THE MAKING OF THE MODERN METROPOLIS: EVIDENCE FROM LONDON*

Stephan Heblich

Stephen J. Redding†

Daniel M. Sturm

April 15, 2020

Abstract

Using newly-constructed spatially-disaggregated data for London from 1801-1921, we show that the invention of the steam railway led to the first large-scale separation of workplace and residence. We show that a class of quantitative urban models is remarkably successful in explaining this reorganization of economic activity. We structurally estimate one of the models within this class and find substantial agglomeration forces in both production and residence. In counterfactuals, we find that removing the entire railway network reduces the population and the value of land and buildings in London by up to 51.5 and 53.3 percent respectively, and decreases net commuting into the historical center of London by more than 300,000 workers.

Keywords: Agglomeration, Urbanization, Transportation

JEL Classification: O18, R12, R40

*We are grateful to Bristol University, the London School of Economics, Princeton University, and the University of Toronto for research support. Heblich also acknowledges support from the Institute for New Economic Thinking (INET) Grant No. INO15-00025. We would like to thank the editor, four anonymous referees, Victor Couture, Jonathan Dingel, Ed Glaeser, Vernon Henderson, Petra Moser, Leah Platt-Boustan, Will Strange, Claudia Steinwender, Jerry White, Christian Wolmar and conference and seminar participants at Berkeley, Canadian Institute for Advanced Research (CIFAR), Center for Economic Policy Research (CEPR), Columbia, Dartmouth, EIEF Rome, European Economic Association, Fed Board, Geneva, German Economic Association, Harvard, IDC Herzilya, LSE, Marseille, MIT, National Bureau of Economic Research (NBER), Nottingham, Princeton, Singapore, St. Gallen, University College London (UCL), University of Pennsylvannia, Urban Economics Association (UEA), Vienna, Yale, Zoom Urban Seminar, and Zurich for helpful comments. We would like to thank David Green for sharing printed copies of the Henry Poole data and T. Wangyal Shawa for his help with the GIS data. We would also like to thank the Cambridge Group for the History of Population and Social Structure, the British Library (BL), the British Library of Political and Economic Science (BLPES) at the London School of Economics, the Guildhall Library, London Metropolitan Archives (LMA), and the Omnibus Society for their help with data. Finally, we are grateful to Charoo Anand, Iain Bamford, Horst Bräunlich, Dennis Egger, Andreas Ferrara, Ben Glaeser, Gordon Ji, Benny Kleinman and Florian Trouvain for excellent research assistance. The usual disclaimer applies.

†Corresponding author: Dept. Economics and WWS, Julis Romo Rabinowitz Building, Princeton, NJ 08544. Tel: 1 609 258 4016. Email: reddings@princeton.edu.

I. Introduction

Modern metropolitan areas include vast concentrations of economic activity, with Greater London and New York City today accounting for around 8.4 and 8.5 million people respectively. These intense population concentrations involve the transport of millions of people each day between their residence and workplace. Today, the London Underground alone handles around 3.5 million passenger journeys per day, and its trains travel around 76 million kilometers each year (about 200 times the distance between the earth and the moon). Yet relatively little is known about the role of these commuting flows in sustaining dense concentrations of economic activity. On the one hand, these commuting flows impose substantial real resource costs, both in terms of time spent commuting and the construction of large networks of complex transportation infrastructure. On the other hand, they are also central to the creation of predominantly commercial and residential areas, with their distinctive characteristics for production and consumption.

In this paper, we use the mid-19th-century transport revolution from the invention of steam railways, a newly-created, spatially-disaggregated dataset for Greater London from 1801-1921, and a quantitative urban model to provide new evidence on the contribution of the separation of workplace and residence to agglomeration. The key idea behind our approach is that the slow travel times achievable by human or horse power implied that most people lived close to where they worked when these were the main modes of transportation. In contrast, the invention of steam railways dramatically reduced the time taken to travel a given distance, thereby permitting the first large-scale separation of workplace and residence. Following this revolution in transport technology, we find that the night-time population in the historical center of the City of London falls by an order of magnitude from around 130,000 in 1831 (before the first railway line) to less than 14,000 in 1921, while its day-time population more than doubles over this period. We provide reduced-form event-study evidence connecting the decline in population in central areas and the rise in population in outlying areas to the arrival of the railway. We show that similar changes in patterns of population and employment growth are observed for other large metropolitan areas following the transport improvements of the 19th century, including Berlin, Paris, Boston, Chicago, New York and Philadelphia.

To interpret these empirical findings, we develop a theoretical framework that holds in an entire class of quantitative urban models that are characterized by a gravity equation in commuting and a land market clearing condition in which payments for commercial and residential floor space are constant multiples of income by workplace and income by residence respectively. In this class of models, the spatial distribution of economic activity is determined by variation in productivity, amenities and transport connections across locations. As commuting costs fall, workers become able to separate their residence and workplace to take advantage of the high wages in places with high productivity relative to amenities (so that these locations specialize as workplaces) and the low cost of living in places with high amenities relative to productivity (so that these locations specialize as residences). Our finding that the central city specializes as a workplace is consistent with it having high productivity relative to amenities compared to the suburbs and with the transition from walking/horses to railways disproportionately reducing travel times into the central city. If productivity and amenities depend on the density of workplace and residence employment respectively through agglomeration forces, this concentration of employment in the center and dispersion of population to the suburbs further magnifies these differences in productivity and amenities across locations.

Our quantitative framework has a recursive structure, such that we undertake our analysis in a sequence of steps, where each step imposes the minimal set of additional assumptions relative to the previous step. In particular, our baseline quantitative analysis uses only a combined land and commuter market clearing condition that holds in this entire class of quantitative urban models. Although we only observe bilateral commuting flows in 1921 at the end of our sample period, we show how our framework can be used to estimate the impact of the construction of the railway network going back to the early-19th century. We first use our bilateral commuting data for 1921 to estimate the parameters that determine commuting costs. We next combine these parameter estimates with historical data on population, land values and the evolution of the transport network from the beginning of the 19th century onwards. Using our framework's combined land and commuter market clearing condition, we solve for unobserved historical patterns of employment by workplace and bilateral commuting flows.

Although we estimate the effect of the railway network on commuting costs using 1921 information alone, we show that our model provides a good approximation to the available historical data on workplace employment and commuting patterns. In particular, we show that it captures the sharp divergence between night-time and day-time population in the City of London from the mid-19th century onwards, and replicates the property of the early commuting data that most people lived close to where they worked at the dawn of the railway age. Despite the City of London experiencing by far the largest *absolute* increase in employment, we find that the highest *percentage* rates of growth of employment (and population) occur in the suburbs, as these areas are transformed from villages and open fields to developed land. As a result, the gradient in log employment density in distance from the Guildhall in the center of the City of London declines between 1831 (before the first railway) and 1921, and the share of the 13 boroughs within 5 kilometers of the Guildhall in total workplace employment in Greater London falls from around 68 percent in 1831 to around 48 percent in 1921. This pattern of results is in line with the findings of employment decentralization in studies of more recent transport improvements, which are based on specifications using log employment or employment shares.

An advantage of our baseline quantitative analysis is that it holds for an entire class of urban models and conditions on observed population and rateable values. Therefore, it controls for a wide range of other potential determinants of economic activity, such as changes in productivity, amenities, the costs of trading goods, the supply elasticity for floor space, and expected utility in the wider economy. Given the data for the initial equilibrium in 1921, we show that the observed historical data on population and land values are sufficient statistics in the model for determining unobserved historical workplace employment and commuting, because they control for changes in these other determinants of economic activity within this class of urban models.

To further explore the respective contributions of changes in commuting costs and these other economic determinants, and to examine the implications of our findings for the strength of agglomeration forces, we choose a version of the canonical urban model of goods trade and commuting as one model from within our class. Using this model, we first recover productivity and amenities for each location and census year, and show that most of the change in net commuting into the City of London is explained by the change in commuting costs alone. Using these solutions for productivity and amenities, we estimate the strength of agglomeration forces by requiring that the observed reorganization of economic activity within Greater London following the invention of the steam railway is explained by the model's mechanism of a reduction in commuting costs and agglomeration forces. In particular, using the identifying assumption that idiosyncratic shocks to productivity and amenities are uncorrelated with distance from the Guildhall conditional on our controls, we estimate elasticities of productivity and amenities to workplace and residence employment density of 0.086 and 0.172 respectively. Our estimate of agglomeration forces in production of 0.086 lies close to the 3-8 range discussed in the survey by Rosenthal and Strange (2004), while our estimate of agglomeration forces in residential decisions of 0.172 is consistent with recent findings of endogenous amenities, such as Diamond (2016).

We use our model to undertake counterfactuals for the removal of the railway network under a range of alternative assumptions about the supply elasticity for floor space and the strength of agglomeration forces. Assuming an inelastic supply of floor space and no agglomeration forces, we find that removing the entire railway network reduces the total population and value of land and buildings in Greater London by 13.7 and 12.5 percent respectively, and decreases net commuting into the City of London by more than 270,000 workers. In comparison, removing only the underground railway network reduces the total population and value of land and buildings in Greater London by 3.5 and 4.0 percent respectively, and decreases net commuting into the City of London by around 80,000 workers. In both cases, the increase in the net present value of land and buildings exceeds historical estimates of the construction cost of the railway network. Introducing a positive floor space supply elasticity and/or agglomeration forces magnifies these effects. Using our calibrated floor space supply elasticity of 1.83 and our estimated agglomeration forces in production and residence, we find that removing the entire railway network reduces the total population and value of land and buildings in Greater London by 51.5 and 53.3 percent respectively, and decreases net commuting into the City of London by more than 350,000 workers .

London during the 19th century is arguably the poster child for the large metropolitan areas observed around the world today. In 1801, London's built-up area housed around 1 million people and spanned only 5 miles East to West. This was a walkable city of 60 squares and 8,000 streets that was not radically different from other large cities up to that time. In contrast, by 1901, Greater London contained over 6.5 million people, measured more than 17 miles across, and was on a dramatically larger scale than any previous urban area. By the beginning of the 20th-century, London was the largest city in the world by some margin (with New York City and Greater Paris having populations of 3.4 million and 4 million respectively at this time) and London's population exceeded that of several European countries.[1] Therefore, 19th-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. First, during this period, there is a revolution in transport technology in the form of the steam locomotive, which dramatically increased travel speeds from around 6 mph for horse omnibuses to 21 mph for railway trains. Steam locomotives were first developed to haul freight at mines (at the Stockton to Darlington Railway in 1825) and were only later used to transport passengers (with the London and Greenwich Railway in 1836 the first to be built specifically for passengers).[2] Second, in contrast to other cities, such as Paris, London developed through a largely haphazard and organic process. Until the creation of the Metropolitan Board of Works (MBW) in 1855, there was no municipal authority that spanned the many different local jurisdictions that made up Greater London, and the MBW's responsibilities were largely centered on infrastructure. Only in 1889 was the London County Council (LCC) created, and the first steps towards large-scale urban planning for Greater London were not taken until the Barlow Commission of 1940. Therefore, 19th-century London provides a setting in which we would expect the size and structure of the city to respond to decentralized market forces.

We contribute to several strands of existing research. Our paper connects with the theoretical and empirical literatures on agglomeration, including Henderson (1974), Fujita, Krugman, and Venables (1999), Fujita and Thisse (2002), Davis and Weinstein (2002), Lucas and Rossi-Hansberg (2002), Davis and Dingel (2019) and Kline and Moretti (2014),

[1] London overtook Beijing's population in the 1820s, and remained the world's largest city until the mid-1920s, when it was eclipsed by New York (Chandler 1987). Greece's population was 2.63 million in the 1907 census and Denmark's 2.59 million in the 1906 census (Webb 1911, p. 457).

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

as reviewed in Rosenthal and Strange (2004), Duranton and Puga (2004), Moretti (2011) and Combes and Gobillon (2015). A key challenge for empirical work in this literature is finding exogenous variation to identify agglomeration forces. Rosenthal and Strange (2008) and Combes, Duranton, Gobillon, and Roux (2010) use geology as an instrument for population density, exploiting the idea that tall buildings are easier to construct where solid bedrock is accessible. Greenstone, Hornbeck, and Moretti (2010) provide evidence on agglomeration spillovers by comparing changes in total factor productivity (TFP) among incumbent plants in "winning" counties that attracted a large manufacturing plant and "losing" counties that were the new plant's runner-up choice. In contrast, we exploit the transformation of the relationship between travel time and distance provided by the invention of the steam locomotive.

Our paper is also related to a recent body of research on quantitative spatial models, including Redding and Sturm (2008), Allen and Arkolakis (2014), Ahlfeldt, Redding, Sturm, and Wolf (2015), Redding (2016), Allen, Arkolakis, and Li (2017), Caliendo, Parro, Rossi-Hansberg, and Sarte (2018), Desmet, Nagy, and Rossi-Hansberg (2018), Monte (2018), Monte, Redding, and Rossi-Hansberg (2018) and Tsivanidis (2018), as reviewed in Redding and Rossi-Hansberg (2017). All of these papers focus on time periods for which modern transportation networks by rail and/or road exist, whereas we exploit the dramatic change in transport technology provided by the steam locomotive. Our main contributions relative to these previous studies are as follows. First, we show that an entire class of urban models feature the same combined land and commuter market clearing condition, which enables us to develop results that are robust across this class of models, and to control for a range of unobserved determinants of the spatial distribution of economic activity. Second, we develop a new structural estimation methodology for this class of urban models, which uses bilateral commuting flows for a baseline year (in our case 1921) and undertakes comparative statics from this baseline year (in our case backwards in time). We show that this estimation procedure can be used to recover unobserved historical employment and commuting data (prior to 1921) from the bilateral commuting data for our baseline year and historical data on population and land rents.[3] This procedure is applicable in other contexts, in which historical data are incomplete or missing, but bilateral commuting data are available for a baseline year. Third, we show that our class of urban models can account quantitatively for the reorganization of economic activity in Greater London following the invention of the steam railway.

Our paper also contributes to the empirical literature on the relationship between the spatial distribution of economic activity and transport infrastructure, as reviewed in Redding and Turner (2015). One strand of this literature has used variation across cities and regions, including Chandra and Thompson (2000), Michaels (2008), Duranton and Turner (2011), Duranton and Turner (2012), Faber (2014), Duranton, Morrow, and Turner (2014), Donaldson and Hornbeck (2016), Donaldson (2018) and Baum-Snow, Henderson, Turner, Zhang, and Brandt (2019). A second group of studies have looked within cities, including Warner (1978), Jackson (1987), McDonald and Osuji (1995), Gibbons and Machin (2005), Baum-Snow and Kahn (2005), Billings (2011), and Brooks and Lutz (2018). A third set of papers has examined the welfare effects of transport infrastructure, as in Allen and Arkolakis (2017) and Fajgelbaum and Schaal (2017). Within this literature, our work is most closely related to research on suburbanization and decentralization, including Baum-Snow (2007), Baum-Snow, Brandt, Henderson, Turner, and Zhang (2017), Gonzalez-Navarro and Turner (2018), and Baum-Snow (2019). Our contributions are again to use the large-scale variation from the transition from

[3]Our use of bilateral commuting data for our baseline year enables us to encompass this class of quantitative urban models and recover the unobserved historical employment and commuting data. In contrast, Ahlfeldt, Redding, Sturm, and Wolf (2015) did not have access to bilateral commuting data between blocks, and focused on one of the theoretical models from within our class in order to use observed data on employment by residence and employment by workplace to predict unobserved commuting between blocks.

human/horse power to steam locomotion and to show that our model can account both qualitatively and quantitatively for the observed changes in city structure.

The remainder of the paper is structured as follows. Section II. discusses the historical background. Section III. summarizes the data sources and definitions. Section IV. presents reduced-form evidence that the invention of the steam railway enabled a large-scale separation of workplace and residence. Section V. formalizes this idea in a class of quantitative urban models. Section VI. undertakes our baseline quantitative analysis. Section VII. chooses one urban model from within our class to both recover productivity and amenities and also estimate the strength of agglomeration forces. Section VIII. undertakes counterfactuals and compares the economic impact of the railway network to its construction cost. Section IX. concludes. A separate online appendix establishes isomorphisms for our class of urban models and contains the proofs of propositions, supplementary empirical results, and further details on the data.

II. Historical Background

London has a long history of settlement that dates back to before the Roman Conquest of England in 43 CE. We distinguish three geographical definitions of its boundaries, as illustrated in Figure Ia. First, we consider Greater London, as defined by the modern boundaries of the Greater London Authority (GLA), which includes a 1921 population of 7.39 million and an area of 1,595 kilometers squared (shown by the thick red line). Second, we examine the historical County of London, which has a 1921 population of 4.48 million and an area of 314 kilometers squared (indicated by the thick purple line). Third, we consider the City of London, which has a 1921 population of 13,709 and an area of around 3 kilometers squared, and whose boundaries correspond approximately to the Roman city wall (denoted by the barely visible thick green line).[4] Data are available at two main levels of spatial aggregation: boroughs (shown by the medium black lines) and parishes (shown by the medium gray lines). Greater London contains 99 boroughs and 283 parishes; the County of London comprises 29 boroughs and 183 parishes; and the City of London is one borough that includes 111 parishes.[5] As apparent from the figure, our data permit a high-level of spatial resolution: the median parish in Greater London has a 1901 population of 1,515 and an area of 0.97 kilometers squared; and the median borough in Greater London has a 1901 population of 26,288 and an area of 11.41 kilometers squared.

In the first half of the 19th-century, there was no municipal authority for the entire built-up area of Greater London, and public goods were largely provided by local parishes and vestries (centered around churches). As a result, in contrast to other cities, such as Paris, London's growth was largely haphazard and organic.[6] In response to the growing public health challenges created by an expanding population, the Metropolitan Board of Works (MBW) was founded in 1855. However, its main responsibilities were for infrastructure, and many powers remained in the hands of the parishes and vestries.[7] 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.[8] As the built-up

[4]From medieval times, the City of London acted as the main commercial and financial center of what became Great Britain, with the neighboring City of Westminster serving as the seat of Royal and Parliamentary government. For historical discussions of London, see for example Ball and Sunderland (2001) and White (2007, 2008).

[5]Parish boundaries in the population census change over time. We use the constant definitions of mappable units (henceforth referred to for simplicity as parishes) provided by Shaw-Taylor, Davies, Kitson, Newton, Satchell, and Wrigley (2010), as discussed further below.

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

[7]See Owen (1982). These public health challenges included cholera outbreaks, as examined in Ambrus, Field, and Gonzalez (2019). The MBW's main achievements were the construction of London's Victorian sewage system and the Thames embankment, as discussed in Halliday (1999).

[8]The LCC continued the MBW's infrastructure improvements, including some new road construction through housing clearance (e.g. Kingsway

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 common mode of transport was walking, with average travel speeds 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 and Bogart 2017). Given these limited transport options, most people lived close to where they worked, as discussed in the analysis of English 18th-century time use in Voth (2001). With the growth of urban populations, attempts to improve existing modes of transport led to the introduction of the horse omnibus from Paris to London in the 1820s. Its main innovation relative to the stage coach was increased passenger capacity for short-distance travel. However, the limitations of horse power and road conditions ensured that average travel speeds remained low at around 6 mph.[9] A further innovation along the same lines was the horse tram (introduced in London in 1860), but average travel speeds again remained low at around 6 mph, in part because of road congestion.[10]

Against this background, the steam passenger railway constituted a major transport innovation, although one with a long and uncertain gestation. The first successful commercial development of a stationary steam engine was by Thomas Newcomen in 1712 to pump mine water. However, the development of the separate condenser and rotary motion by James Watt from 1763-75 substantially improved its efficiency and expanded its range of potential applications. The first commercial use of mobile steam locomotives was to haul freight from mines at the Stockton and Darlington railway in 1825. However, in part as a result of fears about the safety of steam locomotives and the dangers of asphyxiation from rapid travel, it was not until 1833 that carriages with passengers were hauled by steam locomotives at this railway. Only in 1836 did the London and Greenwich railway open as the first steam railway to be built specifically for passengers. The result was a dramatic transformation of the relationship between travel time and distance, with average travel speeds using this new technology of around 21 mph.[11]

Railway development in London, and Great Britain more broadly, was undertaken by private companies in a competitive and uncoordinated fashion.[12] 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

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).

[9] See Section J8 of the online appendix for historical data on average speeds from London County Council (1907) and Barker and Robbins (1963).

[10] A later innovation was the replacement of the horse tram with the electric tram (with the first fully-operational services starting in 1901). But average travel speeds remained low at around 8 mph, again in part because of road congestion (see Section J8 of the online appendix). From 1900, motor buses gradually replaced horse buses, but were again only marginally faster because of road congestion and the need for frequent stops. Private car use was negligible with registered cars per person in the County of London in 1920 equal to 0.01 (London County Council 1921).

