such that $B^* = \left(\prod_{n\in R} B_n\right)^{1/N}$, where $N = |R|$ is the number of locations; MA_n^R is a measure of residents market access, which depends on wages and commuting costs in all surrounding locations; we use units in which to measure amenities such that the geometric mean of amenities is equal to one ($B^* = 1$).

Intuitively, a relatively high residence probability (π_n^R/π^{R*}) in equation (25) must be explained in the model either in terms of relatively attractive amenities (B_n/B^*), relatively low land prices (Q_n/Q^*), or relatively high access to high-wage employment opportunities (MA_n^R/MA^{R*}). Therefore, given the observed residence probabilities (π_n^R), the observed land prices (Q_n) and our solutions for commuting costs (κ_i) and wages (w_n), we can back out from the structure of the model the implied amenities (B_n) that rationalize the observed data as an equilibrium of the model (up to a choice of units in which to measure amenities).

In step six, we use observed land rents (Q_n) and our solutions for wages (w_n), together with the zero-profit condition (14), to solve for relative productivities (A_i):

$$\frac{A_i}{A^*} = \left(\frac{w_i}{w^*}\right)^{\beta}\left(\frac{Q_i}{Q^*}\right)^{1-\beta}, \tag{26}$$

where a superscript asterisk again denotes a geometric mean such that $A^* = \left(\prod_{i\in R} A_i\right)^{1/N}$. Intuitively, a relatively high wage (w_i) and land price (Q_i) must be offset in equilibrium by a relatively high productivity (A_i), in order for the zero-profit condition for positive production and employment to be satisfied.

In Figure 15, we display the key components of the wage equation (22) for each borough in Greater London in 1921. In the top-left panel, we show that there is a strong positively and statistically significant relationship between land rents and workers. In the top-right panel, we find a similar positive and statistically significant relationship for residents and land rents, although the relationship is less tight than for workers. Consistent with these findings, the bottom-left panel shows that residents and workers are strongly positively correlated with one another. Workers are more spatially concentrated than residents, as reflected in the small number of boroughs that have particularly high numbers of workers relative to residents. In the bottom-right panel, we show that there is a relatively flat relationship between wages and land rents. In the absence of productivity differences across locations, the zero-profit condition (14) would imply a tight relationship between wages and land rents. Therefore the bottom-right panel is already suggestive of substantial differences in productivity between the boroughs of Greater London.

In Figure 16, we show the log relative bilateral ease of commuting from equation (23) in 1921 against the log bilateral Euclidean distance between the centroids of boroughs. As discussed above, this measure captures each borough n's ease of commuting to another borough $i \neq n$ relative to its ease of commuting to itself. We omit the bilateral pairs for the own location $n = i$, since the relative bilateral ease of commuting in equation (23) is necessarily equal to one for these pairs. We also omit bilateral pairs $i \neq n$ with zero commuting flows, since the relative bilateral ease of commuting in equation (23) is necessarily equal to zero for these pairs. Consistent with commuting costs increasing with distance, we find a strong negative relationship between the bilateral ease of commuting and bilateral distance, with

a statistically significant elasticity of -1.31. However, this relationship is far from perfect, with a regression R-squared of only 0.15. There are two main potential sources of this imperfect correlation. First, boroughs differ in terms of their own area and commuting costs with themselves, which generates variation in commuting costs to other locations relative to those to themselves. Second, boroughs differ substantially in their 1921 railway connections, which induces heterogeneity in commuting costs to other locations for a given distance between those locations.

In Figure 17, we graph relative amenities from the residential choice probabilities in equation (25) and relative productivities from the zero-profit condition (26). These location characteristics play the key role in the model of rationalizing observed employment and residents given the observed land prices and commuting flows in 1921. In the left-panel, we show a positive and statistically significant relationship between relative amenities and land rents. This relationship is not perfect, because boroughs differ from one another in terms of their access to employment opportunities, as summarized by residential market access (MA_n^R) in equation (25). In the right-panel, we show an even stronger positive and statistically significant relationship between productivity and land prices. The tightness of this relationship reflects the fact that wages vary less than residential market access. Hence high land prices must be offset by high productivity in order for zero profits in production to be satisfied. Amenities vary more than productivity, because of the heterogeneity in worker preferences, as captured by the Fréchet shape parameter ϵ. An increase in this parameter magnifies the impact of differences in land prices (Q_n) on residential choice probabilities in equation (25), thereby inducing larger differences in amenities (B_n).

7 Structural Estimation

We now develop our new structural estimation approach, which again exploits the model's gravity equation predictions. In particular, we use the observed data on bilateral commuting flows for 1921 to undertake counterfactuals for changes in transport technology, such as the development of the steam engine. Using observed data on population, land prices and the transport network back to the nineteenth century, we show how these counterfactuals can be used to structurally estimate the parameters of the model. We examine the extent to which the model can explain the observed changes in city structure, both qualitatively and quantitatively, for the estimated parameters determining the strength of agglomeration and dispersion forces.

Of the model's parameters, the share of consumer expenditure on residential land ($1 - \alpha$) and the share of expenditure on commercial land for the composite production sector ($1 - \tilde{\beta}$) are hard to determine from our data, because information on consumer expenditures and factor payments for each borough is not available over our long historical time period. Therefore, we assume the central values from the existing empirical literature of $1 - \alpha = 0.25$ and $1 - \tilde{\beta} = 0.20$, as discussed in the previous section.

We structurally estimate the remaining parameters of the model, which include the Fréchet shape parameter for the heterogeneity of idiosyncratic amenities (ϵ), the relative importance of production externalities (λ), the relative importance of residential externalities (η), and the determinants of commuting costs (κ_{ni}), the decay of production externalities (d_{ni}^M), and the decay of residential externalities (d_{ni}^R). We model these three spatial frictions in a similar way using a least-cost-path measure 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 crossed by a steam railway. We normalize the cost for points intersected by a

steam railway to one ($\delta^{\text{Rail}} = 1$) and estimate the relative cost for points not connected by a steam railway (δ^{Land}). We measure the transport cost between a pair of centroids of boroughs as the sum of the costs of traveling between each of the intermediate points on the grid in between those centroids along the least-cost path. These transport costs depend on the set of grid points connected by the railway network in a given year ($\mathbb{I}_t^{\text{Rail}}$) and the relative costs of traveling by rail and land (as captured by the vector of costs $\boldsymbol{\delta}$, which we allow to differ across the three spatial frictions). We indicate the role of these determinants by writing the commuting cost as $\kappa_{nit}(\mathbb{I}_t^{\text{Rail}}, \boldsymbol{\delta_\kappa})$, the decay of production externalities as $d_{nit}^M(\mathbb{I}_t^{\text{Rail}}, \boldsymbol{\delta_M})$, and the decay of residential externalities as $d_{nit}^R(\mathbb{I}_t^{\text{Rail}}, \boldsymbol{\delta_R})$, where we have made explicit the dependence of the spatial frictions on time through the evolution of the railway network. The vector of parameters to be estimated is thus $\{\epsilon, \lambda, \eta, \delta_\kappa^{\text{Land}}, \delta_M^{\text{Land}}, \delta_R^{\text{Land}}\}$.

We construct moment conditions to estimate the model's parameters by undertaking counterfactuals for the effect of the change in the railway network on the organization of economic activity within the city for a given value of the parameter vector. We use the recursive structure of the model to compute these moment conditions in four steps. First, suppose that we observe the railway network in 1921 ($\mathbb{I}_{1921}^{\text{Rail}}$) and in an earlier year ($\mathbb{I}_{1921-\tau}^{\text{Rail}}$). Given this observed change in the railway network and the transport cost weights ($\boldsymbol{\delta}$), we can compute the counterfactual change in commuting costs ($\hat{\kappa}_{ni}(\boldsymbol{\delta_\kappa})$), the decay of production externalities ($\hat{d}_{ni}^M(\boldsymbol{\kappa_M})$) and the decay of residential externalities ($\hat{d}_{ni}^R(\boldsymbol{\delta_R})$):

$$\hat{\kappa}_{ni}(\boldsymbol{\delta_\kappa}) = \frac{\kappa_{ni1921-\tau}(\mathbb{I}_{1921-\tau}^{\text{Rail}}, \boldsymbol{\delta_\kappa})}{\kappa_{ni1921}(\mathbb{I}_{1921}^{\text{Rail}}, \boldsymbol{\delta_\kappa})}, \tag{27}$$

$$\hat{d}_{ni}^M(\boldsymbol{\delta_M}) = \frac{d_{ni1921-\tau}^M(\mathbb{I}_{1921-\tau}^{\text{Rail}}, \boldsymbol{\delta_M})}{d_{ni1921}^M(\mathbb{I}_{1921}^{\text{Rail}}, \boldsymbol{\delta_M})}, \tag{28}$$

$$\hat{d}_{ni}^R(\boldsymbol{\delta_R}) = \frac{d_{ni1921-\tau}^R(\mathbb{I}_{1921-\tau}^{\text{Rail}}, \boldsymbol{\delta_R})}{d_{ni1921}^R(\mathbb{I}_{1921}^{\text{Rail}}, \boldsymbol{\delta_R})}, \tag{29}$$

where we use a hat to denote the value of a variable in an earlier year relative to that in 1921: $\hat{x} = x_{1921-\tau}/x_{1921}$.

