

DISCUSSION PAPER SERIES

DP17839

THE ECONOMICS OF CITIES: FROM THEORY TO DATA

Stephen Redding

**INTERNATIONAL TRADE AND
REGIONAL ECONOMICS**

CEPR

THE ECONOMICS OF CITIES: FROM THEORY TO DATA

Stephen Redding

Discussion Paper DP17839
Published 23 January 2023
Submitted 16 January 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- International Trade and Regional Economics

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Stephen Redding

THE ECONOMICS OF CITIES: FROM THEORY TO DATA

Abstract

Economic activity is highly unevenly distributed within cities, as reflected in the concentration of economic functions in specific locations, such as finance in the Square Mile in London. The extent to which this concentration reflects natural advantages versus agglomeration forces is central to a range of public policy issues, including the impact of local taxation and transport infrastructure improvements. This paper reviews recent quantitative urban models, which incorporate both differences in natural advantages and agglomeration forces, and can be taken directly to observed data on cities. We show that these models can be used to estimate the strength of agglomeration forces and evaluate the impact of transportation infrastructure improvements on welfare and the spatial distribution of economic activity.

JEL Classification: R32, R41, R52

Keywords: N/A

Stephen Redding - reddings@princeton.edu
Princeton University, NBER, CEPR and CEPR

Acknowledgements

I am grateful to Princeton University for research support. Thanks to Nan Xiang for excellent research assistance. This paper was commissioned for the Journal of Economic Perspectives. I am very grateful to the editors, Gabriel Ahlfeldt, Treb Allen, Jason Barr, Dave Donaldson and Clifford Winston for helpful comments and suggestions.

The Economics of Cities: From Theory to Data*

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

January 16, 2023

Abstract

Economic activity is highly unevenly distributed within cities, as reflected in the concentration of economic functions in specific locations, such as finance in the Square Mile in London. The extent to which this concentration reflects natural advantages versus agglomeration forces is central to a range of public policy issues, including the impact of local taxation and transport infrastructure improvements. This paper reviews recent quantitative urban models, which incorporate both differences in natural advantages and agglomeration forces, and can be taken directly to observed data on cities. We show that these models can be used to estimate the strength of agglomeration forces and evaluate the impact of transportation infrastructure improvements on welfare and the spatial distribution of economic activity.

Keywords: cities, commuting, transportation, urban economics

JEL Classification: R32, R41, R52

*I am grateful to Princeton University for research support. Thanks to Nan Xiang for excellent research assistance. This paper was commissioned for the *Journal of Economic Perspectives*. I am very grateful to the editors, Gabriel Ahlfeldt, Treb Allen, Jason Barr, Dave Donaldson and Clifford Winston for helpful comments and suggestions.

[†]Dept. Economics and SPIA, JRR Building, Princeton, NJ 08544. Email: reddings@princeton.edu.

1 Introduction

One of the most striking features of our world is the uneven distribution of economic activity across space. This concentration is most evident in the existence of cities. By 2018, 55 percent of the world's population lived in urban areas, with one in eight urbanites residing in 33 megacities with more than 10 million inhabitants. But similar concentration is observed even within cities. In 2010, the top 10 percent of census tracts in New York City (NYC) with the highest land value per square mile accounted for 60 percent of total land value and 64 percent of employment, but only 14 percent of population and 7 percent of land area. More broadly, examples abound of economic functions that cluster in specific locations within cities, such as advertising agencies in midtown Manhattan, or finance in the Square Mile in London.¹

The defining characteristic of cities is arguably the absence of physical space between people and firms, as they cluster together. Economists distinguish between two types of explanations for this concentration: first-nature and second-nature geography. First-nature geography corresponds to exogenous natural advantages or locational fundamentals, such as access to natural water, or proximity to a deep natural harbor. Second-nature geography corresponds to the location of economic agents relative to one another in geographic space. According to this second class of explanations, people and firms endogenously choose to locate together in order to eliminate transport costs for goods, people and ideas, even in the absence of any differences in exogenous natural advantage. From this second perspective, the concentration of economic activity in cities reflects a self-sustaining process of agglomeration, in which the location decisions of agents mutually reinforce one another.