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

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

Road to the North and the Borough and Lambeth Roads to the South.[13] A legacy of this recommendation was the emergence of a series of railway terminals around the edge of this central area, which led to calls for an underground railway to connect these terminals. These calls culminated in the opening of the Metropolitan District Railway in 1863 and the subsequent development of the Circle and District underground lines. While these early underground railways were built using "cut and cover" methods, further penetration of Central London occurred with the development of the technology for boring deep-tube underground railways, as first used for the City and South London Railway, which opened in 1890, and is now part of the Northern Line.[14]

In Figures Ib, Ic and Id, we show the overground and underground railway networks in Greater London for 1841, 1881 and 1921 respectively, where a complete set of maps of the transport network for each census decade is found in Sections J5-J7 of the online appendix. In 1841, which is the first population census year in which any overground railways are present, there are only a few railway lines. These radiate outwards from the County of London, with a relatively low density of lines in the center of the County of London. 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 both overground and underground railway lines.

III. Data

We construct a new spatially-disaggregated dataset on economic activity in Greater London for the period 1801-1921. Our main source of data for Greater London is the population census of England and Wales, which we augment with other sources of data, as summarized below and discussed in further detail in Section J of the online appendix.

Bilateral Commuting: A first key component for our quantitative analysis of the model is the complete matrix of bilateral commuting flows between boroughs within Greater London, which is reported for the first time in the 1921 population census for England and Wales.[15] In this population census, we find that commuting flows between other parts of England and Wales and Greater London were small in 1921, such that Greater London was largely a closed commuting market.[16] Summing across rows in the matrix of bilateral commuting flows for Greater London, we obtain employment by workplace for each borough (which we refer to as "workplace employment"). Summing across columns, we obtain employment by residence for each borough (which we refer to as "residence employment"). We also construct an employment participation rate for each borough in 1921 by dividing residence employment by population.

Population by Residence: We combine our 1921 bilateral commuting data with historical data on population in Greater London from earlier population censuses from 1801-1911. Assuming that the ratio of residence employment to population is stable for a given borough over time, we use the 1921 value of this ratio and the historical population data

[13] 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, often in the form of railway terminals over bridges coming from the south side of the Thames at Victoria (1858), Charing Cross (1864), Cannon Street (1866), and Ludgate Hill (1864), and also at Waterloo (1848).

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

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

[16] 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 lived in the surrounding Home Counties, and the remainder lived in other parts of England and Wales. As residence is based on where one slept on Census night, while workplace is usual place of work, some of this 4 percent could be due to business trips or other travel.

to construct residence employment for earlier census years.[17] Parish boundaries are relatively stable throughout most of the 19th century, but experience substantial change in the early-twentieth century. For our reduced-form empirical analysis using the parish-level data, we construct constant parish boundary data every census decade from 1801-1901 using the classification provided by Shaw-Taylor, Davies, Kitson, Newton, Satchell, and Wrigley (2010), as discussed further in Section J1 of the online appendix. For our quantitative analysis of the model using the borough-level data, we use constant borough definitions every census decade from 1801-1921 based on the 1921 boundaries. For years prior to 1921, we allocate the parish-level data to the 1921 boroughs by weighting the values for each parish by its share of the geographical area of the 1921 borough. Given that parishes have a much smaller geographical area than boroughs, most parishes lie within a single 1921 borough.

Rateable Values: We measure the value of floor space in Greater London using rateable values, which correspond to the annual flow of rent for the use of land and buildings, and equals the price times the quantity of floor space in the model. In particular, these rateable values correspond to "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). With a few minor exceptions, they cover all categories of property, including public services (such as tramways, electricity works etc), government property (such as courts, parliaments etc), private property (including factories, warehouses, wharves, offices, shops, theaters, music halls, clubs, and all residential dwellings), and other property (including colleges and halls in universities, hospitals and other charity properties, public schools, and almshouses). All these categories of properties were assessed, regardless of whether or not their owners were liable for income tax. The main exemptions include roads, canals, railways, mines, quarries, Crown property occupied by the Crown, and places of divine worship. These rateable values have a long history in England and Wales, dating back to the 1601 Poor Relief Act, and were used to raise revenue for local public goods. We construct data on the value of land and buildings for each borough from 1831-1921, as discussed further in Section J2 of the online appendix.

Transport Network: We construct Geographical Information Systems (GIS) data on the transport network in Greater London over time using historical maps of the overground railway network, the underground railway network, and the omnibus and tram network, as discussed further in Sections J5-J8 of the online appendix. We measure bilateral travel times by distinguishing four transport networks based on the historical average travel speeds by transport mode in London reported in London County Council (1907): (a) overground railways (21 mph); (b) underground railways (15 mph), (c) omnibuses and trams (6 mph), and (d) walking (3 mph).[18] Following Donaldson (2018) and Donaldson and Hornbeck (2016), we construct bilateral travel times between parish and borough centroids assuming that workers follow the least cost path in terms of travel time. We assume that workers incur a travel time cost of 3 minutes in changing between modes of transport and can only connect to the railway network at railway stations.

We denote the sets of points connected to each transport network at time t by $\Im_t^{\mathrm{OR}}$, $\Im_t^{\mathrm{UR}}$, and $\Im_t^{\mathrm{OT}}$, where the superscripts OR, UR, and OT indicate overground railways, underground railways, and omnibuses and trams respectively. Similarly, we denote our vector of assumed travel time weights for each transport network by $\boldsymbol{\delta} = [1\ \delta^{UR}\ \delta^{OT}\ \delta^{WA}]$, where we normalize the weight for overground railways to one, and the superscript WA indicates walking. Using this

[17] Empirically, we find relatively little variation in employment participation rates across boroughs in 1921.

[18] In a robustness check, we find similar results if we instead use the travel speeds for England and Wales as a whole from Leunig (2006).

notation, we can write the bilateral travel times between boroughs n and i at time t as $d_{nit}^{W} = d_{nit}^{W}(\Im_t^{\text{OR}}, \Im_t^{\text{UR}}, \Im_t^{\text{OT}}, \boldsymbol{\delta})$, where the superscript W indicates the weighting by transport mode. In our econometric estimation below, we also use an instrumental variable based on bilateral travel times in which walking is assumed to be the only mode of transport, so that bilateral travel times depend solely on straight-line distance. We denote these bilateral travel times in the absence of other modes of transport by d_{ni}^{S}, where the superscript S is a mnemonic for straight-line distance.

Historical Employment by Workplace and Commuting Data: We compare our model's historical predictions for workplace employment within Greater London with data on the "day population" of the City of London that are available from the Day Censuses of 1866, 1881, 1891 and 1911.[19] In the face of a declining residential population ("night population"), the City of London Corporation undertook these censuses of the day population to demonstrate its enduring commercial importance. The day population is defined as "... every person, male or female, of all ages, 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."[20] In other specification checks, we compare our model's predictions for commuting distances with historical commuting data based on the residence addresses of the employees of the tailor Henry Poole, which has been located at the same workplace address in Savile Row in the City of Westminster since 1822.[21]

Data for Other Cities: To show that our findings for London are representative of those for other large metropolitan areas following the improvements in transport technology during the 19th-century, we have also collected historical data on population, employment and commuting distances for Berlin, Paris, Boston, Chicago, New York, and Philadelphia, as discussed in further detail in Sections I5 and J11 of the online appendix.

IV. Reduced-Form Evidence

The key economic mechanism underlying our approach is that a reduction in commuting costs facilitates an increased separation of workplace and residence. In particular, the hub and spoke structure of the railway network disproportionately reduced commuting costs into central locations within Greater London. If these central locations have high productivity relative to amenities compared to suburban locations, this improvement in transport technology leads them to specialize as a workplace, while the suburbs specialize as a residence. Before formalizing this idea in our class of urban models below, this section provides reduced-form evidence of such a change in patterns of specialization following the invention of the steam railway.

In subsection IV.A., we first show that population declines in the City of London and rises in the suburbs following the expansion of the railway network. We then establish that this decline in the City of London's population is combined with an increase in its employment. We next demonstrate that this change in specialization from a residence to a workplace makes the City of London a relatively more valuable location, as reflected in a higher share of the value of land and buildings, before this share again declines as a result of the expansion in the total area of developed land. In subsection IV.B., we report difference-in-differences event-study specifications that tighten the connection between the

[19] For further discussion of the City of London Day Censuses, see Section J4 of the online appendix.
[20] Salmon (1891), page 97.
[21] For further discussion of the Henry Poole data, see Section J9 of the online appendix and Green (1988).

timing of the expansion of the railway network and the changes in the spatial distribution of population. In subsection IV.C., we confirm that our findings for London are representative of those for other large metropolitan areas.

IV.A. City Size and Structure over Time

We begin by illustrating the dramatic changes in the internal structure of economic activity within Greater London between 1801 and 1921. In Figure II, we display residential population over time for the City of London (left panel) and Greater London (right panel). In each case, population is expressed as an index relative to its value in 1801 (such that 1801=1). In the first half of the 19th century, population in the City of London was relatively constant (at around 130,000), while population in Greater London grew substantially (from 1.14 million to 2.69 million). From 1851 onwards, there is a sharp drop in population in the City of London, which falls by around 90 percent to 13,709 in 1921. Over the same period, the population of Greater London as a whole continues to grow rapidly from 2.69 million in 1851 to 7.39 million in 1921.[22] Therefore, consistent with the expansion of the railway network after 1836 facilitating a dispersion of population to the suburbs, we find a precipitous drop in population in the most central parts of Greater London, combined with a rapid increase in population for the metropolitan area as a whole.

In the left panel of Figure III, we again show the City of London's residential (or "night") population, but now as thousands of people rather than an index. Alongside, we also display its day population, as reported in the City of London Day Censuses, except for the 1921 figure, which is workplace employment from our bilateral commuting data for 1921. Consistent with the City of London increasingly specializing as a workplace, we find that the steep decline in its night population from 1851 onwards coincides with a sharp rise in its day population. This evidence of an increasing specialization of locations as either workplaces or residences is also consistent with a sharp rise in public transport journeys per head of population from around 7 in 1834 to just under 400 in 1921, as shown in Figure I.1 in Section I1 of the online appendix.

To examine the implications of this change in specialization for the value of land and buildings, the right panel of Figure III displays the City of London's share of Greater London's rateable value over time. In the early-19th century, this share declines from around 14 to 9 percent, which is consistent with a geographical expansion in the built-up area of Greater London. As this expansion occurs, and undeveloped land becomes developed, the share of already-developed land in overall land values tends to fall, because the denominator of this share increases. In contrast to this pattern, in the years after 1851 when the City of London experiences the largest declines in its residential population, its rateable value share *increases* from 9 to 11 percent. Finally, in the decades at the end of the 19th century, the pattern of a decline in this rateable value share again reasserts itself, consistent with the continuing geographical expansion in the built-up area of Greater London. Therefore, the change in the City of London's pattern of specialization in the decades immediately after 1851 *increases* the relative value of this location.

Finally, in Figure IVa, we compare the evolution of the City of London's population over time to an index of the reduction in rail travel times from other London boroughs to the City. To compute this index, we calculate the bilateral reduction in travel time from the railway network between all other boroughs and the City, measured as travel time with the railway network in a census year divided by the travel time before the arrival of the railways. The index is a weighted sum of these bilateral travel time reductions, where the weights are the population of the origin boroughs

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

in each census year. Following the construction of the first railway in 1836, we observe a sharp reduction in the City of London's population-weighted average travel times to other boroughs, which is followed shortly thereafter by the precipitous decline in its population discussed above. Although the decline in population-weighted average travel times slightly leads the decline in the City's of London's population, both follow a similar trajectory, with their rates of decline slowing towards the end of the 19th century and beginning of the 20th century.

IV.B. Difference-in-Differences Event-Study Specification

In this subsection, we provide further reduced-form evidence on the timing of changes in population growth relative to the construction of the railway network using a difference-in-differences event-study specification and our spatially-disaggregated parish-level data for Greater London from 1801-1901. The main identification challenge in examining the relationship between population growth and the construction of the railway network is that railways are unlikely to be randomly assigned, because they were built by private-sector companies, whose stated objective was to maximize shareholder value. As a result, parishes in which economic activity would have grown for other reasons could be more likely to be assigned railways. We address this identification challenge by considering event-study specifications that include both a parish fixed effect and a parish time trend, and examining the relationship between the timing of deviations from these parish trends and the arrival of the railway.

A key prediction of our economic mechanism of specialization as workplaces or residences is that the treatment effect of the railway network on residential population should be heterogeneous, depending on whether parishes are located in the center or suburbs of Greater London. As a first step, and to provide a point of comparison, we begin by estimating a common average treatment effect for Greater London as a whole, before extending this specification to allow for heterogeneous treatment effects. In particular, we consider the following baseline specification:

$$\ln R_{jt} = \alpha_j + \sum_{\tau=-60}^{\tau=60} \beta_\tau \left(\mathbb{S}_j \times \mathbb{I}_{j\tau} \right) + \left(\mu_j \times \text{Year}_t \right) + d_t + u_{jt} \tag{1}$$

where j indexes parishes; t indicates the census year; R_{jt} is parish population; α_j is a parish fixed effect; $\mathbb{S}_j$ is an indicator variable that equals one if a parish has an overground or underground railway station in at least one census year during our sample period;[23] τ is a treatment year indicator, which equals the census year minus the last census year in which a parish had no railway station;[24] $\mathbb{I}_{j\tau}$ is an indicator variable that equals one in treatment year τ in parish j and zero otherwise; the excluded category is treatment year $\tau = 0$; Year_t is a census-year trend; μ_j is a parish-specific coefficient on this census-year trend; d_t is a census-year dummy; and u_{jt} is a stochastic error. One of the parish-specific census-year trends is colinear with the census-year dummies and hence is omitted without loss of generality. In our baseline specification, we cluster the standard errors on boroughs, which allows the error term to be serially correlated within parishes over time and to be spatially-correlated across parishes within boroughs.[25]

In this specification, the inclusion of the parish fixed effects (α_j) allows parishes treated with a railway to have higher population levels than other parishes in all years (both before and after the arrival of the railway). The census-year

[23] In our baseline specification, we combine both overground and underground railways, although we find a similar pattern of results in robustness checks using only overground railways.

[24] Therefore, $\tau = 0$ corresponds to the last year in which a parish had no railway and positive values of τ correspond to post-treatment years. For example, if the railway arrives in a parish in 1836, census year 1831 corresponds to $\tau = 0$, and census year 1841 corresponds to $\tau = 10$.

[25] In Table I.2 in Section I2 of the online appendix, we show that we typically find somewhat smaller standard errors if we instead cluster on parish, which only allows the error term to be serially correlated within parishes over time. We also experimented with Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors following Conley (1999), and again found these to be typically smaller than the standard errors clustered on borough in our baseline specification.

dummies (d_t) control for secular changes in population across all parishes, including the aggregate growth of Greater London. The parish-specific census-year trends ($\mu_j \times \text{Year}_t$) allow parishes treated with a railway to have higher trend population growth than other parishes in all years (both before and after the arrival of the railway). The key coefficients of interest (β_τ) are those on the interaction terms between the railway indicator ($\mathbb{S}_j$) and the treatment year indicator ($\mathbb{I}_{j\tau}$), which capture the treatment effect of a railway on population in parish j in treatment year τ. They correspond to deviations from the parish-specific census-year trends and have a "difference-in-differences" interpretation, where the first difference compares treated to untreated parishes, and the second difference undertakes this comparison before and after the arrival of the railway.

In our baseline specification, we include six interaction terms for decades from 10 to 60 years before and after a parish receives a railway station. The inclusion of these interactions both before and after the arrival of a railway station allows us to check non-parametrically the extent to which these deviations from the parish-specific census-year trends coincide with the arrival of the railway. We aggregate treatment years more than 60 years before and more than 60 years after the arrival of a railway station to ensure that these initial and final categories have a sufficient number of observations.[26] As the railway arrives in some parishes in a different calendar years from others, parishes differ in terms of the number of treatment years before and after the arrival of a railway station. Therefore, as we vary the number of years before and after the treatment, we change the composition of the treatment group of parishes. In Section I2 of the online appendix, we report a robustness test in which we hold the composition of this treatment group constant, by restricting attention to treatment windows 30 years before and after the arrival of a railway station, and dropping any treated parishes with less than 30 years before and after the arrival of a railway station.

In the top panel of Figure V, we display the estimated treatment effects (β_τ) from equation (1) and the 95 percent confidence intervals, with the full regression results reported in Table I.1 in Section I2 of the online appendix. We find positive and significant deviations in log population from the parish-specific time trends immediately after the arrival of the railway and no evidence of significant deviations from these trends before the arrival of the railway. This pattern of results provides strong evidence that the changes in population growth occur immediately after the arrival of the railway and suggests that the parish fixed effects and time trends largely control for the non-random placement of the railway. We next generalize this baseline specification to allow for heterogeneous treatment effects, distinguishing between parishes in the center of Greater London and those in its outlying suburbs:

$$\log R_{jt} = \alpha_j + \sum_{\tau=-60}^{\tau=60} \beta_\tau \left(\mathbb{S}_j \times \mathbb{I}_{j\tau}\right) + \sum_{\tau=-30}^{\tau=30} \gamma_\tau \left(\mathbb{S}_j \times \mathbb{I}_{j\tau} \times \mathbb{I}_j^{\text{Center}}\right) + d_t + u_{jt}, \tag{2}$$

where $\mathbb{I}_j^{\text{Center}}$ is an indicator variable that equals one for parishes in central London and zero otherwise; the railway treatment effect for parishes in central London is now given by $(\beta_\tau + \gamma_\tau)$; and the railway treatment effect for other parts of Greater London remains equal to β_τ.

In this specification allowing for heterogeneous treatment effects, we consider two definitions of central London, one based on parishes in the City of London, and another based on parishes with centroids within five kilometers of the Guildhall in the center of the City of London. A legacy of the parliamentary exclusion zone is that relatively few

[26] As the first railway was constructed in 1836, the maximum possible value for τ is 70 for parishes that receive a railway station before 1841, for which $\tau = 10$ in 1841 and $\tau = 70$ in 1901. Similarly, as our parish-level sample ends in 1901, the minimum possible value for τ is -90 for parishes that receive a railway after 1891, for which $\tau = 10$ in 1901 and $\tau = -90$ in 1801. Of our 3,113 (283 × 11) parish-year observations, 1,408 involve parishes that have a railway station in at least one census year during the sample period. The distribution of these 1,408 observations across the treatment years is: $\tau <= -60$ (133); $\tau = -50$ (83) ; $\tau = -40$ (106); $\tau = -30$ (128); $\tau = -20$ (128); $\tau = -10$ (128); $\tau = 0$ (128); $\tau = 10$ (128); $\tau = 20$ (119); $\tau = 30$ (104); $\tau = 40$ (96); $\tau = 50$ (60); and $\tau >= 60$ (67).

parishes within the City of London are treated with overground railways. Furthermore, the treatments for underground railways occur relatively late in the sample period. For both of these reasons, there is a relatively short interval after the arrival of railway stations in these parishes. Therefore, we only include interaction terms for 30 years before and after a parish receives a railway station for parishes in central London.[27]

In the middle and lower panels of Figure V, we display the estimated treatment effects (β_τ) and (γ_τ) respectively from equation (2) and the 95 percent confidence intervals, with the full regression results again reported in Table I.1 in Section I2 of the online appendix. The first specification (gray circle markers) shows results using the City of London as our definition of Central London. The second specification (black triangle markers) shows results using parishes with centroids within 5 kilometers of the Guildhall as our definition of central London. In both cases, we find positive and statistically significant treatment effects for the outer parts of Greater London in the middle panel, which are marginally larger than those shown in the top panel. For both specifications, we also find that the estimated treatment effects for central London are substantially and significantly smaller than those for the outer parts of Greater London, as shown in the lower panel. In Table I.3 in Section I2 of the online appendix, we show that we find the same pattern of results using our subsample with a constant composition of the treatment group over the period 30 years before and after the arrival of a railway station.

For both outer and central London, we find estimated treatment effects that increase in absolute magnitude over time up to 60 years after the arrival of the railway. One reason could be that the value of a connection to the railway network increases over time as the network expands. To examine this possibility, we use a measure of travel time reductions similar to that introduced in Section IV.A. above. First, we compute the bilateral reduction in travel time from the railway network between each pair of parishes in each year, measured as travel time with no railways divided by travel time with railways (the inverse of the measure of the fall in travel times used in Figure IVa). Second, we calculate for each parish the population-weighted average of these reductions in travel times to other parishes. In census years before the construction of any railway lines, this variable is equal to one for all parishes. In census years after the construction of the first railway, this variable is greater than or equal to one, with higher values corresponding to greater reductions in travel time. As our model predicts heterogeneous effects of reductions in travel time on parish population, depending on whether a parish is located in the center or the suburbs, we include both the log of this variable and its interaction with the log of distance from the Guildhall in the center of the City of London.

Consistent with our economic mechanism, we find a negative and significant coefficient on the population-weighted average reduction in travel time from the construction of the railway network, and a positive and significant coefficient on its interaction with distance from the Guildhall, as reported in Table I.1 in Section I2 of the online appendix. Therefore, after conditioning on our railway-treatment-year interactions, we find that reductions in travel times decrease population growth in central London and increase population growth in the outlying suburbs of Greater London. In the middle and lower panels of Figure V, the third specification (gray triangle markers) shows the corresponding estimated coefficients on these railway-treatment-year interactions. Comparing with the first and second specifications in these panels, we find that the estimated treatment effects fall by around one half and are often no longer statistically significant once we control for our travel time measure and its interaction with distance from the Guildhall, consistent with an important role for the reductions in travel times from the expansion of the railway network.