Second, we use the counterfactual change in commuting costs from equation (27) together with the land market clearing condition and equilibrium in commuting markets to solve for the counterfactual change in wages ($\hat{w}_n$). From the land market clearing condition (17), we have

$$\hat{\mathbb{Q}}_n \mathbb{Q}_n = (1-\alpha)\hat{v_n} v_n \hat{H}_n^R H_n^R + \left(\frac{1-\tilde{\beta}}{\tilde{\beta}}\right) \hat{w}_n w_n \hat{H}_n^M H_n^M, \tag{30}$$

where we observe or can solve for the initial equilibrium values of $\{\mathbb{Q}_n, H_n^M, H_n^R, w_n, v_n\}$ in our baseline year of 1921, as shown in the previous section.

We also observe the historical changes in rateable values and population for years before 1921. We use the historical population changes to construct historical changes in employment by residence ($\hat{H}_n^M$) by assuming the same ratio of employment by residence to population in each borough as in 1921. Given these observed changes in rateable values ($\hat{\mathbb{Q}}_n$) and residents ($\hat{H}_n^R$), there are three unobserved changes for each borough in the land market clearing condition in equation (30): wages ($\hat{w}_n$), expected residential income ($\hat{v}_n$) and employment ($\hat{H}_n^M$). However, we now use the conditions for commuter market equilibrium to show that the unobserved changes in expected residential income ($\hat{v}_n$) and employment ($\hat{H}_n^M$) can be written in terms of the observables $\{\pi_{ni|n}, \mathbb{Q}_n, H_n^M, H_n^R, w_n, v_n, \hat{\mathbb{Q}}_n, \hat{H}_n^R\}$, the counterfactual change in commuting costs $\{\hat{\kappa}_{ni}(\boldsymbol{\delta_\kappa})\}$ and the unobserved change in wages $\{\hat{w}_n\}$. From equation (18),

the change in expected residential income ($\hat{v}_n$) solves the following system of equations:

$$\hat{v}_n v_n = \sum_{i \in R} \frac{\pi_{ni|n} \hat{w}_i^{\epsilon} \hat{\kappa}_{ni} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}}{\sum_{s \in R} \pi_{ns|n} \hat{w}_s^{\epsilon} \hat{\kappa}_{ns} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}} \hat{w}_i w_i. \tag{31}$$

From the commuter market clearing condition (15), the change in employment ($\hat{H}_n^M$) solves the following system of equations:

$$\hat{H}_i^M H_i^M = \sum_{n \in R} \frac{\pi_{ni|n} \hat{w}_i^{\epsilon} \hat{\kappa}_{ni} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}}{\sum_{s \in R} \pi_{ns|n} \hat{w}_s^{\epsilon} \hat{\kappa}_{ns} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}} \hat{H}_n^R H_n^R. \tag{32}$$

Using equations (31) and (32) in land market clearing (30), we obtain a system of equations that can be solved for the unobserved change in wage for each location $\{\hat{w}_n\}$ given the observables $\{\pi_{ni|n}, \mathbb{Q}_n, H_n^M, H_n^R, w_n, \hat{\mathbb{Q}}_n, \hat{H}_n^R\}$:

$$\begin{aligned} \hat{\mathbb{Q}}_n \mathbb{Q}_n = (1-\alpha) \left[\sum_{i \in R} \frac{\pi_{ni|n} \hat{w}_i^{\epsilon} \hat{\kappa}_{ni} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}}{\sum_{s \in R} \pi_{ns|n} \hat{w}_s^{\epsilon} \hat{\kappa}_{ns} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}} \hat{w}_i w_i \right] \hat{H}_n^R H_n^R \\ + \left(\frac{1-\tilde{\beta}}{\tilde{\beta}} \right) \hat{w}_n w_n \left[\sum_{n \in R} \frac{\pi_{ni|n} \hat{w}_i^{\epsilon} \hat{\kappa}_{ni} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}}{\sum_{s \in R} \pi_{ns|n} \hat{w}_s^{\epsilon} \hat{\kappa}_{ns} (\boldsymbol{\delta}_{\boldsymbol{\kappa}})^{-\epsilon}} \hat{H}_n^R H_n^R \right]. \end{aligned} \tag{33}$$

Using the resulting solutions for changes in wages ($\hat{w}_n$) in equation (32), we can also recover the unobserved changes in employment ($\hat{H}_n^M$) for each location.

Note that the model has a recursive structure, such that the changes in wages ($\hat{w}_n$) and employment ($\hat{H}_n^M$) in the land market clearing condition (33) depend only on the commuting cost parameters ($\boldsymbol{\delta}_{\boldsymbol{\kappa}}$, ϵ), given the observed changes in rateable values ($\hat{\mathbb{Q}}_n$) and residents ($\hat{H}_n^R$). In particular, the model's predictions for impact of the change in the railway network on wages ($\hat{w}_n$) and employment ($\hat{H}_n^M$) are invariant to the parameters determining the strength of agglomeration forces for production (λ, δ_M) and residence (η, δ_R). The intuition for this recursive property of the model is that payments to land are determined by employment, residents and wages. Therefore, given observed changes in land prices and residents, and using the fact that employment can expressed in terms of residents and wages, we can solve for the values that wages must taken in order to be consistent with land market clearing. We estimate the commuting cost parameters ($\boldsymbol{\delta}_{\boldsymbol{\kappa}}$, ϵ) using our bilateral commuting data for 1921. Given these estimates ($\boldsymbol{\delta}_{\boldsymbol{\kappa}}$, ϵ), we solve for predicted changes in wages ($\hat{w}_n$) and employment ($\hat{H}_n^M$) from the land market clearing condition (33). As an overidentification check, we compare the model's predictions for changes in workplace employment ($\hat{H}_n^M$) for the City of London to the observed values of these changes from the Day Censuses for the City of London.

Third, we use these counterfactual changes in wages ($\hat{w}_n$) and the observed changes in land prices ($\hat{Q}_n$), together with the zero-profit condition (14), to recover counterfactual change in productivity for each location ($\hat{A}_n$)

$$\hat{A}_n = \hat{w}_n^{\beta} \hat{Q}_n^{1-\beta}. \tag{34}$$

Therefore, large increases in land prices ($\hat{Q}_n$) and wages ($\hat{w}_n$) must be offset by large increases in productivity in order for zero profits to continue to be made. As this relationship uses only land and commuter market clearing (to solve for wages) and zero-profits, we can determine the change in overall productivity ($\hat{A}_n$) regardless of the relative importance of changes in production fundamentals ($\hat{a}_n$) and production externalities ($\hat{E}_n^M$), and hence regardless of the parameters (λ, δ_M) that determine the relative importance of these two components of productivity.

Fourth, we use our counterfactual changes in wages ($\hat{w}_n$) and the observed changes in residents (H_n^R) and land prices ($\hat{Q}_n$), together with the residential choice probabilities in equation (9), to pin down the counterfactual change in amenities for each location ($\hat{B}_n$)

$$\frac{\hat{B}_n}{\hat{B}^*} = \frac{\hat{H}_n^R}{\hat{H}^{R*}} \left(\frac{\hat{Q}_n}{\hat{Q}^*} \right)^{\epsilon(1-\alpha)} \left(\frac{M \hat{A} R_n}{M \hat{A} R^*} \right)^{-\epsilon}, \tag{35}$$

$$M\hat{A}R_n = \left[\sum_{s \in R} \pi_{ns|n} \hat{w}_n^{\epsilon} \hat{\kappa}_{ns}^{-\epsilon}\right]^{\frac{1}{\epsilon}} \tag{36}$$

where the superscript asterisk again denotes a geometric mean. Therefore, a large increase in residents (H_n^R) in the face of a large increase in land prices (Q_n) and a small increase in residents market access ($M\hat{A}R_n$) must be explained by a large increase in amenities ($\hat{B}_n$). This relationship uses only land and commuter market clearing (to solve for wages) and residential choice probabilities. Hence we can determine the change in overall amenities ($\hat{B}_n$) regardless of the relative importance of changes in residential fundamentals ($\hat{b}_n$) and residential externalities ($\hat{E}_n^R$), and hence regardless of the parameters (η, δ_R) that determine the relative importance of these two components of amenities.