Which of these two explanations drives the observed concentration of economic activity in cities is central to a range of economic issues and public policy debates. Explanations based on agglomeration forces typically feature externalities, such that when one agent makes a location decision, she does not take into account its effect on another agent's location decision. These externalities can be either technological (e.g., knowledge spillovers between agents) or pecuniary in the sense that they are mediated through markets (e.g., demand for locally-traded goods and services). In the presence of these externalities, the market equilibrium is generically inefficient, and there is the potential for public policy interventions to be welfare improving. Therefore, determining the strength of agglomeration forces is central to understanding the impact of local taxation, policies differentiated by location (place-based policies), zoning and building regulations, and transport infrastructure improvements, among other policies. If these agglomeration forces are sufficiently strong, there can be multiple equilibrium patterns of economic activity,

¹Figures in this first paragraph from [United Nations \(2018\)](#), Primary Land Use Tax Lot Output, NYC Department of City Planning, US population census, and LEHD Origin-Destination Employment Statistics (LODES).

such that even small public policy interventions can have substantial effects, by shifting the location of economic activity between multiple equilibria.

Alongside these agglomeration forces, the concentration of economic activity also gives rise to offsetting dispersion forces. As economic activity clusters in one location, this bids up the price of local factors of production that are in inelastic supply, such as land. Economic agents can respond to higher land prices in a number of ways. First, they can live in smaller dwellings, consuming less floor space per person. Second, they can increase the supply of floor space for a given amount of land by building taller buildings. Third, they can commute from other less densely-populated locations with lower land prices. But each of these responses is costly, because lower floor space use reduces welfare, taller buildings incur greater construction costs, and commuting involves real resource and time costs. More broadly, the concentration of economic activity can give rise to a variety of sources of congestion, or facilitate the spread of disease between people, both of which can act as dispersion forces.²

The observed distribution of economic activities within cities reflects the three-way interaction between natural advantages, agglomeration forces and dispersion forces. A key concept for understanding this three-way interaction is the notion of spatial equilibrium. In the simplest setting in which people are all the same, in order for some of them to be willing to pay the higher land prices to live in densely-populated locations, these higher land prices must be offset by either higher wages or higher amenities, such that people are indifferent across all populated locations. Additionally, if all firms produce the same homogenous good and markets are competitive, then in order for firms to pay the higher land prices and wages in densely-populated locations, these higher costs must be offset by higher productivity, such that firms make zero profits across all locations with positive production. Therefore, a key insight from spatial equilibrium is that the concentration of economic activity ultimately must be explained by either higher productivity or higher amenities, where both are influenced by first-nature geography (natural advantages) or second-nature geography (agglomeration forces).

The complexity of modeling agglomeration forces in spatial equilibrium has meant that the theoretical literature on cities has traditionally focus on stylized settings, such a one dimensional city on a line, or a perfectly symmetric circular city. In the canonical Alonso-Muth-Mills model, all employment is assumed to be concentrated in a central business district (CBD), and workers face commuting costs in travelling to work there. As a result of these commuting costs, the most attractive places to live are those closest to the CBD. Therefore, in a spatial equilibrium in which workers are indifferent across locations, workers trade off lower land prices further from the CBD with higher commuting costs. In such an equilibrium, land prices exhibit a monocentric struc-

²For a recent discussion of the implications of the spread of disease, social distancing and remote work for the future development of cities, see [Glaeser and Cutler \(2021\)](#).

ture, with a land price peak at the CBD, and a land price gradient that declines monotonically with distance from the CBD. Subsequent theoretical research relaxed the assumption that all employment is concentrated in the CBD, but retained the assumption of a stylized geography, such as a line or a symmetric circle. In these more general specifications, non-monocentric patterns of economic activity can emerge, with alternating residential and commercial land use.