[27]Of the 1,221 (111 $\times$ 11) parish-year observations for the City of London, only 154 of these observations involve parishes that have a railway station in at least one census year during the sample period. The distribution of these 154 observations across the treatment years is $\tau <= -30$ (73); $\tau = -20$ (14) ; $\tau = -10$ (14); $\tau = 0$ (14); $\tau = 10$ (14); $\tau = 20$ (11) and $\tau >= 30$ (14).

The fact that some of the estimated coefficients on the railway-treatment-year interactions remain significant and continue to increase in magnitude over time has two sets of natural explanations within our class of urban models below. First, population-weighted average travel time reductions are not in general a sufficient statistic in these models for residence choice probabilities, which also depend on changes in wages, the price of floor space, productivity and amenities. Therefore the continued significance of the railway-treatment-year interactions could reflect the omission of controls for these other variables. Second, there could be a gradual response to the construction of the railway network, because of adjustment costs for investments in durable building capital. In Section F of the online appendix, we develop a dynamic model that features a gradual response because of such adjustment costs, and show that it falls within the class of urban models for which our baseline quantitative analysis holds.

In Section I3 of the online appendix, we report a robustness check using an alternative non-parametric specification, in which we estimate a separate railway treatment effect for each parish, using constant composition 30-year windows before and after the arrival of a railway station. Again this specification has a "difference-in-differences" interpretation, in which the railway treatment effect is identified from deviations from parish-specific time trends. Although we do not impose any relationship with geographical location, we find that these estimates exhibit a sharp non-linear relationship with distance from the Guildhall, with parishes close to the center of London experiencing decreased population growth relative to trend, and outlying parishes experiencing increased population growth relative to trend.

Therefore, these event-study results provide further support for the idea that the reduction in population growth in central London and its increase in the outlying parts of Greater London is closely connected to the railway. Indeed it is hard to think of confounding factors that are timed to coincide precisely with the arrival of the railway, and are structured to have exactly the same pattern of opposite effects on population in inner versus outer London.

IV.C. External Validity and Generalizability

A key advantage of our empirical setting of Greater London is the existence of rich historical data before and after the arrival of the railway and the availability of bilateral commuting data for 1921. In this section, we confirm that our findings for London are representative of those for other large metropolitan areas following the improvement in transport technology from the invention of the steam railway during the 19th century.

In Section I5 of the online appendix, we show that the same pattern of declining downtown population and rising metropolitan area population is observed for Berlin, Paris, Boston, Chicago, New York, and Philadelphia. We also provide evidence that the same mechanism of a change in specialization of the central city from a residence to a workplace operates for these other cities. Using workplace census data for Berlin, we show that the decline in population in the central city is accompanied by a rise in employment by workplace. Using journey-to-work data for Boston, New York and Philadelphia, we show that the decline in downtown population and increase in metropolitan area population involves an increase in the distances travelled to work in the central city.

Taken together, these findings paint a remarkably consistent picture across a wide range of different contexts. Our findings are also in line with the discussion in the existing economic history literature of a change in specialization of central cities from residential to commercial activity following 19th-century transport improvements, including Leyden (1933). Warner (1978), Hershberg (1981), Jackson (1987), Fogelson (2003), and Angel and Lamson-Hall (2014).

V. Theoretical Framework

We now develop our theoretical framework to rationalize these observed changes in the organization of economic activity within Greater London. We show that this theoretical framework encompasses an entire class of urban models that satisfy the following three properties: (i) a gravity equation for bilateral commuting flows; (ii) land market clearing, such that income from the ownership of floor space equals the sum of payments for residential and commercial floor space; (iii) payments for residential floor space are a constant multiple of residence income (the total income of all residents) and payments for commercial floor space are a constant multiple of workplace income (the total income of all workers). Within this class of models, workplace incomes are sufficient statistics for the demand for commercial floor space; residence incomes are sufficient statistics for the demand for residential floor space; and commuting costs regulate the difference between workplace and residence incomes.

In Section D of the online appendix, we develop a number of isomorphisms, in which we show that this class of urban models encompasses a wide range of different assumptions about consumption, production, goods' trade costs and the supply of floor space: (i) the canonical urban model with a single final good and costless trade (as in Lucas and Rossi-Hansberg 2002 and Ahlfeldt, Redding, Sturm, and Wolf 2015); (ii) an extension of the canonical urban model to incorporate non-traded goods; (iii) multiple final goods with costly trade and Ricardian technology differences (as in Eaton and Kortum 2002 and Redding 2016); (iv) final goods that are differentiated by origin with costly trade (as in Armington 1969, Allen and Arkolakis 2014 and Allen, Arkolakis, and Li 2017); (v) horizontally-differentiated firm varieties with costly trade (as in Helpman 1998, Redding and Sturm 2008, and Monte, Redding, and Rossi-Hansberg 2018). In Section F of the online appendix, we show that our approach also encompasses a dynamic model, in which there is a gradual response to the new transport technology, because of adjustment costs for investments in durable building capital. In this setting, the spatial distribution of economic activity responds sluggishly to the new transport technology, but our baseline quantitative analysis continues to hold. The reason is that this gradual response is captured in the observed historical population and rateable values and our baseline quantitative analysis conditions on these variables.

We consider a city (Greater London) that is embedded within a wider economy (Great Britain). The economy as a whole consists of a discrete set of locations $\mathbb{M}$. Greater London comprises a subset of these locations $\mathbb{N} \subset \mathbb{M}$. Time is discrete and is indexed by t. In our data for 1921, we only observe aggregate bilateral commuting flows for all workers together. Furthermore, the main effect of the steam railway was to permit the first large-scale separation of workplace and residence for workers in all occupations and industries. Therefore, we abstract from *ex ante* heterogeneity across different types of workers in the model by assuming a single worker type, although we do allow for *ex post* idiosyncratic heterogeneity across workers.[28] The economy as a whole is populated by an exogenous continuous measure $L_{\mathbb{M}t}$ of workers, who are geographically mobile and endowed with one unit of labor that is supplied inelastically. Workers simultaneously choose their preferred residence n and workplace i given their idiosyncratic draws.[29] With a continuous measure of workers, the law of large numbers applies, and the expected values of variables for a given residence and

[28] In Section I4 of the online appendix, we provide further evidence that commuting is pervasive across workers in all occupations and industries. First, we combine our 1921 bilateral commuting data with 1911 data on occupation by residence to show that there is little relationship between the shares of occupations in employment by residence and net inflows and outflows of commuters as a share of employment and residents. Second, we use data on occupation by residence back to the mid-19th century to show that there is an increase over time in the average distance of residences from the Guildhall in the center of the City of London for workers in all occupations.

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

workplace equal their realized values. Motivated by the fact that net commuting into Greater London is small even in 1921, we assume prohibitive commuting costs across the boundaries of Greater London. Therefore, each worker chooses a residence-workplace pair either in Greater London or in the rest of the economy. We denote the endogenous measure of workers who choose a residence-workplace pair in Greater London by $L_{\mathbb{N}t}$. We allow locations to differ from one another in terms of their attractiveness for production and residence, as determined by productivity, amenities, the supply of floor space, and transport connections, where each of these location characteristics can evolve over time.

V.A. Preferences

Worker preferences are defined over consumption goods and residential floor space. We assume that these preferences take the Cobb-Douglas form, such that the indirect utility for a worker ω residing in n and working in i is:[30]

$$U_{ni}(\omega) = \frac{B_{ni} b_{ni}(\omega) w_i}{\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}}, \qquad 0 < \alpha < 1, \tag{3}$$

where we suppress the time subscript from now onwards, except where important; P_n is the price index for consumption goods, which may include both tradeable and non-tradeable consumption goods; Q_n is the price of residential floor space; w_i is the wage; κ_{ni} is an iceberg commuting cost; B_{ni} captures amenities from the bilateral commute from residence n to workplace i that are common across all workers; and $b_{ni}(\omega)$ 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.[31] Each of the isomorphisms in Section D of the online appendix involves a different specification for the consumption price index (P_n). We show below that our baseline quantitative approach holds regardless of which of these specifications for the consumption price index (P_n) is chosen.

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

$$G(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{4}$$

where we normalize the Fréchet scale parameter in equation (4) to one, because it enters worker choice probabilities isomorphically to common bilateral amenities B_{ni} in equation (3); the Fréchet shape parameter ϵ regulates 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 the shape parameter ϵ, the greater the heterogeneity in idiosyncratic amenities, and the less sensitive are worker location decisions to economic variables.

We allow the common amenities (B_{ni}) to vary bilaterally to capture the fact that the attractiveness of a given commute may depend on characteristics of both the workplace and the residence. In particular, we decompose this bilateral common amenities parameter (B_{ni}) into a residence component that is common across all workplaces ($\mathcal{B}_n^R$), a workplace component that is common across all residences ($\mathcal{B}_i^L$), and an idiosyncratic component ($\mathcal{B}_{ni}^I$) that is specific to an individual residence-workplace pair:

$$B_{ni} = \mathcal{B}_n^R \mathcal{B}_i^L \mathcal{B}_{ni}^I, \qquad \mathcal{B}_n^R, \mathcal{B}_i^L, \mathcal{B}_{ni}^I > 0. \tag{5}$$

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

[31] Although we model the idiosyncratic heterogeneity in worker preferences ($b_{ni}(\omega)$), there is a closely-related formulation in terms of heterogeneity in worker productivity (effective units of labor), as discussed in Section E of the online appendix. Similarly, although we model commuting costs (κ_{ni}) in terms of utility, they enter the indirect utility function (3) multiplicatively with the wage, which implies that there is also a closely-related formulation in terms of the opportunity cost of time spent commuting.

We allow the levels of $\mathcal{B}_n^R$, $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ to differ across residences n and workplaces i, although when we examine the impact of the construction of the railway network, we assume that $\mathcal{B}_i^L$ and $\mathcal{B}_{ni}^I$ are time invariant. In contrast, we allow B_n^R to change over time, and for these changes to be potentially endogenous to the evolution of the surrounding concentration of economic activity through agglomeration forces, as discussed further below.

Conditional on choosing a residence-workplace pair in Greater London, we show in Section C of the online appendix that the probability a worker chooses to reside in location $n \in \mathbb{N}$ and work in location $i \in \mathbb{N}$ is given by:

$$\lambda_{ni} = \frac{L_{ni}/L_{\mathbb{M}}}{L_{\mathbb{N}}/L_{\mathbb{M}}} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{r\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \qquad n,i \in \mathbb{N}, \tag{6}$$

where L_{ni} is the measure of commuters from n to i.[32]

A first implication of our extreme value specification for idiosyncratic amenities is that bilateral commuting flows in equation (6) satisfy a gravity equation. Therefore, the probability of commuting between residence n and workplace i depends on the characteristics of that residence n, the attributes of that workplace i and bilateral commuting costs and amenities ("bilateral resistance"). Furthermore, this probability also depends on the characteristics of all residences k, all workplaces ℓ and all bilateral commuting costs ("multilateral resistance").

Summing across workplaces $i \in \mathbb{N}$, we obtain the probability that a worker lives in residence $n \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_n^R = R_n/L_{\mathbb{N}}$). Similarly, summing across residences $n \in \mathbb{N}$, we obtain the probability that a worker is employed in workplace $i \in \mathbb{N}$, conditional on choosing a residence-workplace pair in Greater London ($\lambda_i^L = L_i/L_{\mathbb{N}}$):

$$\lambda_n^R = \frac{\sum_{i\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \qquad \lambda_i^L = \frac{\sum_{n\in\mathbb{N}}\left(B_{ni}w_i\right)^{\epsilon}\left(\kappa_{ni}P_n^{\alpha}Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{7}$$

where R_n denotes employment by residence in location n and L_i denotes employment by workplace in location i.

A second implication of our extreme value specification is that expected utility conditional on choosing a residence-workplace pair ($\bar{U}$) is the same across all residence-workplace pairs in the economy:

$$\bar{U} = \vartheta\left[\sum_{k\in\mathbb{M}}\sum_{\ell\in\mathbb{M}}\left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{8}$$

where the expectation is taken over the distribution for idiosyncratic amenities; $\vartheta \equiv \Gamma((\epsilon-1)/\epsilon)$; $\Gamma(\cdot)$ is the Gamma function; and this result is derived in Section C2 of the online appendix. Using the probability that a worker chooses a residence-workplace pair in Greater London ($L_{\mathbb{N}}/L_{\mathbb{M}}$), we can rewrite this population mobility condition as:

$$\bar{U}\left(\frac{L_{\mathbb{N}}}{L_{\mathbb{M}}}\right)^{\frac{1}{\epsilon}} = \vartheta\left[\sum_{k\in\mathbb{N}}\sum_{\ell\in\mathbb{N}}\left(B_{k\ell}w_\ell\right)^{\epsilon}\left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{9}$$

where only the limits of the summations differ on the right-hand sides of equations (8) and (9).

Intuitively, for a given common level of expected utility in the economy ($\bar{U}$), locations in Greater London must offer higher real wages adjusted for common amenities (B_{ni}) and commuting costs (κ_{ni}) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity determined by the parameter ϵ.

[32]While we assume that workers make location decisions, we allow each worker to have non-working dependents, where we choose the ratio of workers to non-working dependents to match the ratio of residence employment to population in our census data.

V.B. Production

We assume that consumption goods are produced according to a Cobb-Douglas technology using labor, machinery capital and commercial floor space, where commercial floor space includes both building capital and land. We allow for Hicks-neutral productivity differences across locations, which can be potentially endogenous to the surrounding concentration of economic activity. Cost minimization and zero profits imply that payments for labor, commercial floor space and machinery are constant shares of revenue (X_i):

$$w_i L_i = \beta^L X_i, \qquad q_i H_i^L = \beta^H X_i, \qquad r M_i = \beta^M X_i, \qquad \beta^L + \beta^H + \beta^M = 1, \tag{10}$$

where q_i is the price of commercial floor space; H_i^L is commercial floor space use; M_i is machinery use; and machinery is assumed to be perfectly mobile across locations with a common price r determined in the wider economy. Each of the isomorphisms in Section D of the online appendix involves a different specification for consumption, production and trade costs that affects the determination of revenue (X_i). We show below that our baseline quantitative approach holds regardless of which of these different specifications is used to determine revenue.

Although an advantage of our empirical setting is the absence of any large-scale urban planning in 19th-century London, we allow the price of commercial floor space (q_i) to potentially depart from the price of residential floor space (Q_i) in each location i through a location-specific wedge (ξ_i):

$$q_i = \xi_i Q_i. \tag{11}$$

From the relationship between factor payments and revenue in equation (10), payments for commercial floor space are proportional to workplace income ($w_i L_i$):

$$q_i H_i^L = \frac{\beta^H}{\beta^L} w_i L_i. \tag{12}$$

V.C. Commuter Market Clearing

Commuter market clearing implies that the measure of workers employed in each location (L_i) equals the measure of workers choosing to commute to that location:

$$L_i = \sum_{n \in \mathbb{N}} \lambda_{ni|n}^R R_n, \tag{13}$$

where $\lambda_{ni|n}^R$ is the probability of commuting to workplace i conditional on living in residence n:

$$\lambda_{ni|n}^R = \frac{\lambda_{ni}}{\lambda_n^R} = \frac{\left(B_{ni} w_i / \kappa_{ni}\right)^{\epsilon}}{\sum_{\ell \in \mathbb{N}} \left(B_{n\ell} w_\ell / \kappa_{n\ell}\right)^{\epsilon}}, \tag{14}$$

where all characteristics of residence n (namely Q_n and P_n) have cancelled from the above equation, because they do not vary across workplaces for a given residence.

Commuter market clearing also implies that per capita income by residence (v_n) is a weighted average of the wages in all locations, where the weights are these conditional commuting probabilities by residence ($\lambda_{ni|n}^R$):

$$v_n = \sum_{i \in \mathbb{N}} \lambda_{ni|n}^R w_i. \tag{15}$$

V.D. Land Market Clearing

We assume that floor space is owned by landlords, who receive payments from the residential and commercial use of floor space, and consume only consumption goods. Land market clearing implies that total income from the ownership of floor space equals the sum of payments for residential and commercial floor space use:

$$\mathbb{Q}_n = Q_n H_n^R + q_n H_n^L = (1-\alpha)\left[\sum_{i\in\mathbb{N}} \lambda_{ni|n}^R w_i\right] R_n + \frac{\beta^H}{\beta^L} w_n L_n, \tag{16}$$

where H_n^R is residential floor space use; rateable values ($\mathbb{Q}_n$) equal the sum of prices times quantities for both residential floor space ($Q_n H_n^R$) and commercial floor space ($q_n H_n^L$); and we have used the expression for per capita income by residence (v_n) from the commuter market clearing in equation (15).

From this combined land and commuter market clearing condition (16), payments for residential floor space are a constant multiple of residence income ($v_n R_n$), and payments for commercial floor space are a constant multiple of workplace income ($w_n L_n$). Importantly, we allow the supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) to be endogenous, and we allow the prices of residential and commercial floor space to potentially differ from one another through the location-specific wedge ξ_i ($q_i = \xi_i Q_i$). In our baseline quantitative analysis below, we are not required to make assumptions about these supplies of residential and commercial floor space or this wedge between commercial and residential floor prices. The reason is that we condition on the observed rateable values in the data ($\mathbb{Q}_n$) and the supplies and prices for residential and commercial floor space (H_n^R, H_n^L, Q_n, q_n) only enter the land market clearing condition (16) through these observed rateable values.

VI. Quantitative Analysis

In this section, we undertake our baseline quantitative analysis. In Subsection VI.A., we introduce our general methodology, in which we use our combined land and commuter market clearing condition to evaluate the impact of changes in the transport network on the spatial distribution of economic activity. In the remaining Subsections VI.B.-VI.F., we implement this general methodology using our observed data for London. In particular, we use our bilateral commuting data for our baseline line year of $t = 1921$, the changes in residence employment and rateable values going back to the early-19th century in the data, and our estimates of changes in commuting costs to generate predictions for employment by workplace and commuting patterns back to the early-19th century.

VI.A. Combined Land and Commuter Market Clearing

We evaluate the impact of changes in the transport network by using an "exact hat algebra" approach similar to that used in the quantitative international trade literature following Dekle, Eaton, and Kortum (2007). In particular, we rewrite our combined land and commuter market clearing condition (16) for another year $\tau \neq t$ in terms of the values of variables in a baseline year of t and the relative changes of variables between years τ and t:

$$\hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} = (1-\alpha)\hat{v}_{nt} v_{nt} \hat{R}_{nt} R_{nt} + \frac{\beta^H}{\beta^L} \hat{w}_{nt} w_{nt} \hat{L}_{nt} L_{nt}, \tag{17}$$

where $\hat{x}_{nt} = x_{n\tau}/x_{nt}$ for the variable x_{nt} and we now make explicit the time subscripts. Similarly, using equations (13), (14) and (15), the relative change in employment ($\hat{L}_{it}$) and the relative change in average per capita income by residence ($\hat{v}_{nt}$) for year τ can be expressed as:

$$\hat{L}_{it}L_{it} = \sum_{n\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}. \tag{18}$$

$$\hat{v}_{nt}v_{nt} = \sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}, \tag{19}$$

where these equations include terms in changes in wages ($\hat{w}_n$) and commuting costs ($\hat{\kappa}_{ni}$) but not in amenities, because we assume that the workplace and bilateral components of amenities are constant ($\hat{\mathcal{B}}^L_{it} = 1$ and $\hat{\mathcal{B}}^I_{nit} = 1$), and changes in the residential component of amenities ($\hat{\mathcal{B}}^R_{nt} \neq 1$) cancel from the numerator and denominator of the fractions.

Using equations (18) and (19) to substitute for the terms in employment ($\hat{L}_{nt}L_{nt}$) and expected income by residence ($\hat{v}_{nt}v_{nt}$) in equation (17), we obtain our combined land and commuter market clearing condition for year τ:

$$\begin{aligned} \hat{\mathbb{Q}}_{nt}\mathbb{Q}_{nt} \quad &= (1-\alpha)\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{w}_{it}w_{it}\right]\hat{R}_{nt}R_{nt} \\ &+\frac{\beta^H}{\beta^L}\hat{w}_{nt}w_{nt}\left[\sum_{i\in\mathbb{N}} \frac{\lambda^R_{int|i}\hat{w}^{\epsilon}_{nt}\hat{\kappa}^{-\epsilon}_{int}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{i\ell t|i}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{i\ell t}}\hat{R}_{it}R_{it}\right]. \end{aligned} \tag{20}$$

Suppose that we observe the values of all variables in the initial equilibrium in our baseline year of t: the commuting probabilities conditional on residence ($\lambda^R_{nit|n}$), employment (L_{nt}), residents (R_{nt}), wages (w_{nt}), and average per capita income by residence (v_{nt}). Suppose also that we observe relative changes in residents ($\hat{R}_{nt}$) and rateable values ($\hat{\mathbb{Q}}_{nt}$) between years τ and t. Given these observed variables and known values for changes in commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), this combined land and commuter market clearing condition (20) provides a system of N equations that determines unique values for the N unknown relative changes in wages in each location ($\hat{w}_{nt}$).