Fifth, we use our specification for productivity (19) to separate the changes in overall productivity ($\hat{A}_n$) into the contributions of changes in production fundamentals ($\hat{a}_n$) and changes in production externalities (E_n^M):

$$\hat{a}_n = \hat{A}_n \left(\hat{E}_n^M\right)^{-\lambda}, \tag{37}$$

$$\hat{E}_n^M = \sum_{i \in R} \left(\frac{\frac{H_i^M}{d_{ni}^M}}{\sum_{i \in R} \frac{H_i^M}{d_{ni}^M}}\right) \frac{\hat{H}_i^M}{\hat{d}_{ni}^M}. \tag{38}$$

Sixth, we use our specification of residential amenities (20) to separate the changes in overall amenities ($\hat{B}_n$) into the contributions of changes in residential fundamentals ($\hat{b}_n$) and changes in residential externalities (E_n^R):

$$\frac{\hat{b}_n}{\hat{b}^*} = \frac{\hat{B}_n}{\hat{B}^*} \left(\frac{\hat{E}_n^R}{\hat{E}^{R*}}\right)^{-\eta}, \tag{39}$$

$$\hat{E}_n^R = \sum_{i \in R} \left(\frac{\frac{H_i^R}{d_{ni}^R}}{\sum_{i \in R} \frac{H_i^R}{d_{ni}^R}}\right) \frac{\hat{H}_i^R}{\hat{d}_{ni}^R}. \tag{40}$$

We estimate the parameters determining the strength of agglomeration forces for production (λ, δ_M) and residence (η, δ_R) using moment conditions in terms of changes in production fundamentals ($\hat{a}_n$) from equation (37) and residential fundamentals ($\hat{b}_n$) from equation (39). These changes in fundamentals are structural residuals that ensure that the model exactly rationalizes the observed changes in residents $\{\hat{H}_n^R\}$ and rateable values $\{\hat{\mathbb{Q}}_n\}$ as an equilibrium of the model. Given the model parameters $\{\epsilon, \lambda, \eta, \delta_\kappa^{\text{Land}}, \delta_L^{\text{Land}}, \delta_R^{\text{Land}}\}$, there is a unique mapping from the observables $\{\pi_{ni|n}, \mathbb{Q}_n, H_n^M, H_n^R, w_n, \hat{\mathbb{Q}}_n, \hat{H}_n^R\}$ to values of the structural residuals $\{\hat{a}_n, \hat{b}_n\}$ irrespective of whether the model exhibits a unique equilibrium or multiple equilibrium. Intuitively, we are able to structurally estimate the model's parameters by conditioning on the observed equilibrium and recovering the unobserved changes in fundamentals, regardless of whether there could have been another (unobserved) equilibrium.

To provide some first evidence on the ability of the model to explain quantitatively the observed changes in the organization of economic activity within Greater London, we undertake counterfactuals for assumed values of the parameters determining commuting costs. In particular, we set the costs of traveling by rail to 1 ($\delta_\kappa^{\text{Rail}} = \delta_M^{\text{Rail}} = \delta_R^{\text{Rail}} = 1$) and the cost of traveling by foot to 7 ($\delta_\kappa^{\text{Land}} = \delta_M^{\text{Land}} = \delta_R^{\text{Land}} = 7$), which corresponds to the relative speed of travel by rail and walking (around 21 versus 3 miles per hour). We set the Fréchet shape parameter determining the heterogeneity in workers idiosyncratic preferences across locations equal to 3, which is consistent with Bryan and Morten (2015) and Galle, Yi, and Rodriguez-Clare (2015). We calibrate the model in our baseline year of 1921, such that the model's predictions for workers, residents and land prices are equal to the observed values in the data, as shown

in the previous section. Using our assumed parameters for commuting costs and the observed changes in the railway network, land prices and residents between 1921 and 1852, we solve for the model's counterfactual predictions for wages and workplace employment for each borough from the land market clearing condition (33).

In Figure 18, we show employment by both workplace and residence for the City of London in 1852 and 1921. Employment by residence in both years and employment by workplace in 1921 are observed in the data. Employment by workplace in 1852 is the model's counterfactual prediction. We compare this counterfactual prediction for workplace employment in 1852 to the observed values of day population (workplace employment) from the Day Censuses for the City of London. For our assumed parameters, we find that the change in transport technology from the arrival of the steam railway goes a long way to explaining the observed changes in the workplace employment in the City of London. Therefore, the change in transport technology, viewed through the lens of our model, is able to successfully explain the increasing specialization of the City of London as a center of employment rather than of residents.

In Figure 19, we show rateable values, employment by workplace, wages and productivity in the City of London in 1852 and 1921. Rateable values in both years and employment by workplace in 1921 are observed in the data. Employment by workplace in 1852 and wages and productivity are the predictions of the model. Each variable is expressed as an index such that its 1921 value is equal to one. From the production technology (11), the proportional change in productivity is a weighted average of the proportional changes in rateable values and the proportional changes in wages. Given our assumed parameters, the observed changes in the data when viewed through our model imply substantial changes in productivity in the City of London, which falls by a factor of more than five going backwards in time from 1921 to 1852. Both rateable values and wages are nominal variables, which scale with goods price inflation. Therefore, some of these changes in rateable values, wages and productivity in Figure 19 reflect inflation. To abstract from this inflation, we also examine the change the City of London's share of rateable value, employment and wages in Greater London. Again from 1921 to 1852, the City of London's rateable value share falls from 9 to 4 percent, while its employment share falls from 11 to 7 percent, and its wagebill share falls from 7 to 4 percent. The fall in the City of London's productivity from 1921 to 1852 is also larger than for most other locations, with 70 of the 99 boroughs having smaller proportional declines in productivity than the City of London. This pattern of results is consistent with agglomeration economies in production, such that the City of London's productivity falls as its density of employment falls in 1852 relative to 1921, and we plan to explore these agglomeration economies in production further in our ongoing work.

In Figure 20, we show rateable values, employment by residence, and amenities in the City of London in 1852 and 1921. Rateable values and employment by residence in both years are observed in the data. Amenities are the predictions of the model. Each variable is again expressed as an index such that its 1921 value is equal to one. Given our assumed parameters, the rise in residents in the City of London going backwards in time to 1852 cannot be fully explained with the model by the observed fall in rateable values, which implies a substantial increase in amenities in the City of London. Again these changes in amenities are influenced by inflation. Furthermore, the changes in amenities with inflation are dramatically larger than the changes in productivity, because the effect of changes in rateable values on amenities is magnified by the Fréchet shape parameter $\epsilon = 3$ in equation (25). The rise in the City of London's amenities from 1921 to 1852 is larger than for most other locations, with only one of the 99 boroughs having a larger proportional increase in amenities than the City of London. This pattern of results is consistent with agglomeration economies in residence, such that the City of London's amenities rise as its density of residents

increases in 1852 relative to 1921, and agglomeration economies in residence further in our ongoing work.

8 Conclusions

We provide new evidence on the role of agglomeration economies and changes in transport technology 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 peoples' 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 of its historic and 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, and a steep increase in employment and land value shares, in the City of London. To better connect these changes in locations' specialization as residence or workplace 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 shaping patterns of specialization by workplace and residence.

To further interpret these empirical findings, we develop a quantitative urban model that incorporates agglomeration forces (in both production and residence) and dispersion forces (through an inelastic supply of land and commuting costs). Methodologically, we show how data on bilateral commuting can be used together with the model's gravity equation for commuting to decompose economic activity into the contributions of commuting costs, productivity and amenities. We also develop a new structural estimation methodology for the model's parameters based on the gravity equation for commuting flows. We undertake counterfactuals backwards in time for the change in transport technology, and recover historical employment by workplace data from our gravity equation data in our baseline year of 1921, together with historical data on land prices and employment by residence. We show that our model can account both qualitatively and quantitatively for the observed changes in the pattern of economic activity for plausible parameter values. Although we do not use any information on employment by workplace before our baseline year in the 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 that the observed changes in the specialization of locations by workplace and residence involve substantial changes in both productivity and amenities, consistent with agglomeration forces in both production and residence.

References

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.

Ball, M., and D. Sunderland (2001): *An Economic History of London 1800-1914*. Routlege, Abingdon.

Barker, T. (1980): "Towards an Historial 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 Affaires: 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.

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.

Davis, M. A., and F. Ortalo-Magné (2011): "Housing Expenditures, Wages, Rents," *Review of Economic Dynamics*, 14(2), 248–261.

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 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.

Galle, S., M. Yi, and A. Rodriguez-Clare (2015): "Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade," University of Calirfornia, 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 (1982): "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.

Jackson, K. T. (1987): *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press, Oxford.

Klaus Desmet, D. K. N., and E. Rossi-Hansberg (2017): "The Geography of Development," *Journal of Political Economy*, forthcoming.

Kynaston, D. (2012): *City of London: The History*. Vintage, London.

London County Council (1907): *London Statistics*. London County Council, London.

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.

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.