Although these stylized settings reveal important mechanisms, real world cities are not well approximated by a one-dimensional line or a perfectly symmetric circle, which limits the usefulness of these theories for empirical work. In observed data on cities, land prices can fluctuate dramatically between high and low values across proximate neighborhoods. Moving outwards a city's center, the land price gradient can vary substantially between different segments of the city, as for example between the West and East Ends of London. Some parts of a city may have access to natural water and be well suited for heavy industrial use. Other parts of a city may have access to open space and scenic views and be well disposed for residential use. Yet other parts of a city may have good transport connections and be accessible for retail activity. Even with each of these different parts of the city, as one walks from one city block to another, land use can change sharply, from residential to commercial land use, and back again.

A key breakthrough in recent research has been the development of quantitative urban models that are able to rationalize these observed features of the data. These frameworks can accommodate many locations that differ in productivity, amenities, land area, the supply of floor space and transport connections. They allow for differences in productivity and amenities because of both natural advantages and agglomeration forces. At the same time, these frameworks are sufficiently tractable that they permit a mathematical analysis of their properties, such as the conditions under which there is a unique equilibrium versus multiple equilibria in the model. Additionally, these models have a small number of structural parameters that can be estimated using credible sources of exogenous variation. Given the estimated parameter values, these models typically have the property that they can be inverted to recover the unobserved natural advantages that exactly rationalize the observed data on land prices, workers and residents as an equilibrium of the model. Since these frameworks are able to rationalize the observed data, they can be used to undertake counterfactuals for realistic public policy interventions, such as the construction of a new subway line between specific locations.

Given the richness and flexibility of these quantitative urban models, they have been used to analyze a whole host of issues in urban economics, including the strength of agglomeration forces, the impact of transport infrastructure improvements, the sorting of heterogeneous groups of workers across space, and congestion pricing, among many others. Given an exogenous shock within-sample, such as the division of Berlin by the Berlin Wall or the invention of a new transport technology, the counterfactual predictions of these models for the impact of the shock on the

location of economic activity can be compared to the observed impact in the data in a model validation exercise. Naturally, any model is an abstraction, and there are many other idiosyncratic factors that can change over time in the data, but the ability of these quantitative urban models to successfully capture the observed impact of such within-sample shocks provides support for their use in predicting the impact of counterfactual public policies.

The remainder of the paper is structured as follows. Section 2 reviews a number of stylized facts about the organization of economic activity within cities, using NYC as an example. Section 3 summarizes the traditional theoretical literature on the internal structure of cities. Section 4 introduces recent quantitative urban models. Section 5 provides two applications of these quantitative urban models to estimate the strength of agglomeration and dispersion forces and evaluate the impact of transport infrastructure improvements. Section 6 concludes.

2 Stylized Facts

This section of the paper reviews key stylized facts about the spatial distribution of economic activity within cities, using census tract data for the five boroughs of NYC.

2.1 Land Prices

Land prices provide a summary statistic of the relative attractiveness of locations, because they are determined by competition between alternative uses of land. Many cities with long histories of settlement (e.g., London) have a well-defined central business district (CBD) with a single peak of land prices (e.g., the Square Mile), and are often referred to as monocentric. In contrast, in other cities that developed more recently (e.g., Los Angeles), the land price gradient has multiple peaks (e.g., Downtown and Beverly Hills), and hence these cities are termed polycentric. Although these terms are informative, they do not do full justice to the rich asymmetric patterns of variation in land prices, both within and across neighborhoods within cities.