Lemma 1 *Suppose that ($\hat{\mathbb{Q}}_{nt}$, $\hat{R}_{nt}$, L_{nit}, $\lambda^R_{nit|n}$, $\mathbb{Q}_{nt}$, v_{nt}, R_{nt}, w_{nt}, L_{nt}) are known. Given known values for model parameters $\{\alpha,\ \beta^L,\ \beta^H,\ \epsilon\}$ and the change in bilateral commuting costs ($\hat{\kappa}^{-\epsilon}_{nit}$), the combined land and commuter market clearing condition (20) determines a unique vector of relative changes in wages ($\hat{w}_{nt}$) in each location.*

Proof. See Section B of the online appendix. ■

Using these solutions for the relative changes in wages ($\hat{w}_{nt}$), we can immediately recover the unique relative change in employment ($\hat{L}_{nt}$) from the commuter market equilibrium condition (18). Similarly, we can solve for the unique relative change in average per capita income by residence ($\hat{v}_{nt}$) from equation (19). Finally, we can obtain the unique relative change in commuting flows ($\hat{L}_{nit}$) using the conditional commuting probabilities (14):

$$\hat{L}_{nit}L_{nit} = \frac{\lambda^R_{nit|n}\hat{w}^{\epsilon}_{it}\hat{\kappa}^{-\epsilon}_{nit}}{\sum_{\ell\in\mathbb{N}}\lambda^R_{n\ell t|n}\hat{w}^{\epsilon}_{\ell t}\hat{\kappa}^{-\epsilon}_{n\ell t}}\hat{R}_{nt}R_{nt}. \tag{21}$$

Combining these estimates of relative changes in wages, workplace employment and commuting flows $\{\hat{w}_{it},\ \hat{L}_{it},\ \hat{L}_{nit}\}$ with the values of these variables in our baseline year of t $\{w_{it},\ L_{it},\ L_{nit}\}$, we immediately recover wages, workplace employment and commuting flows in another year $\tau \neq t$ $\{w_{i\tau},\ L_{i\tau},\ L_{ni\tau}\}$.

Our combined land and commuter market clearing condition (20) makes clear why our approach holds in the entire class of urban models discussed above and examined in Section D of the online appendix. As long as payments for residential floor space are proportional to residence income, payments for commercial floor space are proportional to workplace income, and gravity in commuting holds, we are not required to make assumptions about other determinants of economic activity, such as goods' trade costs, productivity, residential amenities, consumption goods price indexes, market structure, agglomeration forces, expected utility in the wider economy, the floor space supply elasticity, or adjustment costs in the construction sector. In our combined land and commuter market clearing condition (20), the observed changes in rateable values and residents are sufficient statistics within this class of models that control for changes in these other determinants of economic activity.

A final advantage of conditioning on the observed commuting probabilities conditional on residence ($\lambda^R_{nit|n}$) in our baseline year of t in equation (20) is that these observed probabilities control for unobserved differences in the level of bilateral commuting costs across residence-workplace pairs. When we undertake our comparative static for removing the railway network in earlier years $\tau < t$, bilateral commuting flows necessarily remain zero for earlier years $\tau < t$ for all pairs of boroughs with zero flows in our baseline year ($\lambda^R_{nit|n} = 0$), which is consistent with the much more primitive commuting technology in earlier decades going back to the beginning of the 19th century.

VI.B. Commuting Probabilities (Step 1)

We now apply this general methodology to the construction of the railway network in 19th-century London. In a first step, we use the observed data on bilateral commuting flows (L_{nit}) from the population census in our baseline year $t = 1921$ to directly compute the following variables in that baseline year: (i) total city employment, $L_t = \sum_{n \in \mathbb{N}} \sum_{i \in \mathbb{N}} L_{nit}$; (ii) the commuting probability conditional on choosing a workplace-residence pair in Greater London, $\lambda_{nit} = L_{nit}/L_{\mathbb{N}t}$; (iii) workplace employment, $L_{it} = \sum_{n \in \mathbb{N}} L_{nit}$; (iv) residence employment, $R_{nt} = \sum_{i \in \mathbb{N}} L_{nit}$; (v) the commuting probabilities conditional on residence, $\lambda^R_{nit|n} = L_{nit}/R_{nt}$.

VI.C. Wages in the Initial Equilibrium (Step 2)

In a second step, we solve for wages (w_{nt}) and average per capita income by residence (v_{nt}) in the initial equilibrium in year $t = 1921$ using the observed workplace employment (L_{nt}), residence employment (R_{nt}) and rateable values ($\mathbb{Q}_{nt}$). In order to do so, we calibrate the model's utility and production function parameters based on historical data for our sample period. First, we assume a value for the share of housing in consumer expenditure of $1 - \alpha = 0.25$, which equals the average share of rent in income across occupations in the Registrar General's survey of 30,000 workers in Greater London in 1887, as reported to the House of Commons (Parliamentary Papers 1887).[33] Second, we assume a value for the share of labor in production costs of $\beta^L = 0.55$, which lies in the middle of the range of 0.43-0.63 considered for the period 1770-1860 in Antrás and Voth (2003).[34] Third, we assume a share of machinery and equipment in production costs of $\beta^K = 0.20$, and a share of land and building structures in production costs of $\beta^H = 0.25$, which are in line with the data on factor shares for 1856-1913 reported in Matthews, Feinstein, and Odling-Smee (1982), as discussed further in Section J10 of the data appendix.

Given values for these parameters, our combined land and commuter market clearing condition (16) for our baseline year of $t = 1921$ provides a system of N equations in the N unknown wages (w_{nt}). As this system of equations is linear in the unknown wages (w_{nt}), it determines a unique equilibrium value for the wage in each location in our baseline year (w_{nt}), as long as the rows of the observed matrix of bilateral commuting flows (L_{nit}) are linearly independent. Using

[33] See Section J10 of the data appendix for further discussion of these data and a comparison with other potential sources of historical data on the share of rent in income. Consistent with our assumption of Cobb-Douglas preferences, the average share of rent in income is relatively stable across the 35 occupations reported in Parliamentary Papers (1887).

[34] This value for the share of labor in production costs is close to the values of 0.56 and 0.50 used for similar historical time periods in Matthews, Feinstein, and Odling-Smee (1982) and Crafts and Harley (1992) respectively.

these solutions for wages (w_{nt}) and the observed commuting probabilities conditional on residence ($\lambda^R_{ni|n}$), we can recover average per capita income by residence (v_{nt}) in each location for our baseline year from equation (15). This use of our combined land and commuter market clearing condition to solve for wages is considerably less restrictive than the alternative approach of using the estimated workplace fixed effects from a gravity equation estimation, as in Ahlfeldt, Redding, Sturm, and Wolf (2015). In such an alternative approach, one is required to assume that there are no unobserved workplace components of commuting costs or amenities. In contrast, we control for such unobserved workplace components using the observed commuting probabilities conditional on residence ($\lambda^R_{nit|n}$).

VI.D. Estimating Commuting Costs (Step 3)

In our third step, we use the model's gravity equation predictions to estimate the relationship between commuting flows and travel times, taking into account the endogeneity of the transport network using instrumental variables estimation. In particular, we assume that bilateral commuting costs are a constant elasticity function of bilateral travel times and a stochastic error for all pairs of boroughs with positive commuting flows:

$$-\epsilon \log \kappa_{nit} = -\epsilon\phi \log d^W_{nit} + u^R_{nt} + u^L_{it} + u^I_{nit}, \tag{22}$$

where d^W_{nit} is our measure of travel time based on our assumed travel speeds ($\boldsymbol{\delta}$) and the observed transport network ($\Im^{\text{OR}}_t$, $\Im^{\text{UR}}_t$, $\Im^{\text{OT}}_t$); the composite elasticity $-\epsilon\phi$ captures both the elasticity of commuting flows to commuting costs ($-\epsilon$) and the elasticity of commuting costs to travel times (ϕ); u^R_{nt} is an unobserved residence component of commuting costs; u^L_{it} is an unobserved workplace component of commuting costs; and u^I_{nit} is an unobserved idiosyncratic component of commuting costs that is specific to individual residence-workplace pairs. In our baseline specification, we assume prohibitive commuting costs ($\kappa_{nit} \to \infty$) for all pairs of boroughs with zero commuting flows.

From equations (6), (9) and (22), we obtain the following gravity equation for the log commuting probability in our baseline year t:

$$\log \lambda_{nit} = \mathbb{W}_{it} + \mathbb{B}_{nt} - \epsilon\phi \log d^W_{nit} + e_{nit} \tag{23}$$

where the workplace fixed effect includes the wage and any workplace component of amenities or commuting costs, as well as the common expected utility in Greater London ($\mathbb{W}_{it} = \epsilon \log w_{it} + \epsilon \log \mathcal{B}^L_{it} - \log\left(L_{\mathbb{N}t}/L_{\mathbb{M}t}\right) - \epsilon \log\left(\bar{U}_t/\vartheta\right) + u^L_{it}$); the residence fixed effect includes the prices for goods consumption and floor space and any residence component of amenities or commuting costs ($\mathbb{B}_{nt} = -\epsilon\alpha \log P_{nt} - \epsilon(1-\alpha) \log Q_{nt} + \epsilon \log \mathcal{B}^R_{nt} + u^R_{nt}$); and the error term captures idiosyncratic shocks to bilateral amenities and commuting costs ($e_{nit} = \epsilon \log \mathcal{B}^I_{nit} + u^I_{nit}$).

A challenge in estimating equation (23) is that the transport network is endogenous, because overground and underground railways, omnibuses and trams were operated by private-sector companies. Therefore, the sets of points connected to the railway and hence the bilateral travel times (d^W_{nit}) are non-random. In particular, bilateral pairs that have more commuters for unobserved reasons in the error term (u_{nit}) could have more bilateral transport connections, and hence lower bilateral travel times (d^W_{nit}). To address this concern, we instrument our bilateral travel times based on the transport network (d^W_{nit}) with our bilateral travel times from a world in which walking is assumed to be the only mode of transport (d^S_{ni}), in which case bilateral travel times depend solely on straight-line distance. Conditional on the workplace and residence fixed effects, our identifying assumption is that the unobserved factors that affect commuting in the error term (u_{nit}) are orthogonal to the straight-line distance (d^S_{ni}) between locations. In our empirical setting, Greater London is relatively homogeneous in terms of other economic and geographic features that could be correlated with straight-line distance, and we provide empirical checks on this identifying assumption below.

In Column (1) of the top panel of Table I, we begin by estimating the gravity equation (23) using ordinary least squares (OLS). We find an elasticity of commuting flows with respect to travel times of around -4.90, which is statistically significant at the 1 percent level, and a regression R-squared of more than 0.8. In Column (2), we estimate the same specification using two stage least squares (2SLS), instrumenting our log bilateral travel times with log straight-line distance. We find a somewhat larger elasticity of commuting flows with respect to travel times of around -5.20, which is again statistically significant at the 1 percent level. This marginal increase in the coefficient between the OLS and IV specifications suggests that a greater incentive to invest in routes with more commuters for unobserved reasons in the error term may have been offset by other factors. In particular, the historical literature emphasizes the non-cooperative behavior of the private-sector railways, and their attempts to carve out geographical territories of dominance through a proliferation of branch lines. This struggle for areas of geographic dominance could have led to over-investment in routes that were less attractive in terms of their unobserved characteristics in the error term, thereby resulting in a marginally larger IV coefficient. As shown in the first-stage estimates reported in the bottom panel of Table I, we find that travel time increases less than proportionately with straight-line distance (with an elasticity of 0.43), because railways reduce travel time by more over longer straight-line distances. As also shown in this bottom panel, straight-line distance is a powerful instrument for travel times, with a first-stage F-statistic well above the conventional threshold of 10. Consistent with this, we comfortably reject the null hypothesis of underidentification in the Kleibergen-Paap underidentification test reported in the top panel.

In Section I6 of the online appendix, we report a number of robustness checks on this specification. First, we examine potential heterogeneity in the local average treatment effect (LATE) implied by the instrumental variables estimates. Second, we consider an overidentified specification, in which we include both the log of straight-line distance and the square of this log of straight-line distance as instruments. Third, we examine potential heterogeneity in the determinants of commuting patterns in different parts of Greater London by including fixed effects for bilateral flows between different geographical regions of Greater London (Central London, the rest of the County of London, and the rest of Greater London). Fourth, we control for potential heterogeneity in the determinants of commuting flows over long versus short distances by including fixed effects for quintiles of straight-line distance. Fifth, instead of assuming prohibitive commuting costs ($\kappa_{nit} \rightarrow \infty$) for all pairs of boroughs with zero commuting flows, we estimate the commuting gravity equation including the zeros and using the Poisson Pseudo Maximum Likelihood estimator of Santos Silva and Tenreyro (2006). Across each of these specifications, we obtain a similar qualitative and quantitative pattern of results.

Using our baseline instrumental variables estimates of the elasticity of commuting flows with respect to travel times ($-\epsilon\phi$) from Column (2), we construct measures of the change in commuting costs from the removal of the railway network between our baseline year of $t = 1921$ and an earlier year $\tau < t$:

$$\hat{\mathcal{K}}_{nit} = \hat{\kappa}_{nit}^{-\epsilon} = \left(\frac{\kappa_{ni\tau}}{\kappa_{nit}}\right)^{-\epsilon} = \left(\frac{d_{ni\tau}^{W}(\Im_{\tau}^{\mathrm{OR}}, \Im_{\tau}^{\mathrm{UR}}, \Im_{\tau}^{\mathrm{OT}}, \boldsymbol{\delta})}{d_{nit}^{W}(\Im_{t}^{\mathrm{OR}}, \Im_{t}^{\mathrm{UR}}, \Im_{t}^{\mathrm{OT}}, \boldsymbol{\delta})}\right)^{-\epsilon\phi}, \tag{24}$$

where the only source of changes in commuting costs between years τ and t is changes in the set of points connected to the transport network ($\Im_{\tau}^{\mathrm{OR}} \neq \Im_{t}^{\mathrm{OR}}$, $\Im_{\tau}^{\mathrm{UR}} \neq \Im_{t}^{\mathrm{UR}}$, $\Im_{\tau}^{\mathrm{OT}} \neq \Im_{t}^{\mathrm{OT}}$).

VI.E. Historical Workplace Employment (Step 4)

Given these estimates of changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and a value for the Fréchet shape parameter (ϵ), we can use our combined land and commuter market clearing condition (20) to solve for unobserved workplace employment back to the early-19th century. We calibrate the Fréchet shape parameter (ϵ) by minimizing the sum of squared deviations between the model's predictions for workplace employment and the data on the day population in the City of London for the census years for which these data are available (1881, 1891 and 1911). We obtain a calibrated Fréchet shape parameter of $\epsilon = 5.25$, which lies in between the estimate using commuting flows of around 6 in Ahlfeldt, Redding, Sturm, and Wolf (2015) and the estimate using migration flows of around 3 in Bryan and Morten (2019).

In Figure IVb, we compare the model's predictions for workplace employment in the City of London using $\epsilon = 5.25$ against the available data, which correspond to day population from the City of London day censuses for 1866, 1881, 1891 and 1911, and workplace employment from our bilateral commuting data for 1921. Our quantitative analysis conditions on the observed historical changes in employment by residence, which implies that employment by residence is identical in the model and data for all years. Similarly, our quantitative analysis conditions on workplace employment in the initial equilibrium in our baseline year, which implies that for 1921 our model's predictions and the data necessarily coincide with one another. Additionally, we calibrate the Fréchet shape parameter (ϵ) by minimizing the sum of squared deviations between the model's predictions and the data for the census years for which these data are available. Nonetheless, the model not only provides a good fit to the data on average, but also tracks the evolution of workplace employment across census years relatively closely.

In Figure VI, we display the log of residence employment density and workplace employment density against distance from the Guildhall in the center of the City of London, both for 1921 at the end of our sample period (when both variables are observed), and for 1831 before the first railway in Greater London (when residence employment is observed, but workplace employment is a model prediction). In each case, we display the fitted values from a locally-weighted linear least squares regression of each variable against distance from the Guildhall. In 1831, we find a steep gradient in both residence and workplace employment density with distance from the Guildhall. Between 1831 and 1921, residence employment falls and workplace employment rises in the most central parts of Greater London, as these locations specialize as a workplace, and become large net importers of commuters. In contrast, in the outlying parts of Greater London, residence employment rises more rapidly than workplace employment, such that these locations become net exporters of commuters. This change in specialization in central locations from a residence to a workplace is consistent with these central locations having high productivity relative to amenities compared to the suburbs and the rail transport network disproportionately reducing travel times into the central city.

Using data from more recent transport improvements (typically for highways but sometimes also for railways), an existing empirical literature has found evidence that these transport improvements decentralize employment, population and night lights, including Baum-Snow (2007), Baum-Snow, Brandt, Henderson, Turner, and Zhang (2017), Gonzalez-Navarro and Turner (2018), and Baum-Snow (2019). In principle, the effect of these more recent transport improvements on the location of economic activity could be quite different, depending on the extent to which they reduce travel times into central locations relative to those between peripheral locations. For example, as argued in Glaeser and Kohlhase (2004), rail transport is infrastructure-heavy, which favored a hub and spoke structure that reduced travel times into central locations. In contrast, road transport is relatively infrastructure-light, with dense networks of lateral connections, which is likely to have reduced travel times between outlying locations.

Despite these potential differences in the effect of transport improvements depending on their impact on bilateral travel times, our findings for steam railways during the 19th century are in fact consistent with the evidence from these more recent transport improvements. Although we find that the City of London experiences the largest *absolute* increase in employment during our sample period, we find the largest *percentage* rates of growth of both employment and population in the suburbs, as these areas are transformed from villages and open fields to developed land. As a result, the gradient of *log* workplace employment density in distance from the Guildhall declines in Figure VI between 1831 and 1921, and the *share* of the 13 boroughs within 5km of the Guildhall in the total workplace employment of Greater London also declines from around 68 percent in 1831 to around 48 percent in 1921. This pattern of results is in line with the findings of employment decentralization in these studies of more recent transport improvements, which are based on regression specifications using log employment or employment shares.

A key insight from our analysis is that what matters for the extent to which a location is specialized as a workplace versus a residence is the ratio of workers to residents (L_n/R_n). Following the invention of the steam railway during the 19th century, the increased specialization of the central city as a workplace rather than a residence (a rise in L_n/R_n) was achieved in a particularly dramatic form through an absolute *increase* in workplace employment (L_n) and an absolute *decrease* in residence employment (R_n). However, this finding of a rise in the ratio of workers to residents (L_n/R_n) in the central city relative to the suburbs is again consistent with existing studies of more recent transport improvements, which find greater decentralization of population than employment, and hence also find a rise in the ratio of workers to residents in central locations relative to peripheral locations.

VI.F. Specification Check (Step 5)

As a specification check, we now compare our model's predictions for commuting patterns during the 19th century to historical data that were not used in its calibration. In particular, we use data on commuting distances from the personnel ledgers of Henry Poole bespoke tailors, as discussed further in Section J9 of the online appendix. We focus on the model's predictions for commuters into the workplace of Westminster, in which this company is located. In the left panel of Figure VII, we compare the model's predictions for 1861 with the commuting distances of workers who joined Henry Poole between 1857 and 1877.[35] In the right panel of this figure, we compare the model's predictions for 1901 with commuting distances of workers who joined Henry Poole between 1891 and 1911. In these comparisons, there are several potential sources of discrepancies between the model and data, including the fact that this company is located in a specific site within Westminster, whereas the model covers all of that borough. Nevertheless, we find that our model is remarkably successful in capturing the change in the distribution of commuting distances between these two time periods. In the opening decades of the railway age, most workers in Westminster in both the model and the data lived within 5 kilometers of their workplace. In contrast, by the turn of the twentieth-century, we find substantial commuting over distances ranging up to 20 kilometers in both the model and data. This pattern of results is consistent with a wealth of historical evidence that most people lived close to where they worked before the railway age. Further evidence of an increase in the distance travelled to work following the transport improvements of the 19th century is provided for Boston, New York and Philadelphia in Section I5 of the online appendix.

[35] As discussed in more detail in section J9 of the Appendix, we use the residential address of each worker at the time they joined Henry Poole to compute their commuting distance.

VII. Productivity, Amenities and the Supply of Floor Space

In our baseline quantitative analysis so far, we have evaluated the impact of the railway network on employment by workplace and commuting patterns within an entire class of urban models, controlling for other potential determinants of economic activity, such as the supply of floor space, productivity and amenities. In this section, we consider one theoretical model from within our class in order to explicitly recover productivity, amenities and the supply of floor space, and to examine the implications of our findings for the strength of agglomeration forces.

Again we show that the model has a recursive structure, such that we can proceed in a number of steps, where each step involves the minimal set of additional assumptions. First, we use assumptions on the floor space supply elasticity to decompose rateable values into the contributions of the price and quantity of floor space. Second, we use the resulting solutions for changes in the price of floor space and additional assumptions about production, preferences and market structure to solve for changes in productivity and amenities. Finally, we parameterize agglomeration forces in production and residence to separate out these changes in productivity and amenities into the contributions of agglomeration forces and idiosyncratic shocks to locational fundamentals. All derivations for the results in this section are reported in Section G of the online appendix.