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. Duration, 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.

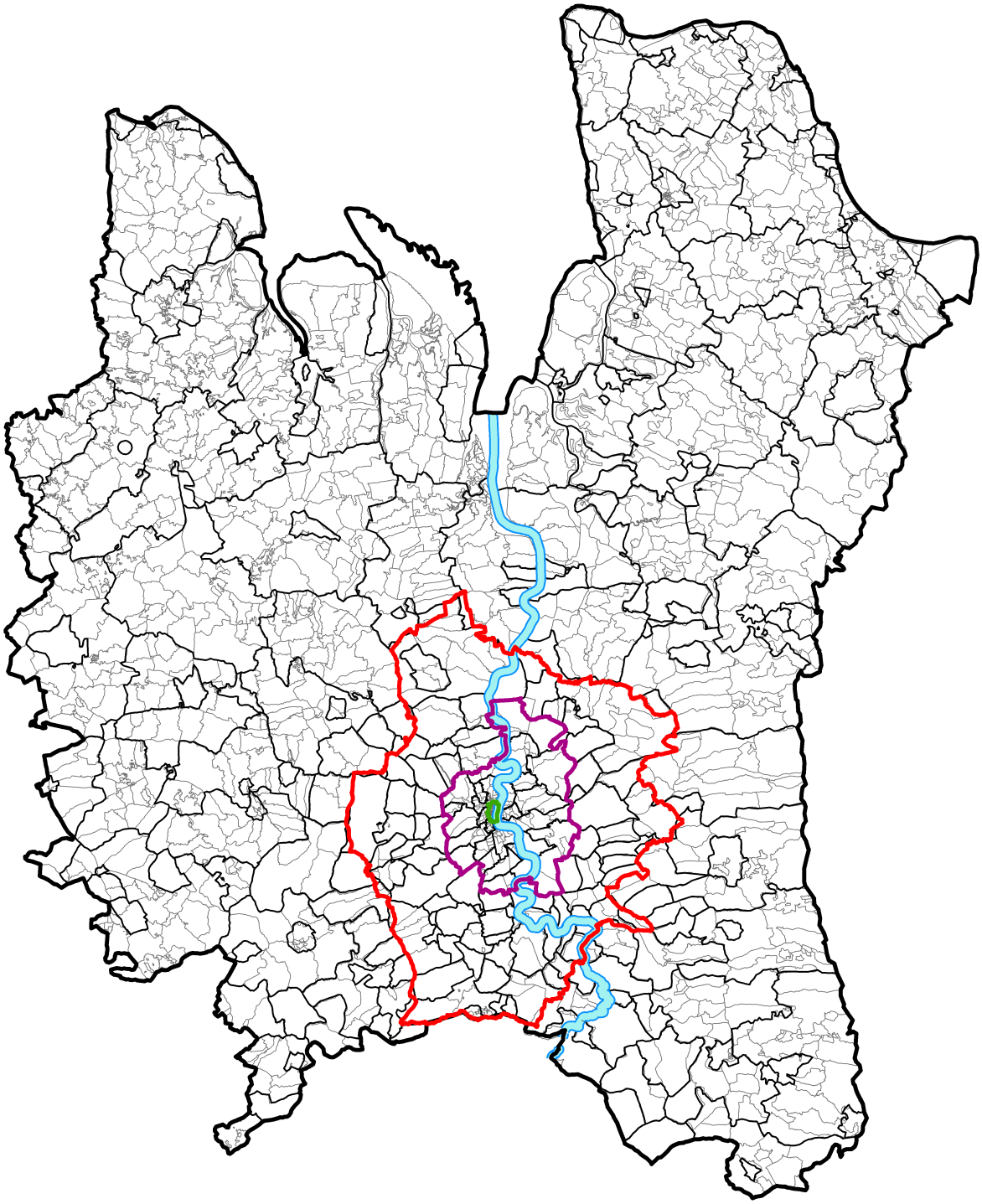
Warner, S. B. (1978): *Streetcar Suburbs: The Process of Growth in Boston, 1870-1900*. Harvard University Press, Cambridge MA.

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.

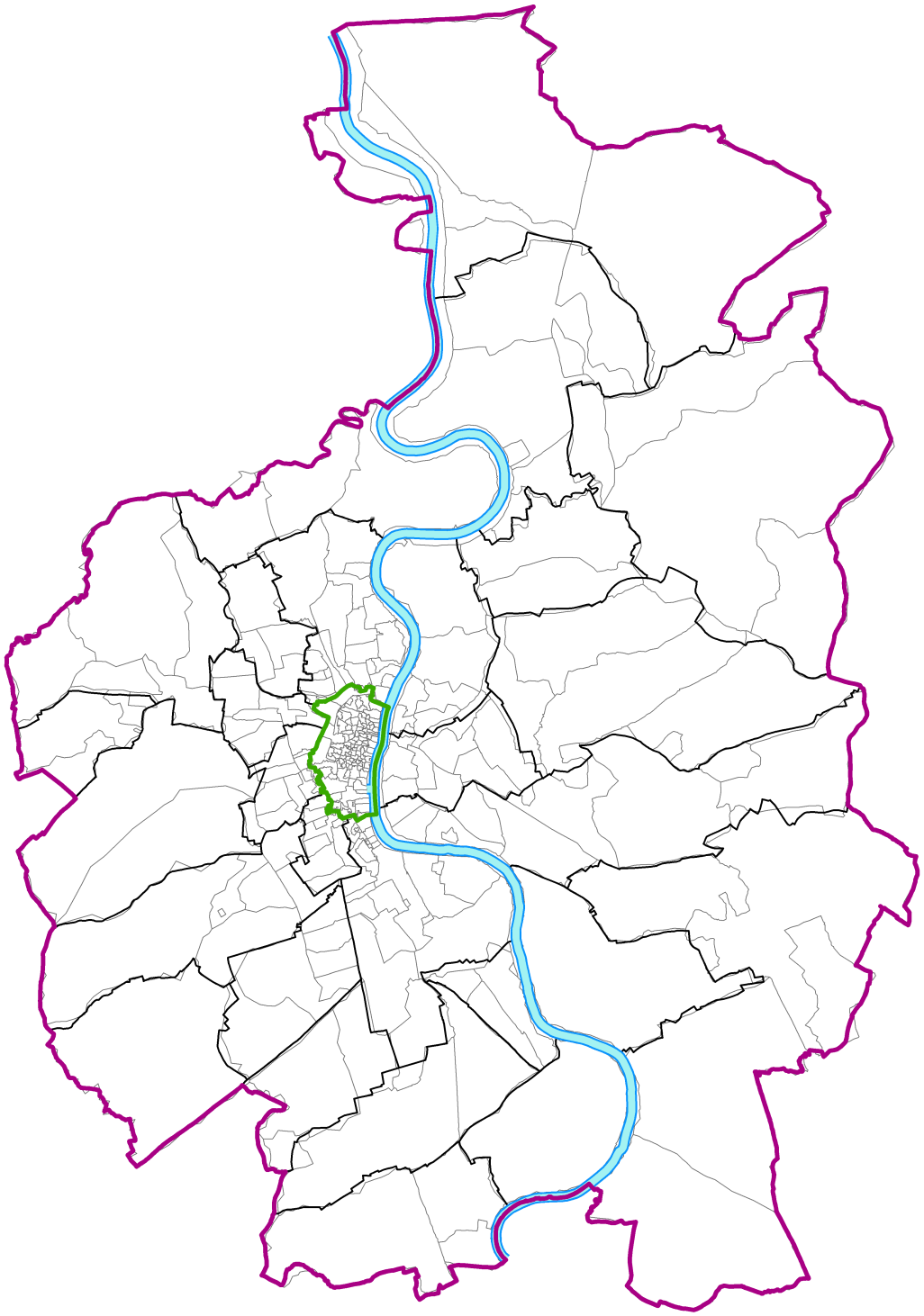
——— (2012): *London in the Eighteenth Century: A Great and Monstrous Thing*. Bodley Head, London.