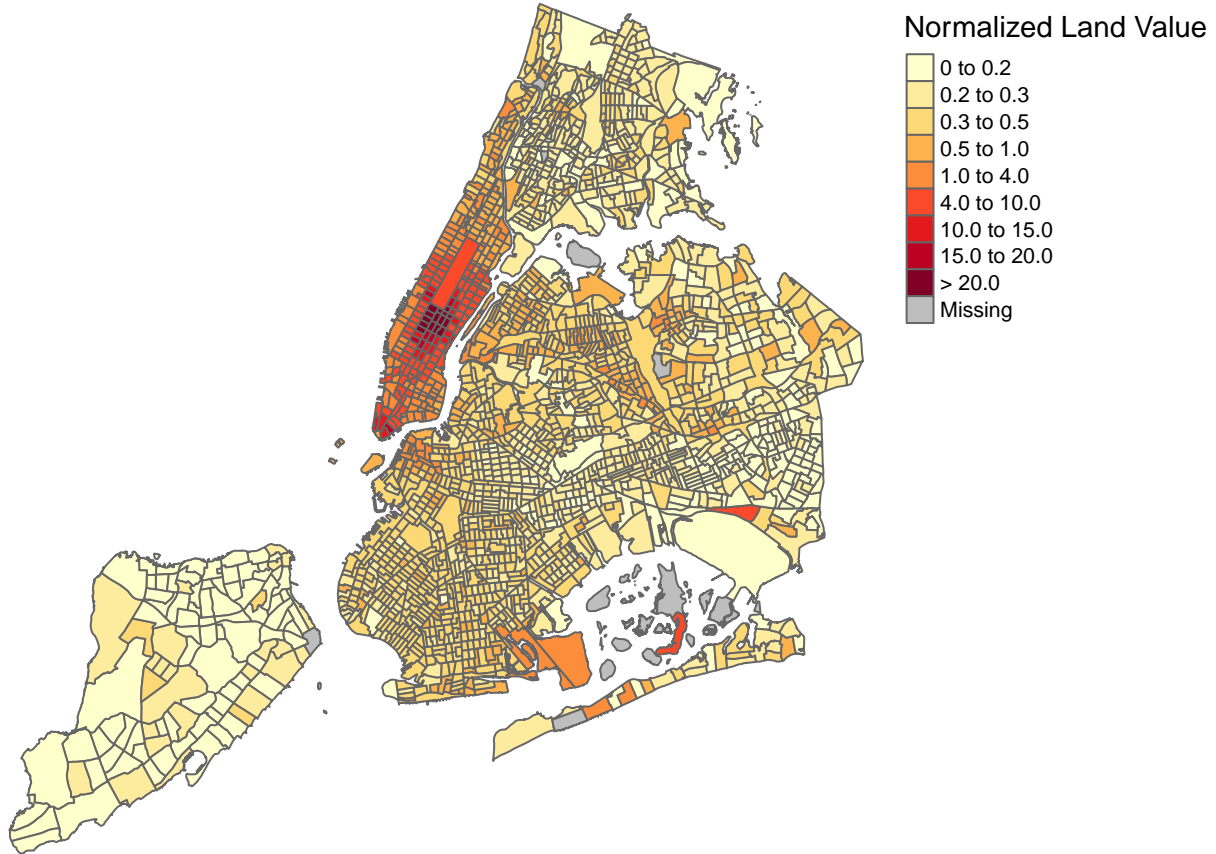
Figure 1 provides an illustration by displaying land prices in 2011 for each census tract in the five boroughs of NYC.³ Land prices are measured using the assessed tax value of the land per square meter, but similar patterns are observed using land prices estimated from property transactions data.⁴ As apparent from the figure, Midtown Manhattan has by far the highest land prices, with a smaller secondary peak in downtown Manhattan, and an area of lower land values in between. Towards the bottom of Manhattan, the Lower East Side has noticeably lower land prices than other nearby neighborhoods. Across the East River, central Brooklyn is the site of

³We use 2010 census tracts and report results for 2011, before the recent Covid-19 pandemic.

⁴For evidence from land prices estimated using property transactions data, see for example [Barr and Kulkarni \(2018\)](#) and [Haughwout and Bedoll \(2008\)](#).

another smaller peak in land prices. Finally, the areas bordering Central Park are relatively more expensive than those further away from the park.