VII.A. Supply of Floor Space

We separate rateable values into the price and quantity of floor space by making the following two additional assumptions. First, we assume no-arbitrage between commercial and residential floor space use ($q_{nt} = Q_{nt}$), which is consistent with the positive values for residents and workers for all boroughs in our data and the absence of large-scale urban planning in 19th-century London. Second, we model the supply of floor space (H_{nt}) as depending on geographical land area (K_n) and a constant elasticity function of the price of floor space (Q_{nt}) following Saiz (2010):

$$H_{nt} = hQ_{nt}^{\mu} K_n, \tag{25}$$

where h is a constant; $\mu \geq 0$ is the floor space supply elasticity; and $\mu = 0$ corresponds to the special case of a perfectly inelastic supply of floor space.

Using these assumptions in our definition of rateable values ($\mathbb{Q}_{nt} = Q_{nt}H_{nt}^{R} + q_{nt}H_{nt}^{L} = Q_{nt}H_{nt}$), we recover both the price and quantity of floor space as constant elasticity functions of rateable values per land area ($\mathbb{Q}_{nt}/K_n$):

$$Q_{nt} = \left(\frac{\mathbb{Q}_{nt}}{hK_n}\right)^{\frac{1}{1+\mu}}, \qquad H_{nt} = hK_n\left(\frac{\mathbb{Q}_{nt}}{hK_n}\right)^{\frac{\mu}{1+\mu}}. \tag{26}$$

We calibrate the floor space supply elasticity (μ) using data on the contribution of new buildings towards changes in rateable values that were compiled by London County Council for part of our time period (1871-1921) for the subset of our boroughs in the County of London. Using the overall change in rateable value and the contribution of new buildings, we compute the change in the price and supply of floor space and hence the floor space supply elasticity for each of these boroughs ($\ln \hat{H}_{nt}/\ln \hat{Q}_{nt}$), as discussed further in Section I7 of the online appendix. We set the floor space supply elasticity in the model to the median value of these floor space supply elasticities across this subset of boroughs in the data ($\mu = 1.83$).

In the top-left panel of Figure VIII, we display log rateable value per land area against distance from the Guildhall for both 1831 and 1921. In each year, we show the fitted values from a locally-weighted linear least squares regression, and we normalize these fitted values by their value for the City of London (such that the variable takes the value zero for the City of London). As apparent from this figure, we observe the largest percentage growth in rateable value per land area between these years for the areas that become the inner suburbs of Greater London (from 8-15 kilometers from the Guildhall), as they are transformed from villages and open fields to developed land. Since both the price and supply of floor space in equation (26) are constant elasticity functions of rateable values per land area, their evolution over time displays the same patterns as shown in the top-left panel of Figure VIII.

As a check on our assumption of a common floor space supply elasticity ($\mu = 1.83$), Section I7 of the online appendix reports a specification check, in which we compare our model's predictions for the change in the supply of floor space from 1871-1921 for each borough in the County of London to the data on the construction of new buildings compiled by London County Council. As discussed further in the online appendix, we find a strong correlation between the model's predictions and the data, with a correlation coefficient of 0.77.

VII.B. Productivity and Amenities

To recover productivity and amenities, we choose one urban model from within our class, and impose additional assumptions on preferences, production and market structure. In particular, we choose an extension of the canonical urban model to incorporate non-traded goods, as developed in Section D2 of the online appendix. This framework permits a particularly tractable and transparent approach to recovering productivity and amenities and estimating the strength of agglomeration forces. It also allows us to undertake counterfactuals for removing the railway network under a range of alternative assumptions about the floor space supply elasticity and the strength of agglomeration forces. In the remainder of this subsection, we use this framework to recover productivity and amenities for each location. In the next subsection, we use these values for productivity and amenities to estimate the strength of agglomeration forces. In Section VIII. below, we undertake our counterfactuals for removing the railway network.

Starting with preferences in equation (3), we assume that the consumption goods price index (P_{nt}) is a Cobb-Douglas function of the price of a homogeneous traded good (P^T_{nt}) and a homogeneous non-traded good (P^N_{nt}):

$$P_{nt} = \left(P^T_{nt}\right)^{\nu} \left(P^N_{nt}\right)^{1-\nu}, \qquad 0 < \nu < 1. \tag{27}$$

Within Greater London, we assume that the homogeneous traded good is costlessly traded, such that $P^T_{nt} = P^T_t$ for all $n \in \mathbb{N}$. Between Greater London and the rest of Great Britain, we allow for changes in trade costs for this good, which are reflected in changes in its price at the boundaries of Greater London (P^T_t).

Turning now to the production technology and market structure, we assume that both the traded and non-traded goods are produced with labor and floor space according to the same Cobb-Douglas production technology under conditions of perfect competition. From profit maximization and zero profits, we recover a composite measure of productivity in the traded sector ($\mathbb{A}^T_{nt}$) using our solutions for wages (w_{nt}) and the price of floor space (Q_{nt}):

$$\mathbb{A}^T_{nt} = w_{nt}^{\beta^L} Q_{nt}^{\beta^H}, \qquad\qquad \mathbb{A}^T_{nt} \equiv A^T_{nt} P^T_t r_t^{-\beta^M}, \tag{28}$$

where this composite traded productivity ($\mathbb{A}^T_{nt}$) captures traded productivity (A^T_{nt}), the common price of the traded good (P^T_t), and the common price of machinery (r_t).

In the top right panel of Figure VIII, we display the log of composite productivity against distance from the Guildhall for 1831 and 1921. Again, we show the fitted values from locally-weighted linear least squares regressions, normalized by their value for the City of London. We find that productivity declines monotonically in distance from the Guildhall, with a similar gradient for both years up to around 15 kilometers from the Guildhall. As Greater London becomes increasingly developed between these two years, we observe an increase in the relative productivity of locations within 15 kilometers of the Guildhall compared to those beyond 15 kilometers from the Guildhall.

Using our additional assumptions on preferences, production and market structure in the residence choice probabilities in equation (7), we now recover a composite measure of changes in amenities. We focus on changes in amenities, because our baseline quantitative analysis uses estimates of changes in commuting costs from the construction of the railway network ($\hat{\kappa}_{nit}^{-\epsilon}$) and controls for unobserved determinants of the level of commuting costs in our baseline year using the initial commuting probabilities conditional on residence ($\lambda_{nit|n}^{R}$). Using our estimates of changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), we show in Section G2 of the online appendix that we can recover changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) from the residence choice probabilities as follows:

$$\hat{\lambda}_{nt}^{R}\lambda_{nt}^{R} = \frac{\lambda_{nt}^{R}\hat{\mathbb{B}}_{nt}^{\epsilon}\hat{Q}_{nt}^{-\epsilon(1-\alpha)}\widehat{RMA}_{nt}^{\epsilon}}{\sum_{k\in\mathbb{N}}\lambda_{kt}^{R}\hat{\mathbb{B}}_{kt}^{\epsilon}\hat{Q}_{kt}^{-\epsilon(1-\alpha)}\widehat{RMA}_{kt}^{\epsilon}}, \tag{29}$$

$$\widehat{RMA}_{nt} = \left[\sum_{\ell\in\mathbb{N}}\lambda_{n\ell t|n}^{R}\hat{w}_{\ell t}^{\epsilon}\hat{\kappa}_{n\ell}^{-\epsilon}\right]^{\frac{1}{\epsilon}} \qquad \hat{\mathbb{B}}_{nt} \equiv \hat{\mathcal{B}}_{nt}^{R}\left(\hat{A}_{nt}^{N}\right)^{\alpha(1-\nu)}\left(\hat{A}_{nt}^{T}\right)^{-\alpha(1-\nu)},$$

where recall that $\hat{\lambda}_{nt}^{R} = \lambda_{n\tau}^{R}/\lambda_{nt}^{R}$ for $\tau < t$; we observe the residence probabilities (λ_{nt}^{R}) and the commuting probabilities conditional on residence ($\lambda_{nit|n}^{R}$) in our baseline year of $t = 1921$; we also observe the relative changes in residence probabilities ($\hat{\lambda}_{nt}^{R}$); we solved for changes in wages ($\hat{w}_{nt}$) and floor prices ($\hat{Q}_{nt}$) above; $\widehat{RMA}_{nt}$ denotes the change in residents' commuting market access, as determined by changes in commuting costs ($\hat{\kappa}_{nit}$) and wages ($\hat{w}_{nt}$); changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) capture changes in the residential component of amenities ($\hat{\mathcal{B}}_{nt}^{R}$) and changes in productivity in each sector ($\hat{A}_{nt}^{N}$, $\hat{A}_{nt}^{T}$); and we assume that the workplace and idiosyncratic components of amenities remain constant over time ($\hat{\mathcal{B}}_{nt}^{L} = 1$ and $\hat{\mathcal{B}}_{nt}^{I} = 1$). Using equation (29), we solve for a unique vector of changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) up to a normalization or choice of units. We determine these changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) without making an assumption about the traded goods expenditure share (ν), which only affects the relative contributions of $\hat{A}_{nt}^{N}$ and $\hat{A}_{nt}^{T}$ towards the overall value of $\hat{\mathbb{B}}_{nt}$ pinned down by this equation.

In the bottom left panel of Figure VIII, we display the log changes in composite amenities ($\hat{\mathbb{B}}_{nt}$) against distance from the Guildhall. We show the fitted values from a locally-weighted linear least squares regression, normalized such that the value for the City of London is equal to zero. As apparent from the figure, we find the highest percentage rates of growth of composite amenities for the areas that become the inner suburbs of Greater London at intermediate distances from the Guildhall. From the residence choice probabilities (29), this pattern of results implies that the combination of higher growth rates of residence employment ($\hat{\lambda}_{nt}^{R}$) and the price of floor space ($\hat{Q}_{nt}$) in these locations is not fully explained by higher growth rates of commuting market access ($\widehat{RMA}_{nt}$), and hence we require higher growth rates of composite amenities ($\hat{\mathbb{B}}_{nt}$) to rationalize the observed data, through either higher amenities ($\hat{B}_{nt}$) or higher non-traded productivity ($\hat{A}_{nt}^{N}$) relative to traded productivity ($\hat{A}_{nt}^{T}$).

Our model solutions for changes in the supply of floor space ($\hat{H}_{nt}$), composite traded productivity ($\hat{\mathbb{A}}_{nt}^{T}$), and composite amenities ($\hat{\mathbb{B}}_{nt}$) are structural residuals, in the sense that they exactly rationalize the observed changes in residence employment ($\hat{R}_{nt}$) and rateable values ($\hat{\mathbb{Q}}_{nt}$) in the data and the changes in wages ($\hat{w}_{nt}$), workplace employment ($\hat{L}_{nt}$) and commuting patterns ($\hat{L}_{nit}$) from our baseline quantitative analysis. Therefore, if we start at our initial equilibrium in our baseline year of $t = 1921$, and feed into the system of equations for the general equilibrium of the model our estimated changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$) and the changes in these structural residuals ($\hat{H}_{nt}$, $\hat{\mathbb{A}}_{nt}^{T}$, $\hat{\mathbb{B}}_{nt}$), we exactly replicate the the observed data ($R_{n\tau}$, $\mathbb{Q}_{n\tau}$) and the results of our baseline quantitative analysis ($L_{n\tau}$, $w_{n\tau}$, $L_{ni\tau}$) for an earlier year $\tau < t$, as shown in Section G3 of the online appendix.

We can use this property of the model to undertake decompositions, in which we examine the relative contributions of changes in commuting costs ($\hat{\kappa}_{nit}^{-\epsilon}$), the supply of floor space ($\hat{H}_{nt}$), productivity ($\hat{A}_{nt}$) and amenities ($\hat{B}_{nt}$) towards the observed reorganization of economic activity within Greater London during the 19th-century. Although we use these model-based decompositions to assess the relative importance of different mechanisms, it should be kept in mind that these different mechanisms can influence one another. In particular, these decompositions treat productivity and amenities as exogenous. However, in the presence of agglomeration forces, productivity and amenities can respond endogenously to the reorganization of economic activity induced by the change in commuting costs, as examined in our counterfactuals in Section VIII.. Notwithstanding this limitation of the decompositions in this subsection, we find that much of the increase in net commuting into the City of London can be explained by the change in commuting costs alone ($\hat{\kappa}_{nit}^{-\epsilon}$), as shown in Figure G.1 in Section G3 of the online appendix. Taken together, these results provide further support for the idea that the invention of the steam railway was central to the emergence of the large-scale separation of workplace and residence in Greater London over our sample period.

VII.C. Agglomeration Forces

While the main focus of our analysis is quantifying the impact of the construction of the railway network on the internal organization of economic activity within Greater London, we now briefly examine the implications of our findings for the strength of agglomeration forces. Estimating agglomeration forces in our setting is challenging for several reasons. First, our data only includes 99 boroughs, compared to the thousands of city blocks considered in Ahlfeldt, Redding, Sturm, and Wolf (2015). Second, we do not directly observe land or floor space prices, and have to use assumptions about the supply of floor space to infer them from our rateable value data. Third, we estimate agglomeration forces over a much longer time period than usually considered, over which there is greater scope for other factors to change over time. Nevertheless, in Figures VI and VIII, areas with high rates of growth of composite traded productivity ($\hat{\mathbb{A}}_{n}^{T}$) and composite amenities ($\hat{\mathbb{B}}_{nt}$) also typically have high rates of growth of workplace employment ($\hat{L}_{nt}$) and residence employment ($\hat{R}_{nt}$), which is potentially consistent with the presence of agglomeration forces in both production and residence.

To provide further evidence on these agglomeration forces, we assume that productivity and amenities depend on two components: (i) production and residential externalities from the surrounding concentration of economic activity and (ii) production and residential fundamentals that are unrelated to the surrounding concentration of economic activity (e.g. natural water and green areas). We model production externalities as a constant elasticity function of a borough's own workplace employment density (with elasticity η^{L}) and residential externalities as a constant elasticity function of a borough's own residence employment density (with elasticity η^{R}). We specify these externalities as a function of a borough's own employment density, because of the relatively large area of the boroughs in our data, and the typical high rates of spatial decay of agglomeration economies. Taking differences between $\tau = 1831$ before the first railway line and our baseline year of $t = 1921$, we obtain the following expressions for the log changes in composite productivity and amenities, as shown in Section G4 of the online appendix:

$$\ln \hat{\mathbb{A}}_{nt}^{T} = \varsigma^{L} + \eta^{L} \ln \hat{L}_{nt} + \ln \hat{a}_{nt}, \tag{30}$$

$$\ln \hat{\mathbb{B}}_{nt} = \varsigma^{R} + \eta^{R} \ln \hat{R}_{nt} + \ln \hat{b}_{nt}, \tag{31}$$

where recall $\hat{L}_{nt} = L_{n\tau}/L_{nt}$ for $\tau < t$; the constants ς^{L} and ς^{R} control for any factors that are common across all locations within Greater London, such as common changes in productivity or amenities, or changes in expected utility in the wider economy; and $\ln \hat{a}_{nt}$ and $\ln \hat{b}_{nt}$ capture idiosyncratic shocks to production and residential fundamentals.

A key challenge in estimating the strength of agglomeration forces (η^{L}, η^{R}) in equations (30) and (31) is that workplace and residence employment are endogenous to productivity and amenities. In particular, the workplace choice probabilities (λ_{nt}^{L}) in equation (7) depend on the wage (w_{nt}), which in turn depends on traded productivity ($\mathbb{A}_{nt}^{T}$). Therefore, changes in workplace employment are likely to be positively correlated with idiosyncratic shocks to production fundamentals ($\ln \hat{a}_{nt}$) in the error term, thereby inducing an upwards bias in the estimated production elasticity (η^{L}). Similarly, the residence choice probabilities (λ_{nt}^{R}) in equation (7) are determined by amenities ($\mathbb{B}_{nt}$). Hence, changes in residence employment are likely to be positively correlated with idiosyncratic shocks to residential fundamentals ($\ln \hat{b}_{nt}$), thus imparting an upward bias in the estimated residential elasticity (η^{R}).

To address this challenge, we use the quasi-experimental variation from the invention of the steam railway to estimate the strength of agglomeration forces. In particular, we require that the observed reorganization of workplace and residence employment within Greater London is explained by the model's mechanisms of a change in commuting costs and agglomeration forces, rather than by systematic changes in production and residential fundamentals. Using this assumption, we estimate equations (30) and (31) using two-stage least squares, instrumenting the log changes in workplace and residence employment with indicator variables for 5 km distance grid cells from the Guildhall, where our excluded category is locations more than 20 km away. To allow changes in production and residential fundamentals to depend on the initial density of economic activity, we include as controls initial log employment (workplace and residence respectively) and log land area. Additionally, we include as a control an indicator variable for whether a borough is located in the London County Council (LCC) area to allow for potential differences in the supply of local public goods within and outside the LCC boundaries. Therefore, we estimate the production and residential externalities parameters (η^{L} and η^{R} respectively) using the identifying assumption that conditional on our controls the idiosyncratic shocks to production and residential fundamentals ($\ln \hat{a}_{nt}$, $\ln \hat{b}_{nt}$) are unrelated to distance from the Guildhall, and hence have the same mean value across these distance grid cells.

Table II reports the estimation results. As a point of comparison, in Columns (1) and (2), we estimate equations (30) and (31) using OLS. We find positive and statistically significant relationships between the growth of productivity and amenities and the growth of workplace and residence employment, with elasticities of 0.148 and 0.248 respectively. In Columns (3) and (4), we estimate the same specifications using two stage least squares, instrumenting the growth in workplace and residence employment with our indicator variables for grid cells in distance from the Guildhall. Consistent with the upward bias discussed above from the dependence of workplace and residence employment density on productivity and amenities, our instrumental variables estimates are smaller than our OLS estimates, with an elasticity of productivity (amenities) with respect to workplace (residence) employment density of 0.086 (0.172). We find that our instruments have power in the first-stage regressions, with first-stage F-statistics above the conventional threshold of 10. In Hansen-Sargen overidentification tests, we are unable to reject the model's overidentifying restrictions at conventional levels of significance.

As discussed above, these results come with a number of caveats. Our identifying assumption that the idiosyncratic shocks to productivity and residential fundamentals are uncorrelated with distance from the Guildhall conditional on our controls could be violated if there are changes to other determinants of productivity and amenities that are correlated with distance from the Guildhall and not captured by our controls for initial log employment (workplace and residence respectively), log land area and whether a borough is located in the County of London. Notwithstanding these caveats, we find estimates of the strength of agglomeration forces that are broadly in line with those in the existing empirical literature. Our estimated production agglomeration parameter of 0.086 lies close to the 3-8 percent range discussed in the survey by Rosenthal and Strange (2004) and well within the range of estimates reported in the meta-analysis by Melo, Graham, and Noland (2009). Our estimated residential agglomeration parameter of 0.172 is also close to the value of 0.155 reported in Ahlfeldt, Redding, Sturm, and Wolf (2015), and is consistent with the finding of endogenous amenities in Diamond (2016). Therefore, viewed through the lens of this extension of the canonical urban model and our identifying assumptions, the observed reorganization of economic activity within Greater London following the invention of the steam railway implies substantial agglomeration forces in both production and residence.

VIII. Counterfactuals

A key advantage of our baseline quantitative analysis in Section VI. is that it allows us to use our estimates of changes in commuting costs to recover the unobserved historical values of workplace employment and commuting patterns, while using the observed historical residence employment and rateable values to control for other determinants of economic activity. In Section VII., we used additional structure from the canonical urban model with non-traded goods to recover changes in the supply of floor space, productivity and amenities and estimate the strength of agglomeration forces. In this section, we use this model to examine the counterfactual question of how the spatial distribution of economic activity within Greater London would have evolved if the only thing that changed were the railway network, holding constant other determinants of economic activity.

We undertake our counterfactuals under a range of alternative assumptions about the floor space supply elasticity (μ) and the strength of agglomeration forces (η^L, η^R). In our first specification, we remove the railway network, holding constant the supply of floor space, productivity and amenities at their values in our baseline year of $t = 1921$ ($\mu = \eta^L = \eta^R = 0$). In our second specification, we set the floor space supply elasticity equal to our calibrated value and hold productivity and amenities constant ($\mu = 1.83$ and $\eta^L = \eta^R = 0$), thereby only allowing for an endogenous response in the supply of floor space. In our third specification, we introduce our estimated agglomeration forces in production ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0$), thus capturing endogenous responses in the supply of floor space and productivity. In our fourth specification, we incorporate our estimated agglomeration forces in production and residence ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0.172$), thereby allowing for endogenous responses in the supply of floor space, productivity and amenities. Finally, to address the concern that the floor space supply elasticity in 19th-century London could be larger than in other settings with more building regulations, we report results using a floor space supply elasticity equal to half our calibrated value ($\mu = 1.83/2 = 0.92$) in Section H of the web appendix.