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: Administrative Boundaries (Zooming In)



Note: London County Council (LCC) (purple outer boundary); City of London (green outer boundary); boroughs (medium black lines); and parishes (medium gray lines)

Figure 3: Population Index Over Time (City of London and Greater London, 1801 equals 1)

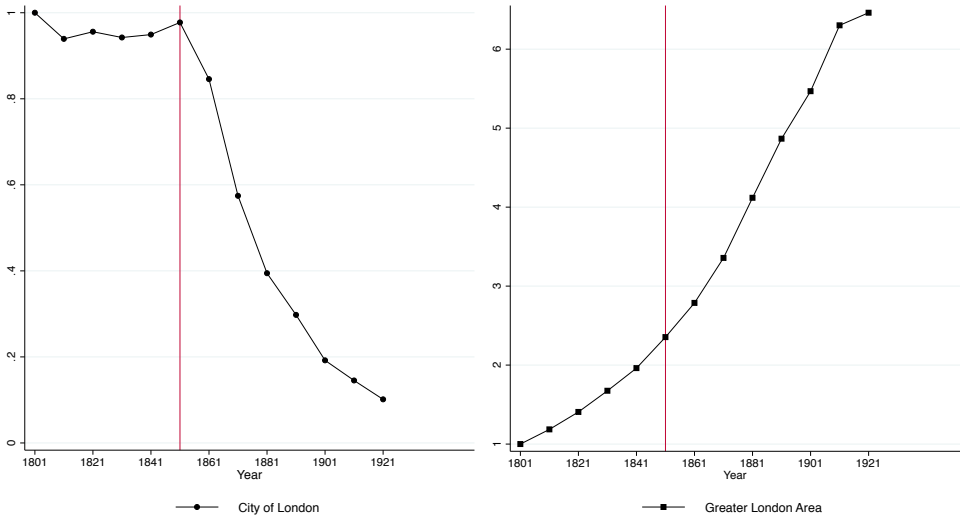


Figure 4: Day and Night Population Over Time (City of London)

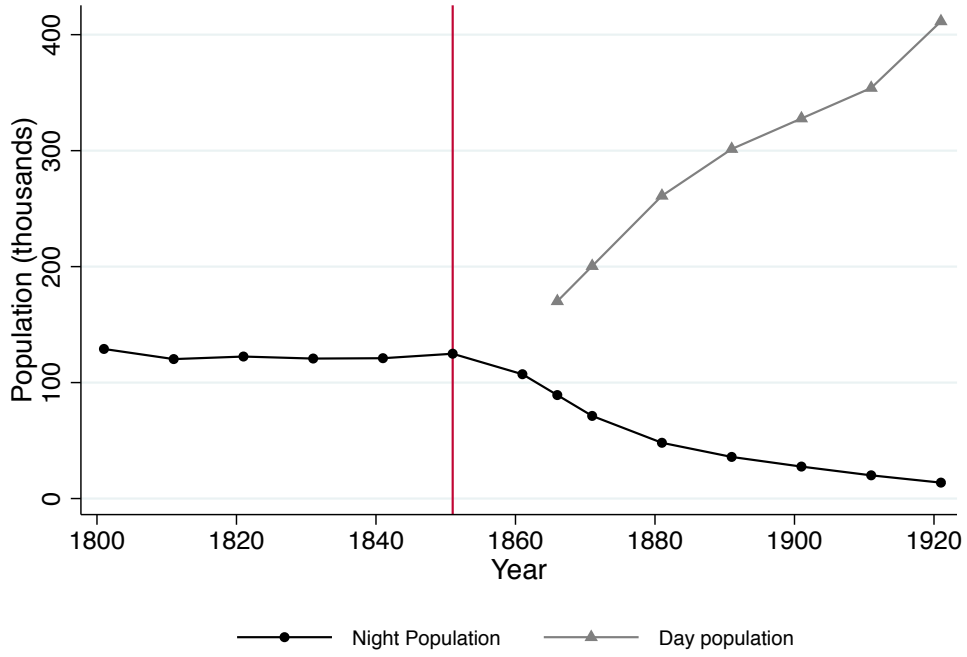
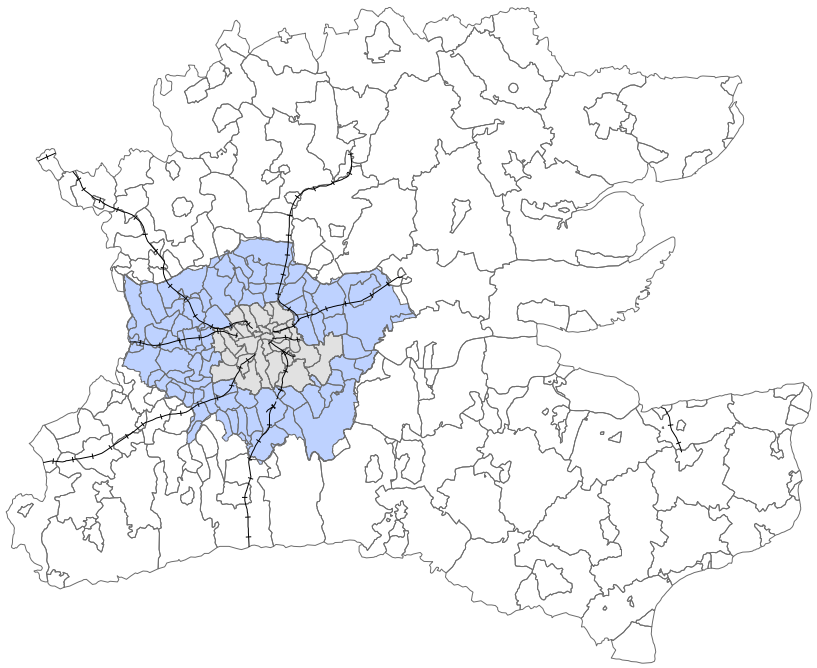


Figure 5: City of London Share of London County Council Rateable Value

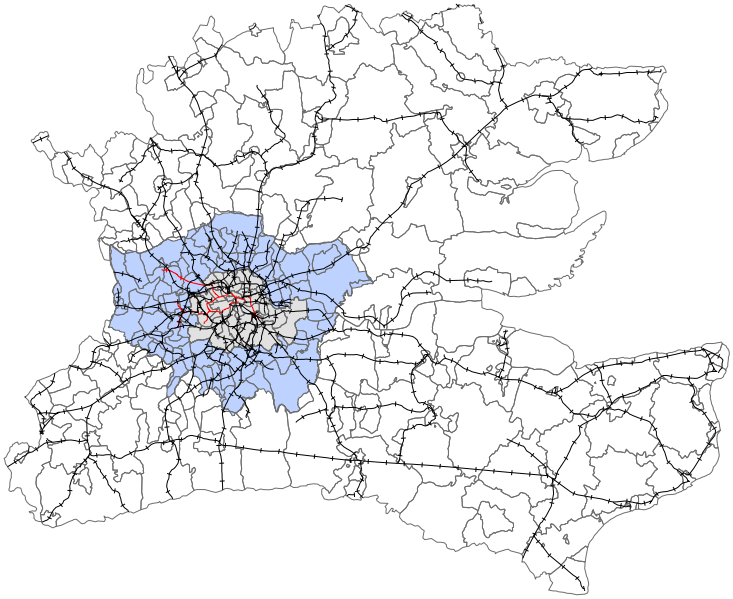


Figure 6: Overground Railway Network 1841



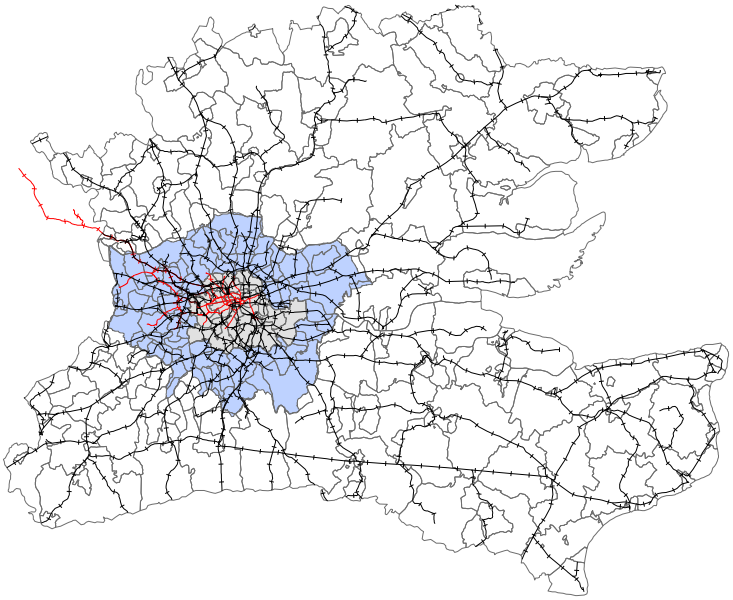
Note: Home counties outside Greater London (white background); Greater London outside London County Council (blue background); London County Council (gray background); overground railway lines shown in black

Figure 7: 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 8: 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 9: Railway Event Study (Baseline Effect for all Parishes)