Figure 1: Relative Land Prices per Square Mile in New York City in 2011

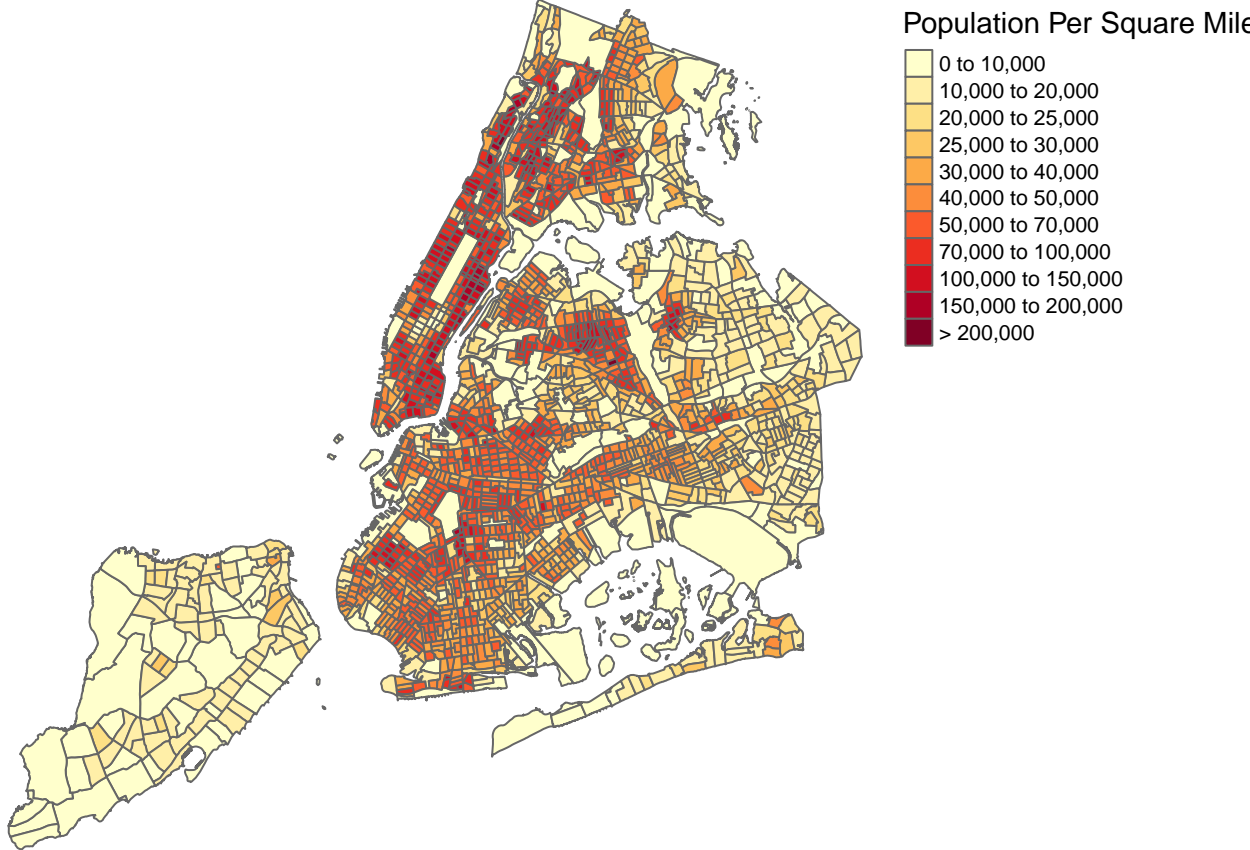


Note: Assessed land values per square mile in 2011, normalized by the mean across census tracts in New York City (the five boroughs of Bronx, Brooklyn, Manhattan, Queens and Staten Island). Therefore a value above one corresponds to an above average land value per square mile. Land values are from property taxation assessments. Source: Primary Land Use Tax Lot Output, NYC Department of City Planning.