For each of these specifications, we undertake two sets of counterfactuals. First, we examine the overall impact of the new transport technology on economic activity within Greater London by removing the entire railway network decade-by-decade back to 1831 (before the first railway line). Second, we examine the role of the underground in enabling passengers to travel into the most central parts of London by removing only the underground network decade-by-decade back to 1861 (before the first underground line). This second counterfactual further isolates a pure change in commuting costs, because the underground network is exclusively used to transport people. In both cases, we hold expected utility ($\bar{U}_t$) and total employment ($L_{\mathbb{M}t}$) in the wider economy constant, and allow the share of the economy's workers that choose residence-workplace pairs in Greater London ($L_{\mathbb{N}t}/L_{\mathbb{M}t}$) to adjust until expected utility in Greater London in the population mobility condition (9) is equal to its unchanged value in the wider economy.

In each case, we hold omnibus and tram routes constant at the 1921 network structure. In principle, one could allow for changes in omnibus and tram routes in response to the removal of the railway network, although the direction of this response is unclear. On the one hand, some of the increase in commuting costs from the removal of the railway network could be offset by an expansion of the omnibus and tram network. On the other hand, these two modes of transport are imperfect substitutes, with quite different average travel speeds. Indeed, over our sample period, omnibuses and trams were largely complementary to railways, expanding in tandem with them, and being more important for shorter journeys (including from railway terminals to final destinations). Therefore, to focus on the direct effect of the railway network, we hold the omnibus and tram network constant.

Given a change in commuting costs ($\hat{\kappa}_{rit}^{-\epsilon}$) implied by the removal of parts of the railway network and the observed variables ($\mathbb{Q}_{nt}$, λ_{nt}^{L}, λ_{nt}^{R}, $\bar{L}_t$, $\lambda_{nit|n}^{R}$, λ_{nit}, w_{it}) in our baseline year of $t = 1921$, we solve for counterfactual changes in the following five endogenous variables: (i) the workplace choice probability ($\hat{\lambda}_{nt}^{L}$); (ii) the residence choice probability ($\hat{\lambda}_{nt}^{R}$); (iii) the price of floor space ($\hat{Q}_{nt}$); (iv) the wage ($\hat{w}_{nt}$); (v) total city population ($\hat{L}_{\mathbb{N}t}$). These counterfactual changes solve the system of five equations for the general equilibrium of the model: (i) the land market clearing condition; (ii) the zero-profit condition; (iii) the workplace choice probability; (iv) the residence choice probability; and (v) the population mobility condition, as shown in Section H of the online appendix.

In the special case of the model in which productivity, amenities and the supply of floor space are exogenous $\eta^L = \eta^R = \mu = 0$, there are no agglomeration forces and the supply of land is perfectly inelastic, which ensures the existence of a unique equilibrium, as shown in Proposition H.1 in Section H of the online appendix. Therefore, our counterfactuals yield unique predictions for the impact of the change in commuting costs on the spatial distribution of economic activity. In the presence of agglomeration forces ($\eta^L > 0$ and $\eta^R > 0$) and an elastic supply of land ($\mu > 0$), whether or not the equilibrium is unique depends on the strength of these agglomeration forces relative to the model's congestion forces and the exogenous differences in production and residential fundamentals across locations. In each of the counterfactuals considered in this section, we obtain the same counterfactual equilibrium regardless of our starting values for the counterfactual changes in the model's endogenous variables.

In Figure IX, we display the model's counterfactual predictions for removing the entire railway network (left panels) and the underground railway network (right panels). As the first underground line was built in 1863, the counterfactuals in the right-hand panels are flat before 1861. We report results for net commuting into the City of London in the top two panels, the total population of Greater London in the middle two panels, and total rateable values in Greater London in the bottom two panels. In each case, the solid black line with no markers (labelled "Baseline") corresponds to our baseline quantitative analysis from Section VI. above. Our first counterfactual holding constant the supply of floor space, productivity and amenities ($\mu = \eta^L = \eta^R = 0$) is shown by the solid gray lines (labelled "Inelastic No Agglom"). In this specification, we find that removing the entire railway network reduces net commuting into the City of London to 98,173 in 1831 (panel (a)), while eliminating only the underground network decreases these net commuting flows to 293,165 in 1861 (panel (b)). As apparent from panel (a), the counterfactual net commuting of 98,173 in 1831 holding floor space, productivity and amenities constant compares to 30,375 in our baseline quantitative analysis from Section VI. above. Therefore, consistent with the results from our model-based decompositions in the previous section, we find that much of the increased separation of workplace and residence in Greater London during the 19th century can be explained by the pure change in commuting costs from the new transport technology rather than by other determinants of economic activity.

Our second counterfactual with our calibrated floor space supply elasticity and constant productivity and amenities ($\mu = 1.83$ and $\eta^L = \eta^R = 0$) is shown by the black dashed line with circle markers (labelled "Elastic No Agglom"). We find that introducing an endogenous response in the supply of floor space magnifies the effect of the new transport technology. In this specification, removing the entire railway network reduces net commuting into the City of London to 65,962 in 1831 (panel (a)), while eliminating only the underground network decreases these commuting flows to 258,115 in 1861 (panel (b)). This pattern of results highlights a complementarity between the development of the built environment and improvements in transport infrastructure.

Our third counterfactual with our calibrated floor space supply elasticity and our estimated production agglomeration force ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0$) is shown by the gray dashed line with circle markers (labelled "Elastic Prod Agglom"). We find that the introduction of production agglomeration forces further enhances the impact of the new transport technology. In this specification, removing the entire railway network reduces net commuting into the City of London to 25,347 in 1831 (panel (a)), while eliminating only the underground network decreases these commuting flows to 205,537 in 1861 (panel (b)). Our fourth counterfactual with our calibrated floor space supply elasticity and our estimated production and residential agglomeration forces ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0.172$) is shown by the black dashed line with triangle markers (labelled "Elastic Prod + Amen Agglom"). In this specification, removing the entire railway network reduces net commuting into the City of London to a value below that in our baseline quantitative analysis (12,099 compared to 30,375 in 1831 in panel (a)). Therefore, the model with an elastic supply of floor space and our estimated agglomeration forces generates a decline in net commuting into the City of London in the second half of the 19th century of a comparable magnitude to that observed in the data.

In panels (c) and (d), we examine the counterfactual predictions of the model for the total population of Greater London, where we compute counterfactual population using the model's predictions for residence employment and our constant employment participation rate for each borough. As shown by the solid black line, the observed population for Greater London rises from 1.92 to 7.39 million between 1831 and 1921. Our model predicts that the removal of the railway network leads to a reduction in the total population of Greater London, because boroughs specialize less according to their comparative advantages as workplaces or residences, which reduces real income in Greater London for a given population. This in turn leads to a population outflow until expected utility in Greater London is equal to its unchanged value in wider economy ($\bar{U}_t$).[36]

[36]To the extent that the overground railway network was also removed in other parts of Great Britain, and this reduced expected utility in the wider economy ($\bar{U}_t$), the effects on Greater London's population would be reduced. Nevertheless, the degree of separation of workplace and residence in Greater London is much greater than in other less-densely-populated locations, implying that Greater London would be more adversely affected by the removal of overground railways than these other locations. Furthermore, the underground railway network is largely specific to Greater London, because Glasgow is the only other city in Great Britain with an underground line.

In our first counterfactual holding constant the supply of floor space, productivity and amenities ("Inelastic No Agglom"), we find that removing the entire railway network reduces the total population of Greater London by 13.72 percent from 7.39 million in 1921 to 6.38 million in 1831 (panel (c)), while eliminating only the underground network decreases total population by 3.51 percent to 7.13 million in 1861 (panel (d)). Again we find that introducing a positive floor space supply elasticity and agglomeration forces magnifies the predicted impact of the new transport technology. With our calibrated floor space supply elasticity and our estimated production agglomeration force ("Elastic Prod Agglom"), removing the entire railway network reduces the total population of Greater London by 26.09 percent to 5.46 million in 1831 (panel (c)).[37] By comparison, with our calibrated floor space supply elasticity and our estimated production and residence agglomeration forces ("Elastic Prod + Amen Agglom"), eliminating the entire railway network decreases the total population of Greater London by 51.45 percent to 3.59 million in 1831 (panel (c)). Therefore, we find sizeable effects of the new transport technology of the steam railway on the total population of Greater London, with both production and residential agglomeration forces making quantitatively relevant contributions towards the impact of this new transport technology.

We find a similar pattern of results for the total value of land and buildings in Greater London (panels (e) and (f)) as for population (panels (c) and (d)), with the following two differences. First, an important advantage of our quantitative analysis is that our estimation results are invariant to goods price inflation, because both rateable values and wages are homogenous of degree one in goods prices. Therefore, multiplying all goods prices by a constant leaves unchanged the allocation of workers and residents and the relative values across locations of the supply of floor space, productivity and amenities. However, in our counterfactuals, we remove the railway network holding goods prices constant. Therefore, it is unsurprising that the counterfactual changes in rateable values are smaller than the observed changes, because these observed changes include goods price inflation, whereas the counterfactual changes do not. Second, under our Cobb-Douglas assumptions on preferences and technology, rateable values and wage bills move by the same proportion, but the percentage changes in rateable values and population differ from one another, because of movements in wages.

Despite these two differences, we obtain similar qualitative and quantitative conclusions for the total value of land and buildings in Greater London as for its total population. We again find that the introduction of a positive floor space supply elasticity and agglomeration forces magnifies the counterfactual effects of the removal of the railway network. With our calibrated floor space supply elasticity and no agglomeration forces ("Elastic No Agglom"), we find that removing the entire railway network reduces the total value of land and buildings in Greater London by 23.61 percent, from 65.86 million in 1921 to 50.31 million in 1831 (panel (e)). Introducing our estimated production agglomeration forces ("Elastic Prod Agglom"), we find a larger decline in the total value of land and buildings of 31.56 percent to 45.08 million in 1831 (panel(e)). Incorporating both our estimated production and residence agglomeration forces ("Elastic Prod + Amen Agglom") further magnifies these effects, with a decline in the total value of land and buildings of 53.25 percent to 30.79 million in 1831 (panel(e)).

We now use these counterfactual predictions of the model to evaluate the welfare effects of the construction of the railway network. Under our assumptions of population mobility and a constant value of expected utility in the wider economy, the total population of Greater London adjusts such that the expected utility of workers in Greater London

[37] Comparing the "Elastic No Agglom" and "Elastic Prod Agglom" specifications, we find a different organization of economic activity within Greater London (as reflected in the results for net commuting into the City of London in panels (a) and (b)), but the differences in the impact on the total population of Greater London are more visible for the removal of the entire railway network in panel (c) than they are for the elimination of only the underground network in panel (d) (where we find only a small difference in total population of 6,828,255 compared to 6,837,185 in 1861).

is unaffected by the construction of the railway network. Therefore, as in the classical approach to valuing public goods using land values following George (1879), the welfare gains from the new transport technology are experienced by landlords through changes in the value of land and buildings. We assess the magnitude of these welfare gains by comparing the counterfactual changes in the net present value of rateable values from the removal of the railway network to its construction costs. We measure these construction costs using historical estimates of authorized capital per mile for the private-sector companies that built these lines, which yields estimates of $\pounds 555,000$ per mile for bored-tube underground railways, $\pounds 355,000$ per mile for cut-and-cover underground railways, and $\pounds 60,000$ per mile for overground railways (all in 1921 prices), as discussed further in Section J6 of the online appendix.

In Table III, we report the results of these comparisons of the economic impact of the removal of the railway network with its construction costs. In the top panel, we remove the entire railway network. In the middle panel, we eliminate only the underground railway network. In the bottom panel, we examine the extent to which there were diminishing returns to the construction of the railway network, by only removing overground and underground railway lines constructed in the final decade of our sample from 1911-1921. In Column (1), we assume an inelastic supply of floor space and constant productivity and amenities ($\mu = \eta^L = \eta^R = 0$). In Column (2), we introduce our calibrated floor space supply elasticity, while continuing to assume constant productivity and amenities ($\mu = 1.83$ and $\eta^L = \eta^R = 0$). In Column (3), we augment this specification with our estimated production agglomeration force ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0$). In Column (4), we incorporate our estimated production and residential agglomeration forces ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0.172$). In each specification, we compute the net present value of changes in rateable values assuming an infinite lifetime and either 3 or 5 percent discount rates, and report the ratio of these reductions in the net present value of rateable values to the savings in construction costs.[38]

As shown in Table III, regardless of whether we assume a 3 or 5 percent discount rate, and irrespective of whether we consider the entire railway network (top panel) or only the underground railway network (middle panel), we find ratios of reductions in the net present value of rateable values to construction costs that are substantially greater than one. This pattern of results suggests that the large-scale investments in the construction of the 19th-century railway network in Greater London can be rationalized in terms of their effects on the net present value of economic activity. Comparing Columns (1) and (2), we find that allowing for a positive floor space supply elasticity substantially increases the economic impact of the railway network, again highlighting the role of complementary expansions in the supply of floor space. Comparing Columns (2), (3) and (4), we also find that incorporating agglomeration economies further magnifies the economic effects of the railway network, illustrating the relevance of endogenous changes in both productivity and amenities for the evaluation of transport improvements. Consistent with diminishing returns to the expansion of the railway network, we find lower ratios of reductions in the net present value of rateable values to savings in construction costs for overground and underground railway lines added in the final decade of our sample (bottom panel) than for the entire railway network (top panel). Nevertheless, even for these late railway lines, we find that the change in the net present value of rateable values exceeds construction costs.

[38] These discount rates of 3-5 percent are in line with the reported report rates of dividends paid on ordinary shares for London railways in the official *London Statistics* publication of the London County Council and are standard values in cost-benefit analyses.

IX. Conclusions

We use the natural experiment of the invention of the steam railway to provide evidence on the role of modern transport technologies in underpinning the concentration of economic activity in large metropolitan areas. The key idea underlying our approach is that a reduction in commuting costs facilitates an increased separation of workplace and residence. In particular, the hub and spoke structure of the railway disproportionately reduced commuting costs into central locations. If these central locations have high productivity relative to amenities compared to the suburbs, this transport improvement leads them to specialize as a workplace, while the suburbs specialize as a residence. In the presence of agglomeration forces, this increased concentration of workers in the center and residents in the suburbs further reinforces these differences in productivity and amenities across locations.

We begin by providing reduced-form evidence in support of this mechanism. First, we show that population declines in the City of London and rises in the suburbs following the construction of the railway network. Second, we establish that this decline in the City of London's population is combined with an increase in its employment. Third, we find that this change in specialization from a residence to a workplace makes the City of London a relatively more valuable location. We demonstrate the same pattern of results for other large metropolitan areas following the transport improvements of the 19th century. Using our parish-level data for London, we find no evidence of deviations in population growth from parish time trends before the arrival of the railway, and substantial deviations from these trends after its arrival. Consistent with our mechanism, we find reductions in parish population growth relative to trend in central London, and increases in parish population growth relative to trend in the suburbs.

To interpret these patterns in the data, we develop a new structural estimation procedure for an entire class of urban models, which satisfy a gravity equation in commuting flows and are characterized by a constant proportional relationship between labor income and payments for floor space. Although we only observe bilateral commuting flows in 1921 at the end of our sample period, we show that this framework can be used to estimate the impact of the construction of the railway network going back to the early-19th century. As this approach holds for an entire class of urban models and conditions on observed population and rateable values, it controls for a wide range of other potential determinants of economic activity, such as changes in productivity, amenities, the costs of trading goods, the supply of floor space, and the expected utility in the wider economy. We show that our model successfully captures the sharp divergence between night and day population in the City of London from the mid-19th century onwards and also replicates the property of early commuting data that most people lived close to where they worked at the dawn of the railway age.

To further explore the role of changes in commuting costs relative to other determinants of economic activity and to examine the implications of our findings for the strength of agglomeration forces, we consider a version of the canonical urban model of goods trade and commuting from within our class of models. Using the identifying assumption that the change in the organization of economic activity within Greater London following the invention of the steam railway is explained by the resulting change in commuting costs and agglomeration forces rather than by a systematic change in locational fundamentals, we estimate substantial agglomeration forces for both production and residence. Undertaking counterfactuals for the removal of the railway network, we find that the change in commuting costs alone accounts for most of the observed separation of workplace and residence. Holding the supply of floor space, productivity and amenities constant, we find that removing the entire railway network reduces the total population and rateable value of Greater London by 13.7 and 12.5 percent respectively, and decreases net commuting into the City of London by more than 270,000 workers. Introducing an endogenous supply of floor space and agglomeration forces magnifies these effects, highlighting the relevance of complementary changes in the built environment and agglomeration forces for cost-benefit analyzes of transport improvements. Using our calibrated floor space supply elasticity and estimated agglomeration forces, these declines in total population and rateable value reach 51.5 and 53.3 percent respectively. Across a wide range of different specifications, we find increases in the net present value of land and buildings that exceed historical estimates of the railway network's construction costs.

Taken together, we find that a class of quantitative urban models is remarkably successful in explaining the large-scale changes in the organization of economic activity observed in 19th-century London, and our findings highlight the role of modern transport technologies in sustaining dense concentrations of economic activity.

UNIVERSITY OF TORONTO
PRINCETON UNIVERSITY, NBER AND CEPR
LONDON SCHOOL OF ECONOMICS AND CEPR

References

ABERNATHY, S. T. (2015): "Opening up the Suburbs: Workmen's Trains in London 1860-1914," *Urban History*, 42(1), 70–88.

AHLFELDT, G., S. REDDING, D. STURM, AND N. WOLF (2015): "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83(6), 2127–2189.

ALLEN, T., AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129(3), 1085–1140.

——— (2017): "The Welfare Effects of Transportation Infrastructure Improvements," Dartmouth College.

ALLEN, T., C. ARKOLAKIS, AND X. LI (2017): "Optimal City Structure," Yale University, mimeograph.

AMBRUS, A., E. FIELD, AND R. GONZALEZ (2019): "Loss in the Time of Cholera: Long-run Impact of a Disease Epidemic on the Urban Landscape," *American Economic Review*, forthcoming.

ANGEL, S., AND P. LAMSON-HALL (2014): "The Rise and Fall of Manhattan's Densities 1800-2010," Marron Institute of Urban Management Working Paper No. 18.

ANTRÁS, P., AND H.-J. VOTH (2003): "Factor Prices and Productivity Growth During the British Industrial Revolution," *Explorations in Economic History*, 40, 52–77.

ARMINGTON, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers*, 16, 159–178.

BALL, M., AND D. SUNDERLAND (2001): *An Economic History of London 1800-1914*. Routlege, Abingdon.

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. (2019): “Urban Transport Expansions and Changes in the Spatial Structure of US Cities: Implications for Productivity and Welfare,” *Review of Economics and Statistics*, forthcoming.

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*, 99(2), 435–448.

BAUM-SNOW, N., V. HENDERSON, M. A. TURNER, Q. ZHANG, AND L. BRANDT (2019): “Does Investment in National Highways Help or Hurt Hinterland City Growth?,” *Journal of Urban Economics*, forthcoming.

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.

BOGART, D. (2017): “The Turnpike Roads of England and Wales,” *The Online Historical Atlas of Transport, Urbanization and Economic Development in England and Wales c.1680–1911*.

BROOKS, L., AND B. LUTZ (2018): “Vestiges of Transit: Urban Persistence at a Micro Scale,” *Review of Economics and Statistics*, 101(3), 385–399.

BRYAN, G., AND M. MORTEN (2019): “The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia,” *Journal of Political Economy*, forthcoming.

CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2018): “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” *Review of Economic Studies*, 85(4), 2042–2096.

CHANDLER, T. (1987): *Four Thousand Years of Urban Growth: An Historical Census*. The Edwin Mellen Press, Lewiston, NY.

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, chap. 1, pp. 15–66. Chicago University Press, Chicago.

COMBES, P.-P., AND L. GOBILLON (2015): “The Empirics of Agglomeration Economies,” in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. C. Strange, vol. 5A, chap. 5, pp. 247–348. Elsevier North Holland, Amsterdam.

CONLEY, T. G. (1999): “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, 92(1), 1–45.

CRAFTS, N. F. R., AND C. K. HARLEY (1992): “Output Growth and the British Industrial Revolution: a Restatement of the Crafts-Harley View,” *Economic History Review*, 45, 703–730.

CROOME, D. F., AND A. A. JACKSON (1993): *Rails Through the Clay: a History of London's Tube Railways*. Capital Transport Publishing, London.

DAVIS, D. R., AND J. DINGEL (2019): "A Spatial Knowledge Economy," *American Economic Review*, 109(1), 153–70.

DAVIS, D. R., AND D. E. WEINSTEIN (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

DAVIS, M. A., AND F. ORTALO-MAGNÉ (2011): "Housing Expenditures, Wages, Rents," *Review of Economic Dynamics*, 14(2), 248–261.

DEKLE, R., J. EATON, AND S. KORTUM (2007): "Unbalanced Trade," *American Economic Review*, 97(2), 351–355.

DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2018): "The Geography of Development," *Journal of Political Economy*, 126(3), 903–983.

DIAMOND, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 106(3), 479–524.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108(4-5), 899–934.

DONALDSON, D., AND R. HORNBECK (2016): "Railroads and American Economic Growth: A Market Access Approach," *Quarterly Journal of Economics*, 131(2), 799–858.

DURANTON, G., P. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 81(2), 681–724.

DURANTON, G., AND D. PUGA (2004): "Micro-Foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J.-F. Thisse, vol. 4, pp. 2063–2117. Elsevier, Amsterdam.