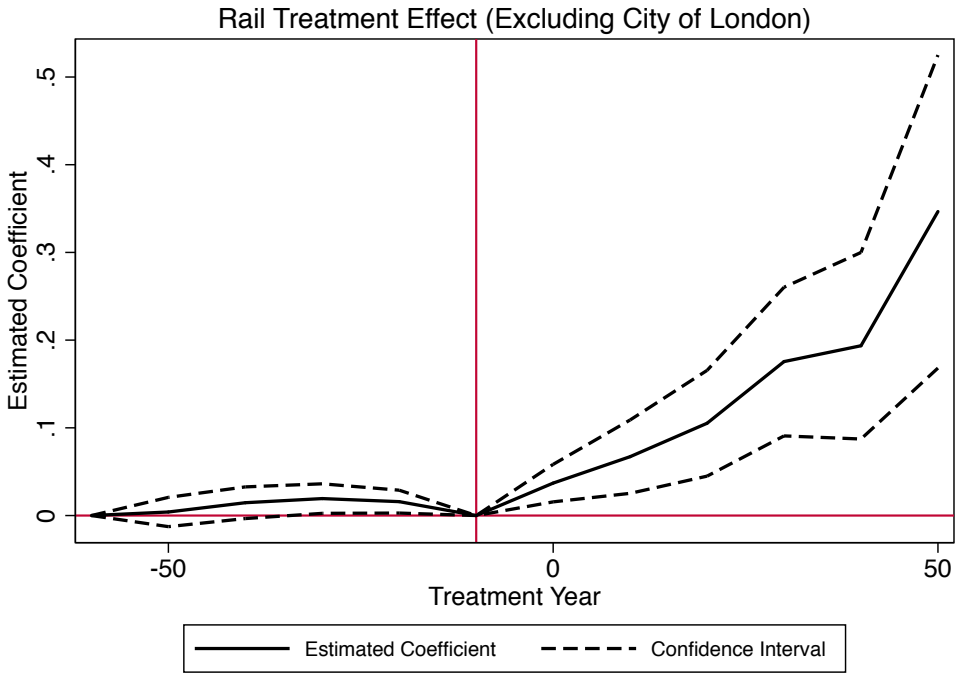


Figure 10: Railway Event Study (Deviation from Baseline Effect for City of London)

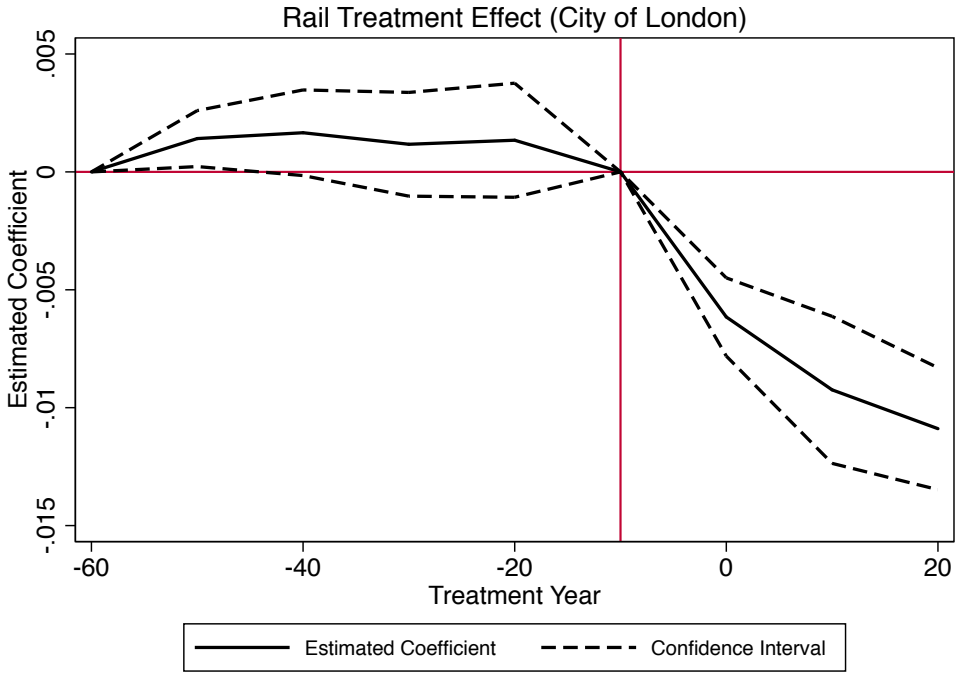
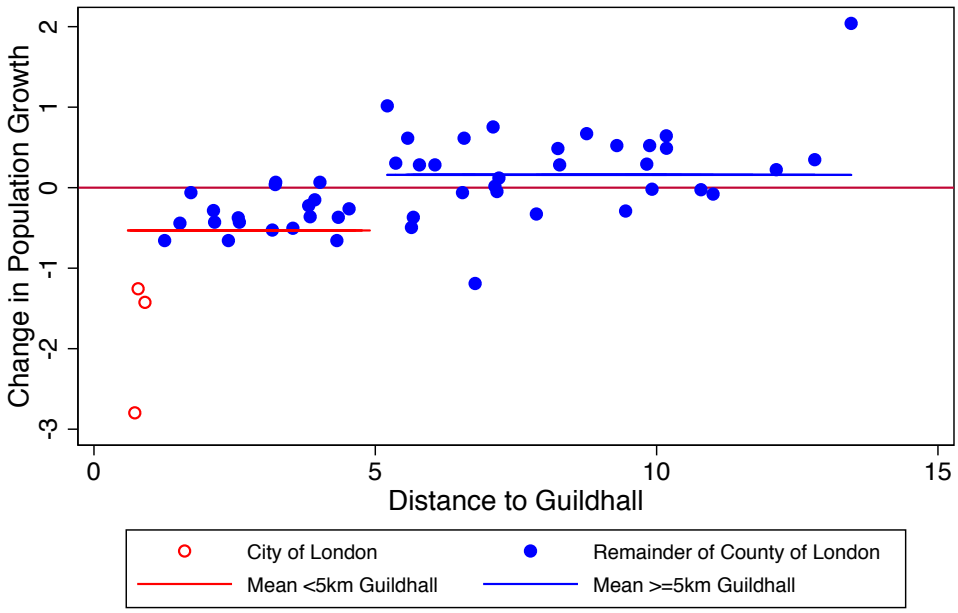


Figure 11: Railway Treatment Heterogeneity Across Parishes



Note: Parishes in the County of London that gain a railway connection during 1801-1911
Population growth in the 30-year period after the arrival of the railway
minus population growth in the 30-year period before the arrival of the railway

Figure 12: Commuting Distances for Barristers (a Type of Lawyer)