DURANTON, G., AND M. A. TURNER (2011): "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, 101(6), 2616–52.

——— (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79(4), 1407–1440.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

FABER, B. (2014): "Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country," *Review of Economic Studies*, 81(3), 1046–1070.

FAJGELBAUM, P. D., AND E. SCHAAL (2017): "Optimal Transport Networks in Spatial Equilibrium," *NBER Working Paper*, 23200.

FOGELSON, R. M. (2003): *Downtown: Its Rise and Fall, 1880-1950*. Yale University Press, New Haven CT.

FOLEY, D. L. (1963): *Controlling London's Growth: Planning the Great Wen*. University of California Press, Berkeley.

FUJITA, M., P. KRUGMAN, AND A. J. VENABLES (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge MA.

FUJITA, M., AND J.-F. THISSE (2002): *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*. Cambridge University Press, Cambridge.

GEORGE, H. (1879): *Progress and Poverty*. Doubleday, Page & Company, New York.

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.

GLAESER, E. L., AND J. E. KOHLHASE (2004): "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science*, 83, 197–228.

GONZALEZ-NAVARRO, M., AND M. A. TURNER (2018): "Subways and Urban Growth: Evidence from Earth," *Journal of Urban Economics*, 108, 85–106.

GREEN, D. R. (1988): "Distance to Work in Victorian London: A Case Study of Henry Poole, Bespoke Tailors," *Business History*, 30(2), 179–194.

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 Publishing, Stroud, Gloucestershire UK.

HELPMAN, E. (1998): "The Size of Regions," in *Topics in Public Economics: Theoretical and Applied Analysis*, ed. by D. Pines, E. Sadka, and I. Zilcha, pp. 33–54. Cambridge University Press, Cambridge.

HENDERSON, J. V. (1974): "The Sizes and Types of Cities," *American Economic Review*, 64(4), 640–56.

HERSHBERG, T. (1981): *Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century*. Oxford University Press, Oxford.

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

KELLET, J. R. (1969): *The Impact of Railways on Victorian Cities*. Routlege and Kegan Paul, London.

KLINE, P., AND E. MORETTI (2014): "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 129(1), 275–331.

LEUNIG, T. (2006): "Time is Money: a Re-assessment of the Passenger Social Savings from Victorian British Railways," *Journal of Economic History*, 66(3), 635–673.

LEYDEN, F. (1933): *Gross-Berlin: Geographie einer Weltstadt*. Gebr. Mann Verlag, Berlin.

LONDON COUNTY COUNCIL (1907): *London Statistics*. London County Council, London.

——— (1921): *London Statistics*. London County Council, London.

LUCAS, R. E., AND E. ROSSI-HANSBERG (2002): "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445–1476.

MATTHEWS, R. C., C. H. FEINSTEIN, AND J. C. ODLING-SMEE (1982): *British Economic Growth 1856-1973*. Clarendon Press.

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.

MELO, P. C., D. J. GRAHAM, AND R. B. NOLAND (2009): "A Meta-analysis of Estimates of Urban Agglomeration Economies," *Regional Science and Urban Economics*, 39(3), 332–342.

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.

MONTE, F. (2018): "The Local Incidence of Trade Shocks," Georgetown University, mimeograph.

MONTE, F., S. REDDING, AND E. ROSSI-HANSBERG (2018): "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 108(12), 3855–3890.

MORETTI, E. (2011): "Local Labor Markets," in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 4b, pp. 1238–1303. Elsevier North Holland, Amsterdam.

OWEN, D. (1982): *The Government of Victorian London: The Metropolitan Board of Works, the Vestries and the City Corporation*. Belknap Press, Cambridge MA.

PARLIAMENTARY PAPERS (1887): "Tabulations of the Statements Made by Men Living in Certain Selected Districts of London in March 1887," London: Her Majesty's Stationery Office.

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*, 9, 21–58.

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, Amsterdam.

——— (2008): "The Attenuation of Human Capital Spillovers," *Journal of Urban Economics*, 64(2), 373–389.

SAIZ, A. (2010): "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125(3), 1253–1296.

SALMON, J. (1891): *Ten Years' Growth of the City of London*. Simkin, Marshall, Hamilton Kent & Company, London.

SANTOS SILVA, J. M. C., AND S. TENREYRO (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641–658.

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," http://www.geog.cam.ac.uk/research/projects/occupations/.

TSIVANIDIS, N. (2018): "The Aggregate And Distributional Effects Of Urban Transit Infrastructure: Evidence From Bogotá's TransMilenio," University of California, Berkeley, mimeograph.

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.

WEBB, A. D. (1911): *The New Dictionary of Statistics*. Routledge and Sons, London.

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.

WOLMAR, C. (2009): *Fire and Steam: How the Railways Transformed Britain*. Atlantic, London Books.

——— (2012): *The Subterranean Railway: How the London Underground was Built and How it Changed the City Forever*. Atlantic, London Books.

TABLE I:
Gravity Equation Estimation Using 1921 Bilateral Commuting Data

	(1)	(2)
Second-stage Regression		
	$\log \lambda_{nit}$	$\log \lambda_{nit}$
$\log d_{nit}^{W}$	-4.899^{***}	-5.203^{***}
	(0.062)	(0.069)
Workplace fixed effects	yes	yes
Residence fixed effects	yes	yes
Kleibergen-Paap (p-value)		0.000
Estimation	OLS	IV
Observations	3023	3023
R-squared	0.851	-
First-stage Regression		
		$\log d_{nit}^{W}$
$\log d_{ni}^{S}$		0.429^{***}
		(0.003)
Workplace and Residence fixed effects		yes
First-stage F-statistic		22,235
Observations		3023
R-squared		0.949

Note: λ_{nit} is the commuting probability from equation (6); d_{nit}^{W} is our least-cost-path travel time measure based on the transport network; d_{ni}^{S} is straight-line travel time based on walking; Kleibergen-Paap is the p-value for the Kleibergen-Paap underidentification test; OLS refers to ordinary least squares; the second-stage R-squared is omitted from the instrumental variables (IV) specification (two-stage least squares), because it does not have a meaningful interpretation; First-stage F-statistic is the F-statistic for the joint significance of the excluded exogenous variables in the first-stage regression; Heteroskedasticity robust standard errors in parentheses: * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

TABLE II:
Estimation of Agglomeration Forces in Production and Residence

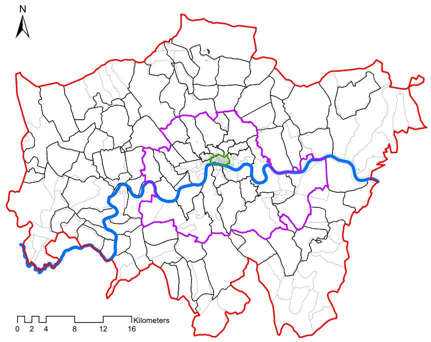
	(1)	(2)	(3)	(4)
	$\ln \widehat{\mathbb{A}}_{nt}^{T}$	$\ln \widehat{\mathbb{B}}_{nt}$	$\ln \widehat{\mathbb{A}}_{nt}^{T}$	$\ln \widehat{\mathbb{B}}_{nt}$
$\ln \widehat{L}_{nt}$	0.148^{***}	-	0.086^{**}	-
	(0.027)		(0.037)	
$\ln \widehat{R}_{nt}$	-	0.248^{***}	-	0.172^{***}
		(0.023)		(0.031)
$\ln L_{nt}$	0.029^{*}	-	0.011	-
	(0.017)		(0.017)	
$\ln R_{nt}$	-	0.033	-	-0.015
		(0.024)		(0.027)
$\ln K_{n}$	0.078^{***}	-0.067	0.092^{***}	-0.056
	(0.020)	(0.042)	(0.023)	(0.038)
$\mathbb{I}_{n}^{\text{LCC}}$	-0.112^{**}	0.085	-0.033	0.252^{***}
	(0.048)	(0.074)	(0.060)	0.089
First-stage F-statistic	-	-	11.26	12.76
Kleibergen-Paap (p-value)	-	-	0.000	0.000
Hansen-Sargen (p-value)	-	-	0.416	0.483
Estimation	OLS	OLS	IV	IV
Observations	99	99	99	99
R-squared	0.428	0.648	-	-

Note: $\ln \widehat{\mathbb{A}}_{nt}^{T} = \ln\left(\mathbb{A}_{n\tau}^{T}/\mathbb{A}_{nt}^{T}\right)$ is the log change in composite traded productivity; $\ln \widehat{\mathbb{B}}_{nt} = \ln\left(\mathbb{B}_{n\tau}/\mathbb{B}_{nt}\right)$ is the log change in composite amenities; $\ln \widehat{L}_{nt} = \ln\left(L_{n\tau}/L_{nt}\right)$ is the log change in workplace employment; $\ln \widehat{R}_{nt} = \ln\left(R_{n\tau}/R_{nt}\right)$ is the log change in residence employment; $\ln L_{nt}$ is initial log workplace employment; $\ln R_{nt}$ is initial log residence employment; $\ln K_{n}$ is log land area; $\mathbb{I}_{n}^{\text{LCC}}$ is an indicator variable for whether a borough is located within the London County Council (LCC) area; instruments are indicator variables for 5 km distance grid cells from the Guildhall in the center of the City of London; the excluded category is > 20 km from the Guildhall; first-stage F-statistic is the F-statistic for the joint significance of the distance grid cell indicators; Kleibergen-Paap is the p-value for the Kleibergen-Paap underidentification test; Hansen-Sargen is the p-value for the Hansen-Sargan overidentification test; OLS refers to ordinary least squares; the R-squared is omitted from the instrumental variables (IV) estimates (two-stage least squares) of the second-stage equation, because it does not have a meaningful interpretation; the results of the first-stage regression are reported in Table G.1 in Section G4 on the online appendix; Heteroskedasticity robust standard errors in parentheses: * denotes statistical significance at the 10 percent level; ** denotes statistical significance at the 5 percent level; *** denotes statistical significance at the 1 percent level.

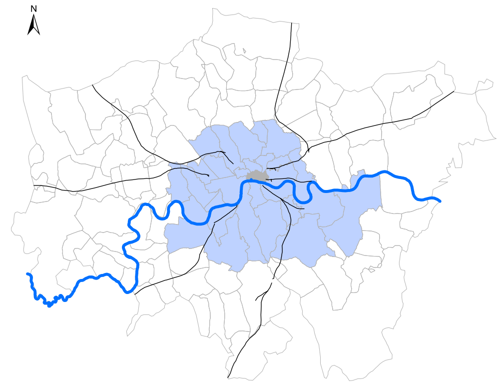
TABLE III:
Counterfactuals for Removing the Entire Railway Network, the Entire Underground Railway Network, or Railway Lines Constructed from 1911-21, Starting from the Initial Equilibrium in our Baseline Year of 1921

	(1)	(2)	(3)	(4)
Floor Space Supply Elasticity	$\mu = 0$	$\mu = 1.83$	$\mu = 1.83$	$\mu = 1.83$
Production Agglomeration Force	$\eta^L = 0$	$\eta^L = 0$	$\eta^L = 0.086$	$\eta^L = 0.086$
Residential Agglomeration Force	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0.172$
Removing the Entire Overground and Underground Railway Network				
Economic Impact				
Rateable Value	$-£8.24$m	$-£15.55$m	$-£20.78$m	$-£35.07$m
NPV Rateable Value (3 percent)	$-£274.55$m	$-£518.26$m	$-£692.76$m	$-£1,169.05$m
NPV Rateable Value (5 percent)	$-£164.73$m	$-£310.96$m	$-£415.66$m	$-£701.43$m
Construction Costs				
Cut-and-Cover Underground		$-£9.96$m		
Bored-tube Underground		$-£22.90$m		
Overground Railway		$-£33.19$m		
Total All Railways		$-£66.05$m		
Ratio Economic Impact / Construction Cost				
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	4.16	7.85	10.49	17.70
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	2.49	4.71	6.29	10.62
Removing the Entire Underground Railway Network				
Economic Impact				
Rateable Value	$-£2.65$m	$-£6.21$m	$-£8.22$m	$-£14.16$m
NPV Rateable Value (3 percent)	$-£88.46$m	$-£206.87$m	$-£274.05$m	$-£471.85$m
NPV Rateable Value (5 percent)	$-£53.08$m	$-£124.12$m	$-£164.43$m	$-£283.11$m
Construction Costs				
Cut-and-Cover Underground		$-£9.96$m		
Bored-tube Underground		$-£22.90$m		
Total All Underground		$-£32.86$m		
Ratio Economic Impact / Construction Cost				
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	2.69	6.30	8.34	14.36
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	1.62	3.78	5.00	8.62
Removing Overground and Underground Railway Lines Constructed from 1911-21				
Economic Impact				
Rateable Value	$-£0.17$m	$-£0.24$m	$-£0.37$m	$-£0.39$m
NPV Rateable Value (3 percent)	$-£5.63$m	$-£8.09$m	$-£12.46$m	$-£12.96$m
NPV Rateable Value (5 percent)	$-£3.38$m	$-£4.86$m	$-£7.47$m	$-£7.77$m
Construction Costs				
Cut-and-Cover Underground		$-£0.00$m		
Bored-tube Underground		$-£2.35$m		
Overground Railway		$-£0.34$m		
Total All Railways		$-£2.69$m		
Ratio Economic Impact / Construction Cost				
$\frac{\text{NPV Rateable Value (3 percent)}}{\text{Construction Cost}}$	2.09	3.01	4.63	4.82
$\frac{\text{NPV Rateable Value (5 percent)}}{\text{Construction Cost}}$	1.26	1.81	2.78	2.89

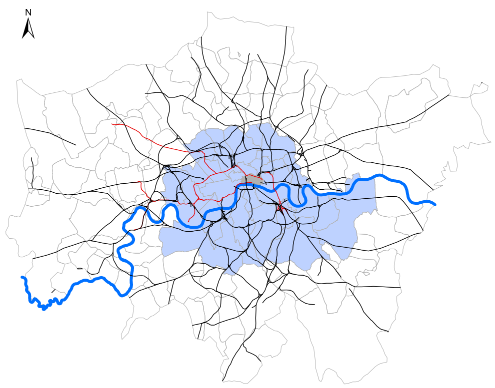
Note: Counterfactuals start in our baseline year of 1921 and remove either the entire railway network or parts thereof; we hold the omnibus and tram network constant at its 1921 structure; all values reported in the table are expressed in millions of 1921 pounds sterling; $\mu = 0$ corresponds to an inelastic supply of floor space; $\mu = 1.83$ is our calibrated floor space supply elasticity; $\eta^L = 0$ corresponds to no production agglomeration force; $\eta^R = 0$ corresponds to no residential agglomeration force; $\eta^L = 0.086$ corresponds to our estimated production agglomeration force; $\eta^R = 0.172$ corresponds to our estimated residential agglomeration force; all specifications assume population mobility between Greater London and the wider economy, with the elasticity of population supply determined by our calibrated Fréchet shape parameter of $\epsilon = 5.25$; net present values are evaluated over an infinite lifetime, assuming either 3 or 5 percent discount rate; construction costs are based on capital issued per mile for cut-and-cover, bored-tube and surface railway lines and the length of lines of each type of railway in Greater London in 1921, as discussed further in Section J6 of the online appendix.



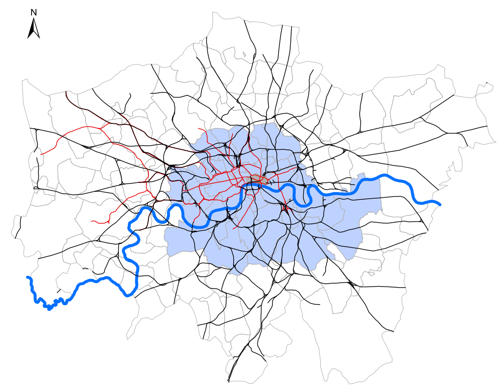
(a) Administrative Boundaries



(b) Railway Network 1841



(c) Railway Network 1881



(d) Railway Network 1921

FIGURE I:
Administrative Boundaries and the Railway Network over Time

Note: Panel (a): Greater London Authority (GLA) referred to as Greater London (red outer boundary); London County Council (LCC) (purple outer boundary); City of London (green outer boundary); River Thames (thick blue); boroughs (medium black lines); and parishes (medium gray lines). Panels (b)-(d): Greater London outside County of London (white background); County of London outside City of London (blue background); City of London (gray background); River Thames shown in blue; overground railway lines shown in black; underground railway lines shown in red.

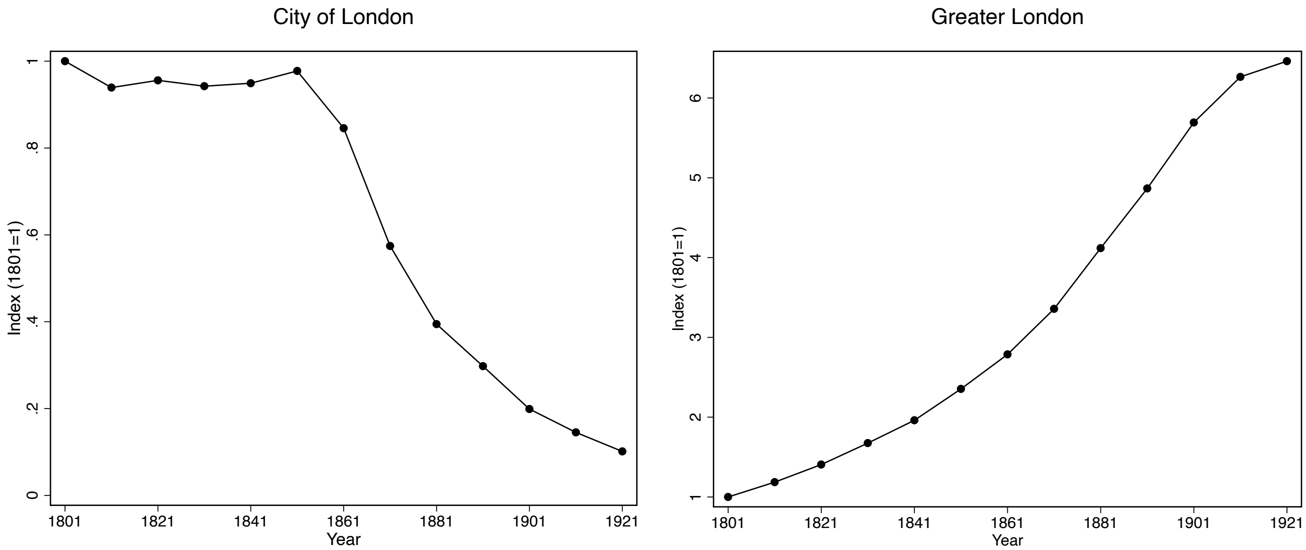


FIGURE II:
Population Indexes Over Time (City of London and Greater London, 1801 equals 1)

Note: Indexes of residence (night) population from the population censuses over time.