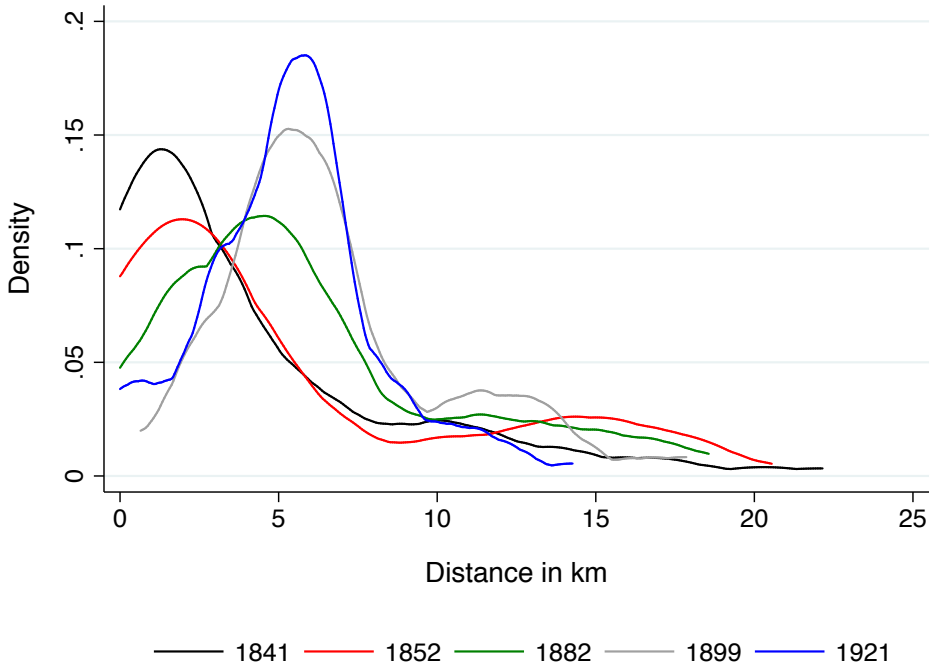


Figure 13: Public Transport Journeys per Head

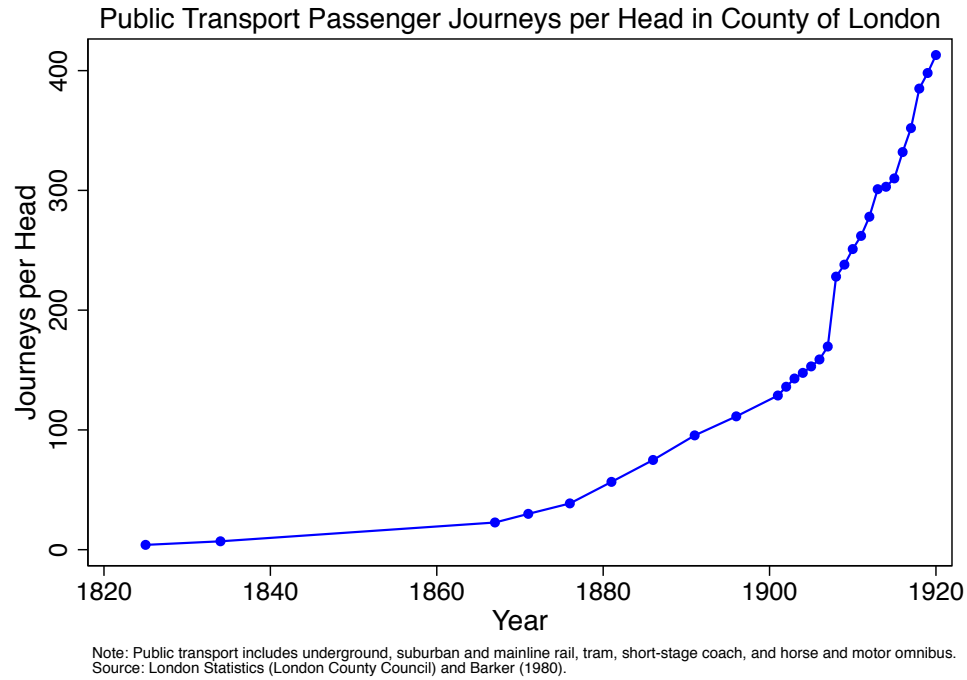


Figure 14: Borough Shares of Workers and Residents in the County of London 1921

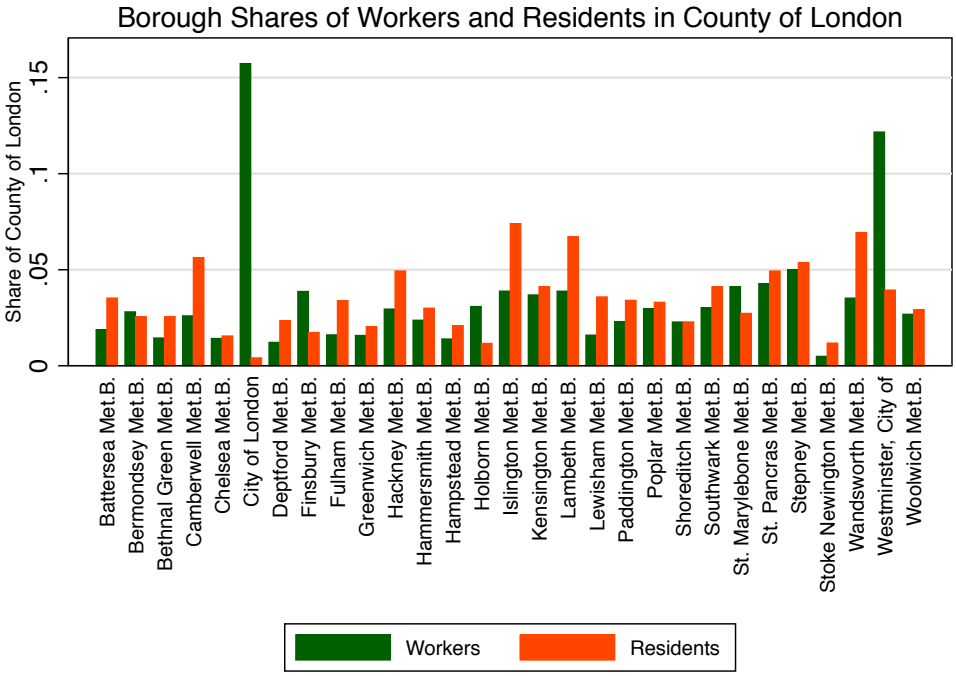
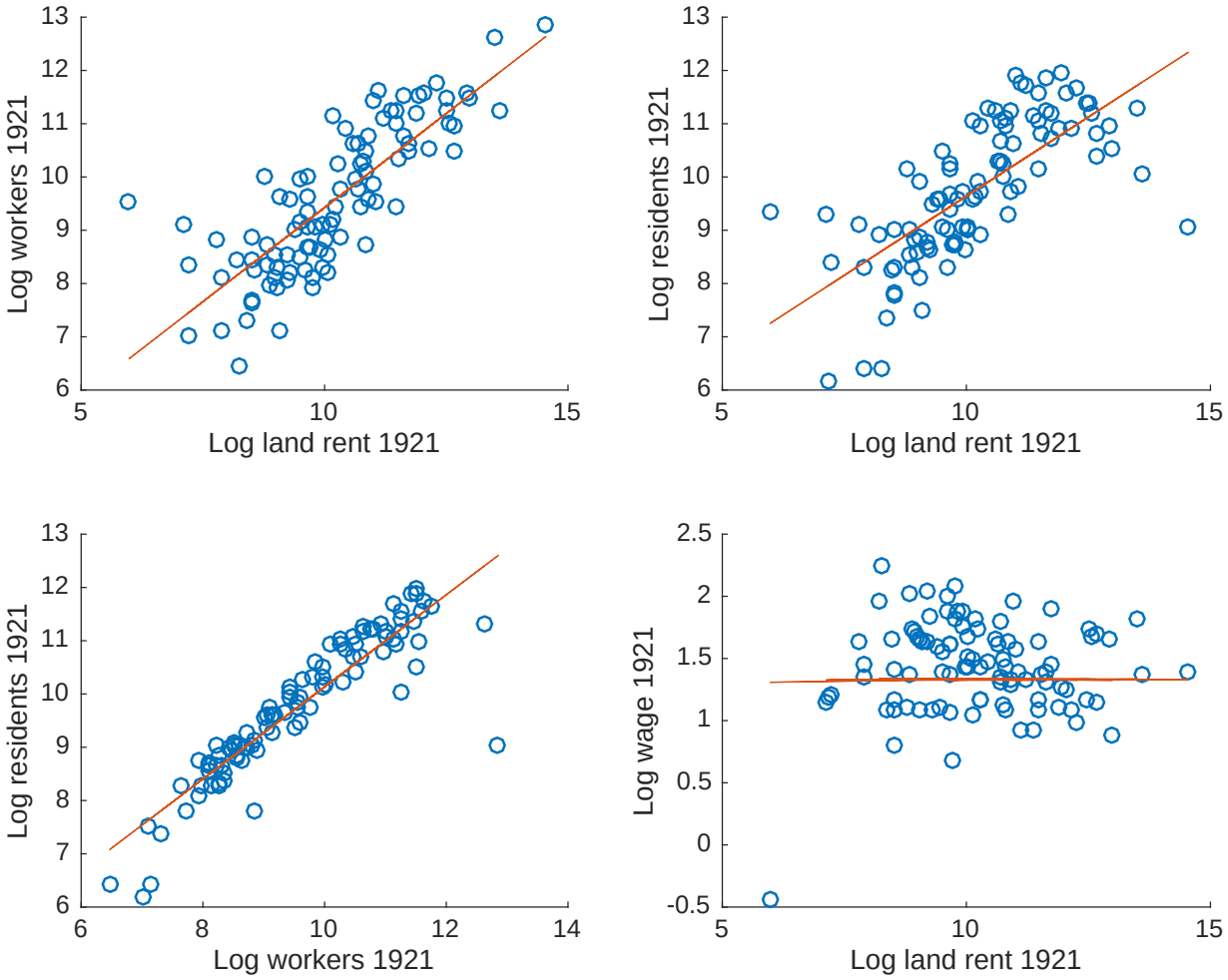
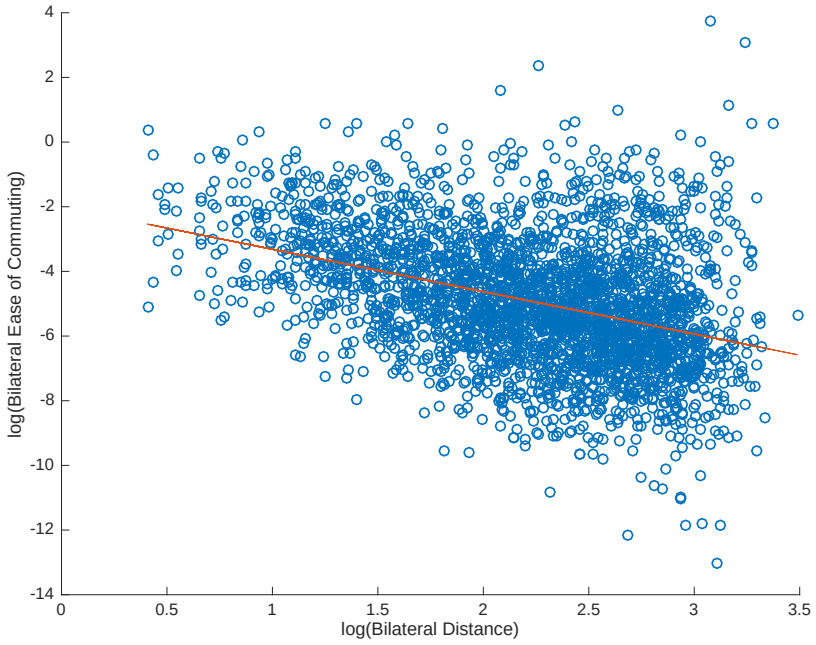