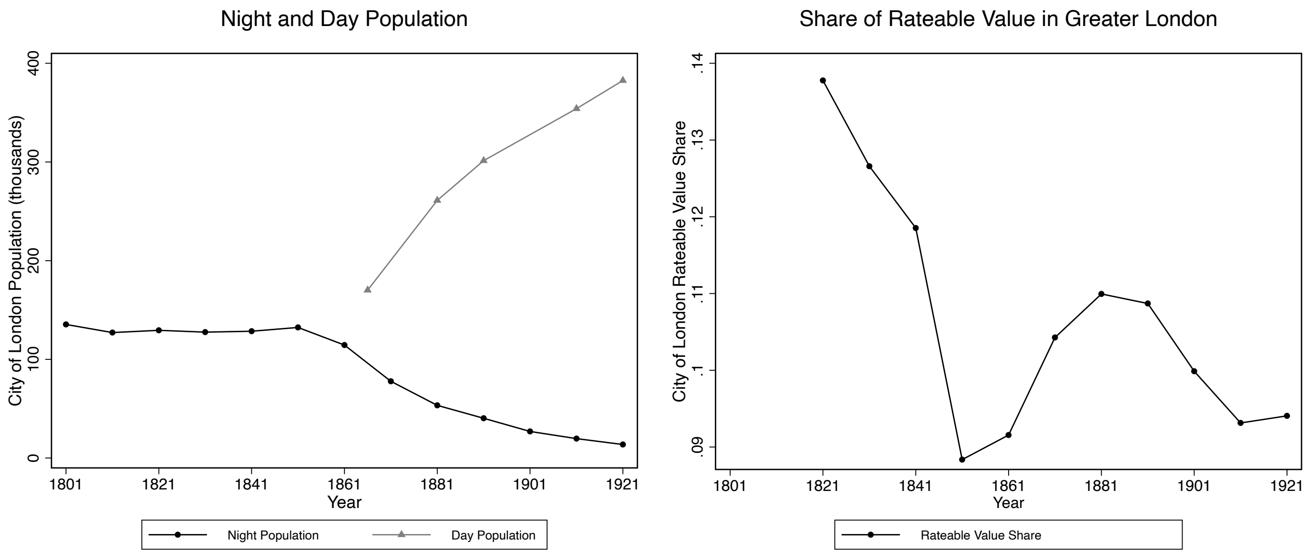
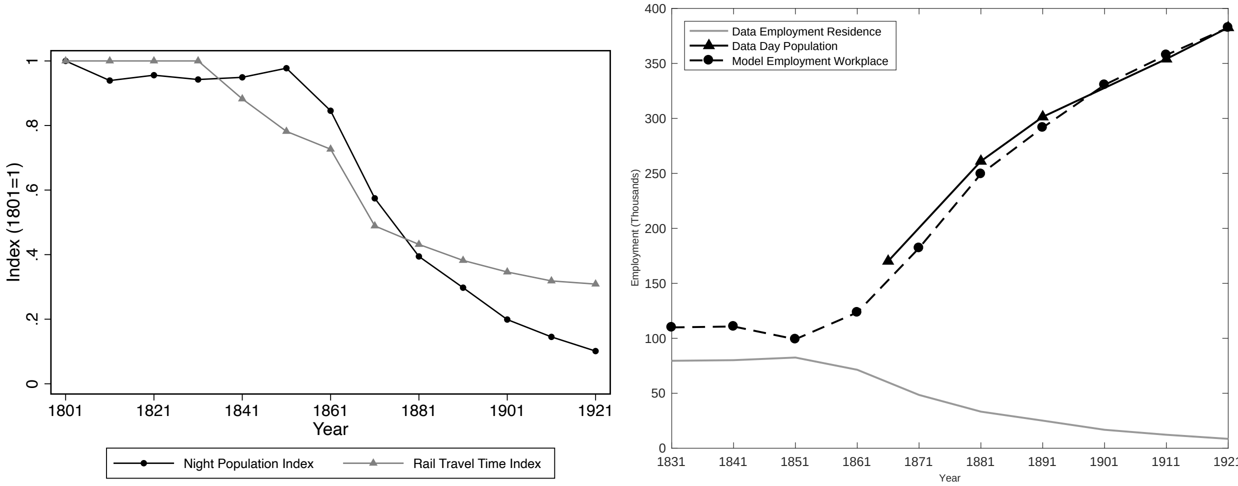


FIGURE III:
City of London Day and Night Population and Rateable Value Share Over Time

Note: Residence (night) population from the population census; day population from the City of London day census for 1866, 1881, 1891 and 1911 and from workplace employment in the population census for 1921; City of London's share of rateable value is its share in the total value of all land and buildings in Greater London.



(a) Population-Share-Weighted Average Travel Time

(b) Night and Day Population in the City of London

FIGURE IV:
Population-Share-Weighted Average Travel Time and Day and Night Population in the City of London

Note: Left panel (a): Night (residence) population index from the population census; rail travel time index is the population-share-weighted average of the reduction in travel time from each borough to the City of London (relative to 1801). Right panel (b): "Data Night Population" is residential population from the population census; "Data Day Population" is day population from the City of London Day Census for 1866, 1881, 1891 and 1911 and workplace employment in the population census for 1921; "Model Employment Workplace" equals the data on workplace employment from the population census for 1921 and the model's predictions for workplace employment using our calibrated Fréchet shape parameter of $\epsilon = 5.25$ for the other years; we calibrate this Fréchet shape parameter of $\epsilon = 5.25$ by minimizing the sum of squared deviations between the model's predictions for workplace employment and the data on the day population of the City of London in the census years for which these data are available (1881, 1891 and 1911); the model's predictions use the estimated change in commuting costs from removing the railway network and condition on the observed changes in population and rateable values in the data.

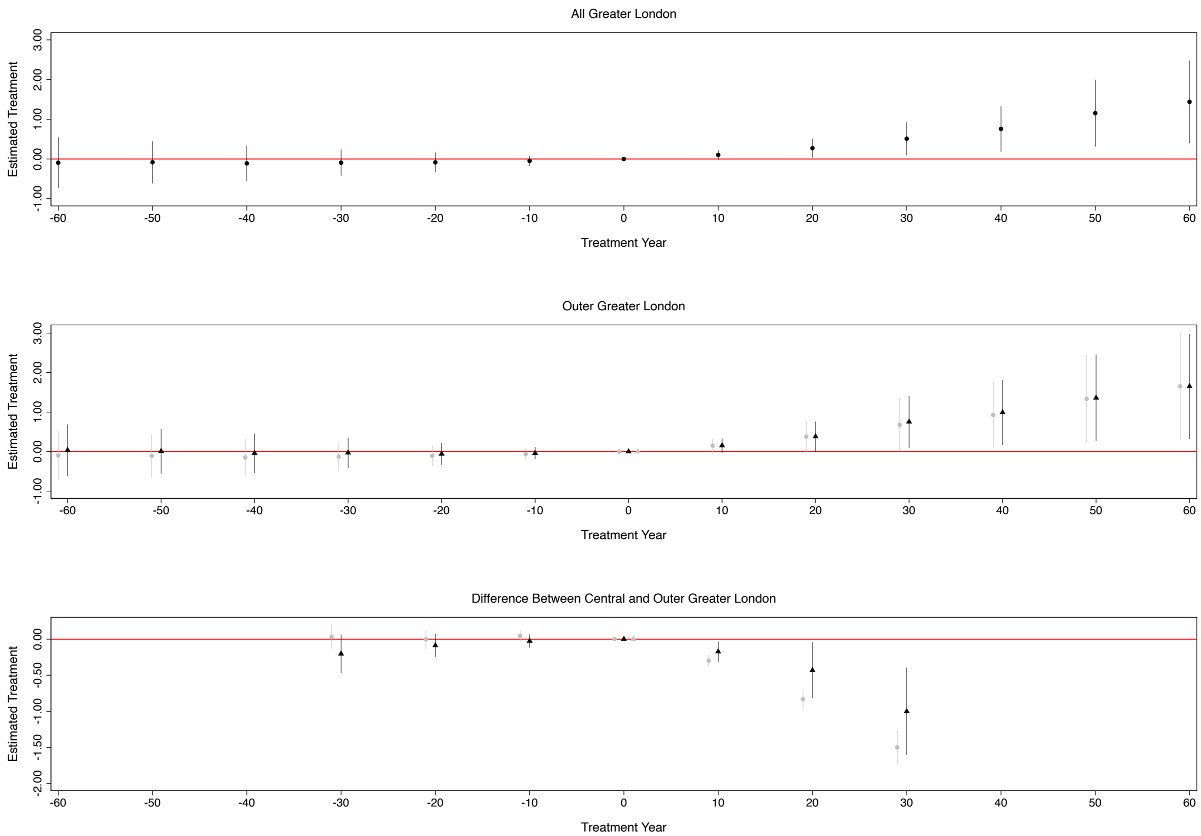


FIGURE V:
Event-Study Treatment Effects for the Arrival of an Overground or Underground Railway Station in Greater London Parishes from 1801-1901

Note: Estimated treatment effects from the arrival of an overground or underground railway station on log parish population; sample includes 283 parishes in Greater London in census years from 1801-1901 (every 10 years); all specifications include parish fixed effects, year fixed effects and parish-specific time trends; estimated coefficients and standard errors are reported in Table I.1 in the online appendix; in all three panels, the horizontal axis shows the treatment year τ, which equals census year minus the last census year in which a parish had no railway, so that positive values of τ correspond to post-treatment years; the excluded category is treatment year $\tau = 0$; top panel shows estimated treatment effects across all parishes in Greater London (the coefficients β_τ from equation (1)); middle panel shows estimated treatment effects for parishes outside central London (the coefficients β_τ from equation (2)); lower panel shows the difference in estimated treatment effects between parishes in central and outer London (the coefficients γ_τ from equation (2)); the vertical lines in each panel show the estimated 95 percent confidence intervals, based on standard errors clustered on boroughs; in the middle and lower panels, the first specification (gray circle markers) estimates equation (2) using the City of London as the definition of central London, while the second specification (black triangle markers) estimates equation (2) using parishes with centroids less than 5 kilometers from the Guildhall as the definition of central London, and the third specification (gray triangle markers) estimates equation (2) augmented with travel time controls and using parishes with centroids less than 5 kilometers from the Guildhall as the definition of central London; the travel time controls are the log population-weighted average travel time reduction to other parishes from the railway network and its interaction with the log of straight-line distance from the centroid of each parish to the Guildhall.

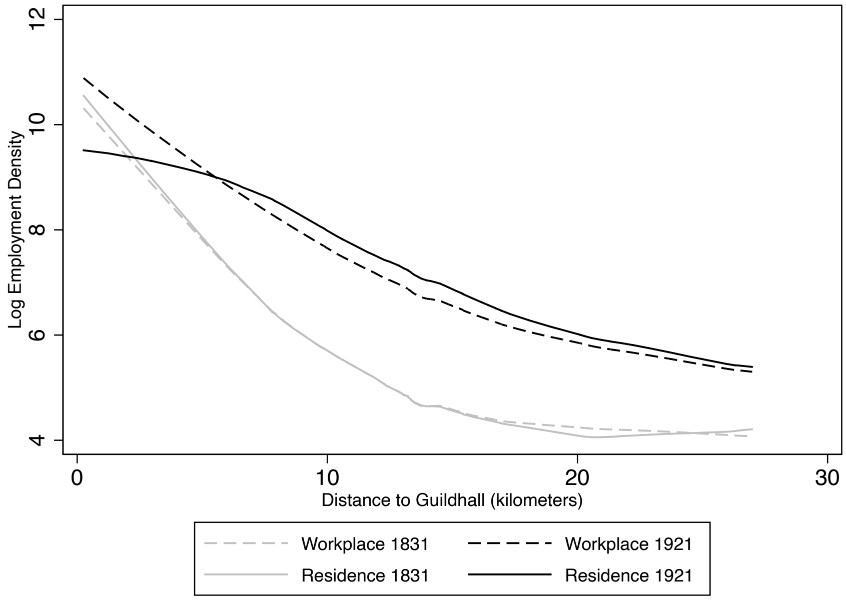
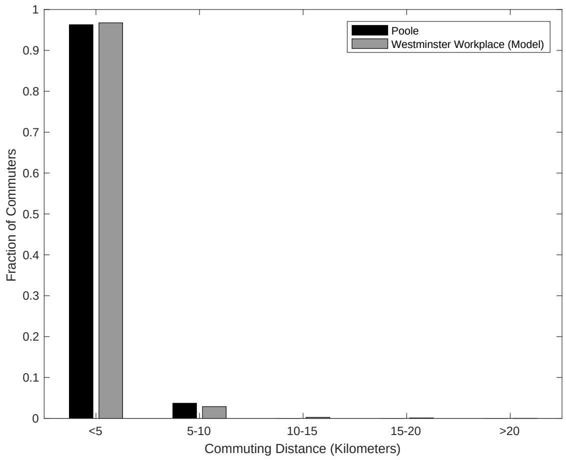
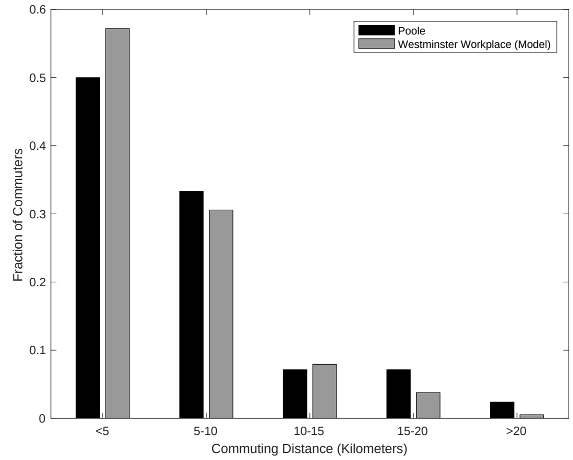


FIGURE VI:
Employment Workplace and Residence Density and Distance to the Guildhall in 1831 and 1921

Note: Employment residence density is measured per unit of land area and the values for both 1831 and 1921 are from the population census data. Employment workplace density is measured per unit of land area; the values for 1921 are from the population census data; the values for 1831 are model predictions from our baseline quantitative analysis. Figure shows the fitted values from locally-weighted linear least squares regressions of the log of each variable on distance to the Guildhall.



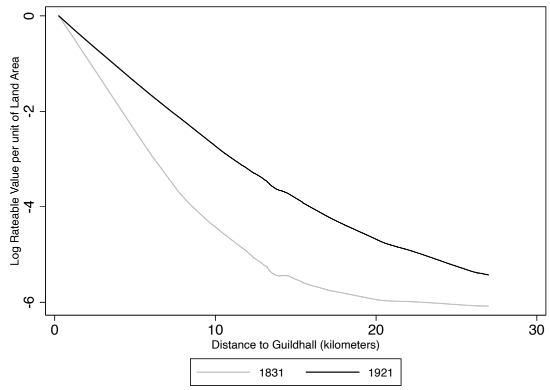
(a) 1857-1877



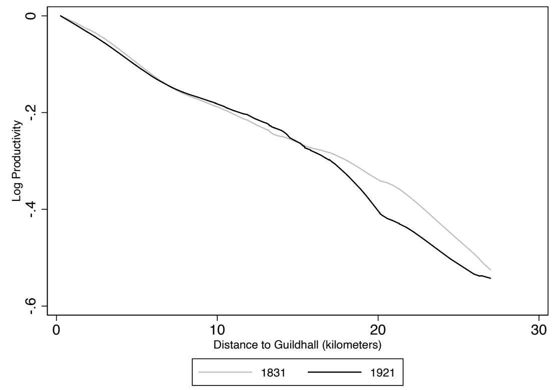
(b) 1891-1911

FIGURE VII:
Commuting Distances in the Model and the Henry Poole Data

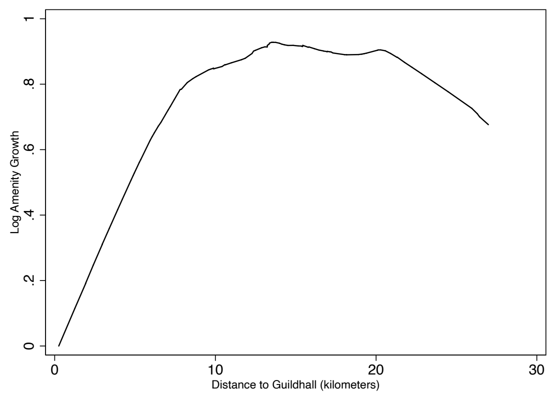
Note: Shares of workers by commuting distance for all workers employed in the borough of Westminster in the model and for workers employed by Henry Poole, Westminster. Model predictions are for 1861 and 1901. Henry Poole data are for workers hired in 1857-1877 and 1891-1911.



(a) Rateable Value per unit of Land Area



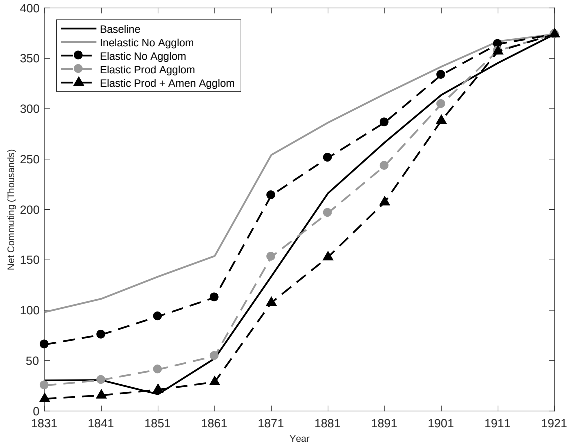
(b) Productivity



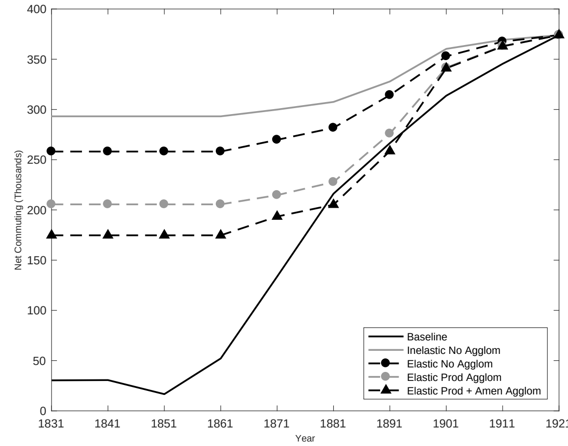
(c) Amenity Growth 1831-1921

FIGURE VIII:
Rateable Values and Productivity in 1831 and 1921 and Amenity Growth from 1831-1921

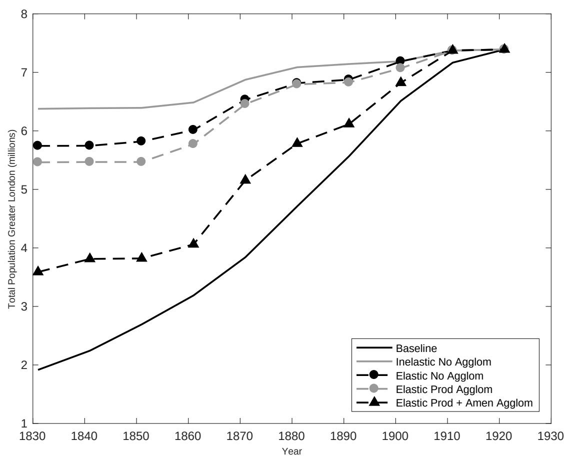
Note: Figure shows the fitted values from locally-weighted linear least squares regressions of the log of each variable on distance to the Guildhall. The fitted values for each variable are normalized such that they take the value zero for the City of London.



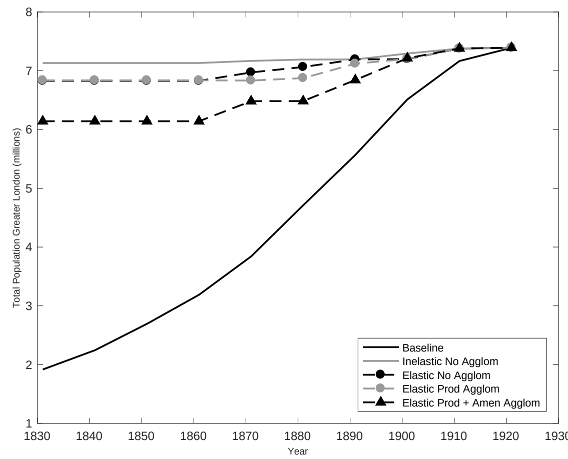
(a) Net Commuting into City of London (All Rail)



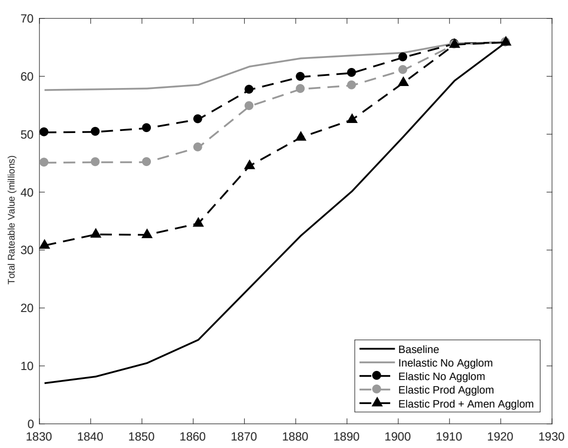
(b) Net Commuting into City of London (Underground)



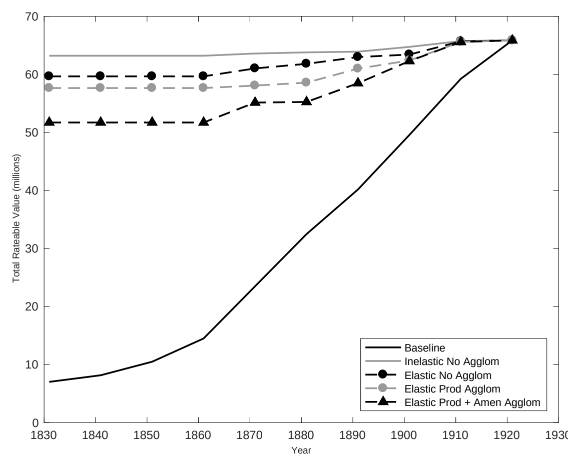
(c) Population Greater London (All Rail)



(d) Population Greater London (Underground)



(e) Rateable Value Greater London (All Rail)



(f) Rateable Value Greater London (Underground)

FIGURE IX:
Counterfactuals for Removing the Railway Network

Note: "Baseline" shows the values of variables from our baseline quantitative analysis from Section VI.; "Inelastic No Agglom" shows the values of variables from our counterfactual with a perfectly inelastic supply of floor space and exogenous productivity and amenities ($\mu = \eta^L = \eta^R = 0$); "Elastic No Agglom" shows the values of variables from our counterfactual using our calibrated floor space supply elasticity and exogenous productivity and amenities ($\mu = 1.83$ and $\eta^L = \eta^R = 0$); "Elastic Prod Agglom" shows the values of variables from our counterfactual using our calibrated floor space supply elasticity, estimated production agglomeration forces and exogenous amenities ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0$); "Elastic Prod + Amen Agglom" shows the values of variables from our counterfactual using our calibrated floor space supply elasticity and estimated production and residential agglomeration forces ($\mu = 1.83$, $\eta^L = 0.086$ and $\eta^R = 0.172$).