Figure 15: Workers, Residents, Land Prices and Wages in Greater London 1921



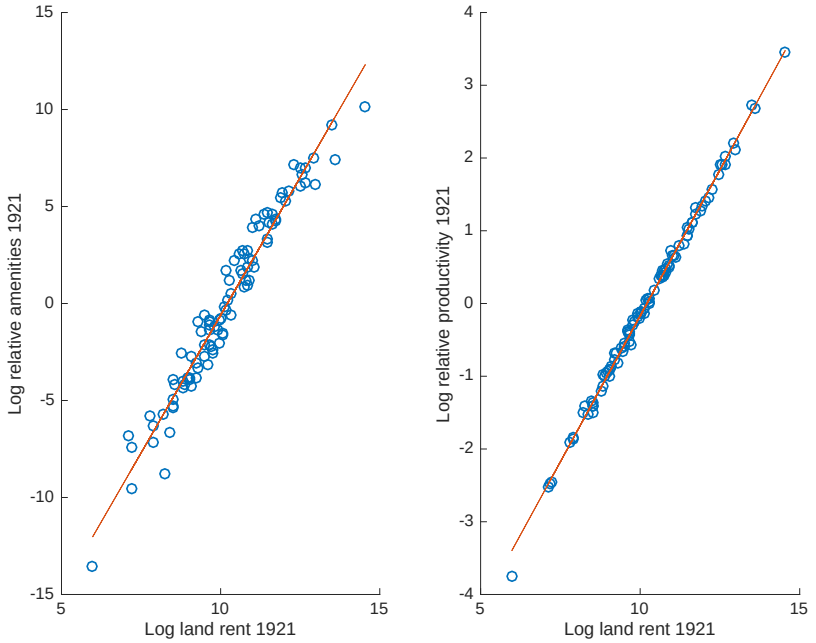
Note: Workers, residents and land rents are observed in the data; wages are solved for using the model (equation (22)).

Figure 16: Relative Bilateral Ease of Commuting and Bilateral Distance Between Boroughs



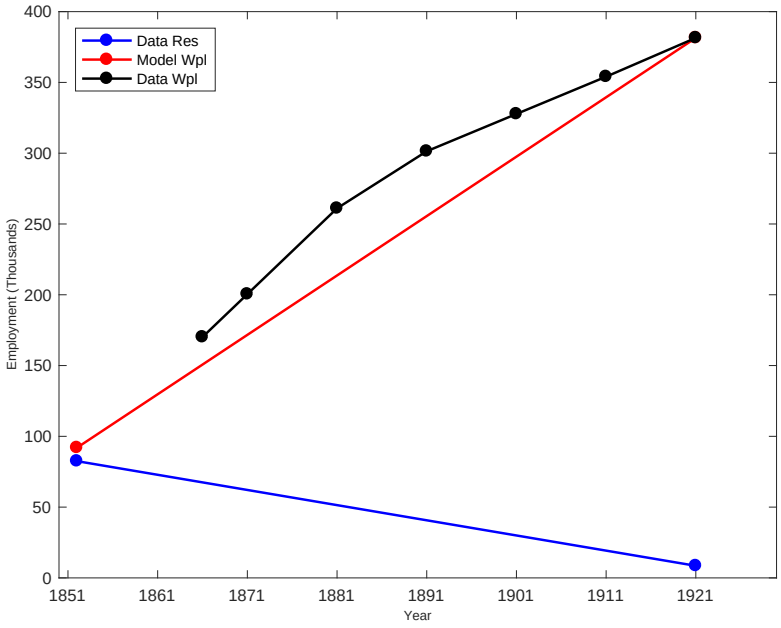
Note: Relative bilateral ease of commuting is recovered from the commuting gravity equation, using observed commuting flows and the solution for wages from the model (equation (23)). Bilateral distance is the Euclidean distance between the centroids of boroughs.

Figure 17: Relative Amenities and Productivities in Greater London 1921



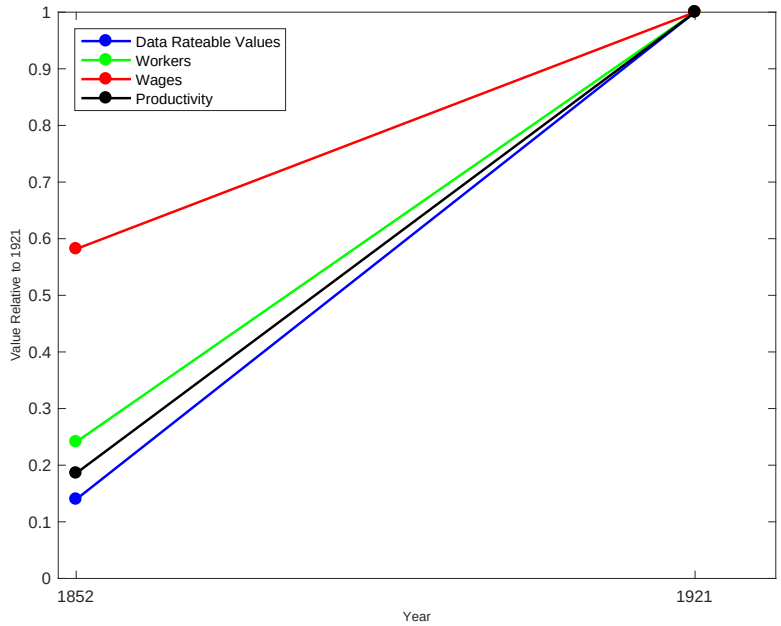
Note: Relative amenities are recovered from the observed residential choice probabilities, observed land prices, and the solutions for wages and commuting costs from the model (equation (25)). Relative productivities are recovered from observed land prices and the solutions for wages from the model (equation (26)).

Figure 18: Workers and Residents in the City of London (Model and Data) 1852-1921



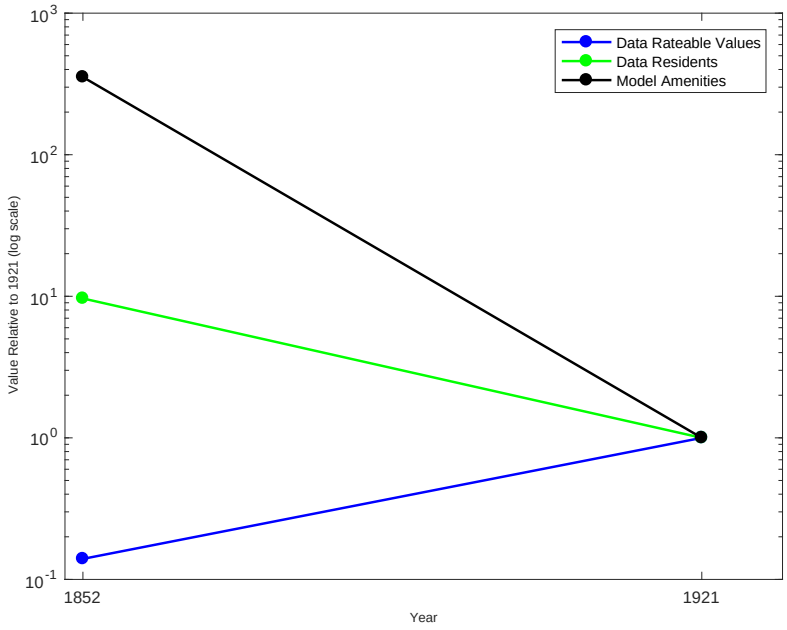
Note: "Data Res" is employment by residence as observed in the data; "Data Wpl" is employment by workplace as observed in the data; "Model Wpl" is the model's prediction for employment by workplace; the model is calibrated to the 1921 values of rateable values, employment by workplace and employment by residence and used to undertake a counterfactual for the change in the railway network back to 1852.

Figure 19: Rateable Values, Employment by Workplace, Wages and Productivity in the City of London 1852-1921 (1921=1)



Note: Rateable Values 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 undertake a counterfactual for the change in the railway network back to 1852. Wages, productivity and workplace employment in 1852 are predictions from the model. Variables expressed as an index where 1921 equals one.

Figure 20: Rateable Values, Employment by Residence, and Amenities 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 undertake a counterfactual for the change in the railway network back to 1852. Amenities are are predictions from the model. Variables expressed as an index where 1921 equals one.