The higher land prices in some neighborhoods than in others can be explained in terms of the demand for either commercial or residential land use. These demands for land use are in turn determined by productivity and amenities, and hence by natural advantages and agglomeration forces. Some of the patterns visible in Figure 1 suggest a role for exogenous natural advantage. For example, the waterfront areas around the edge of Manhattan historically had locational advantages for production, for port facilities, warehousing, and industrial processing, which could have long-lived effects on land values. In contrast, Central Park in the midst of the densely-populated city is an important natural amenity for consumption. However, even among areas with apparently similar natural advantages, we observe substantial differences in land prices, suggesting a potential role for endogenous agglomeration forces.

Population Density The differences in land prices shown in Figure 1 are accompanied by substantial differences in population density. In Figure 2, we display the number of people per square mile in 2010 for each census tract in NYC. Across the five boroughs that make up the city, population density varies substantially. Manhattan is the most densely-populated county in the United States: In 2010, 1,518,500 people lived in an area of 22.8 square miles, with a population density of 66,579 people per square mile. In contrast, Staten Island is relatively sparsely populated, with a population density an order of magnitude smaller at 7,923 people per square mile.⁵ Even within Manhattan, population density displays dramatic variation, with relatively low densities in the commercial districts of midtown and downtown, and relatively high densities in the residential suburbs of the Upper West and Upper East Sides.

Figure 2: Population Density per Square Mile in New York City in 2010



Note: Population density per square mile in 2010 for each census tract in the five boroughs of New York City (the five boroughs of Bronx, Brooklyn, Manhattan, Queens and Staten Island). Source: US population census.

This variation in population density is also influenced by productivity and amenities, as again determined by natural advantages and agglomeration forces. Other things equal, high productiv-

⁵Even so, Staten Island is densely-populated relative to many rural locations in the United States, with the state of Wyoming having a population density in 2010 of 5.63 people per square mile.

ity or amenities in a location attracts people, increasing population density, and bidding up land prices. These higher land prices in turn affect equilibrium city structure in a number of ways. First, residents can substitute away from the use of floor space, resulting in smaller dwellings. Second, higher land prices raise the return to constructing taller buildings, as reflected in the many skyscrapers that make up Manhattan's skyline.⁶ Third, workers can commute from other less densely-populated locations with lower land prices. All three responses are costly, since lower residential floor space use reduces utility, taller buildings incur greater construction costs, and commuting involves monetary and time costs.

Public policy can play an important role in influencing these responses. First, zoning and building regulations affect the elasticity of the supply of floor space with respect to an increase its price (referred to as the housing supply elasticity). Other things equal, the lower this housing supply elasticity, the lower the supply of floor space, the higher the price of floor space, and the lower population density. Second, transport infrastructure investments in the form of highway construction and public transit provision can provide an alternative to increasing housing supply in central cities, by expanding the city's geographical boundaries and allowing people to commute longer distances from further away.

Workers and Residents When workers commute, they separate their residence and workplace to take advantage of high wages at their workplace and a low cost of living net of amenities and housing costs at their residence. This separation of workplace and residence is made possible by modern transportation technologies, such as overground and underground railway networks and highway systems, which are capable of transporting large numbers of people each day between their home and place of work. Indeed, remote working can be viewed as a communication technology that further facilitates this separation of workplace and residence, by enabling workers for example to combine the productivity advantages of a job in San Francisco with the amenity advantages of a home on Lake Tahoe.⁷

An implication of this separation of workplace and residence is the specialization of locations between residential and commercial land use. Areas with high productivity relative to amenities can specialize as workplaces, while those with high amenities relative to productivity can specialize as residences. The result is a rich internal structure of economic activity within cities. This specialization occurs at the macro scale of NYC and its economic hinterland of the three states of Connecticut, New Jersey and New York: In 2010, NYC was a net importer of 0.5 million workers from the rest of the tri-state area (with a inflow of 0.97 million and an outflow of 0.46 million). This specialization also occurs at the intermediate scale of the five boroughs of the city: In the

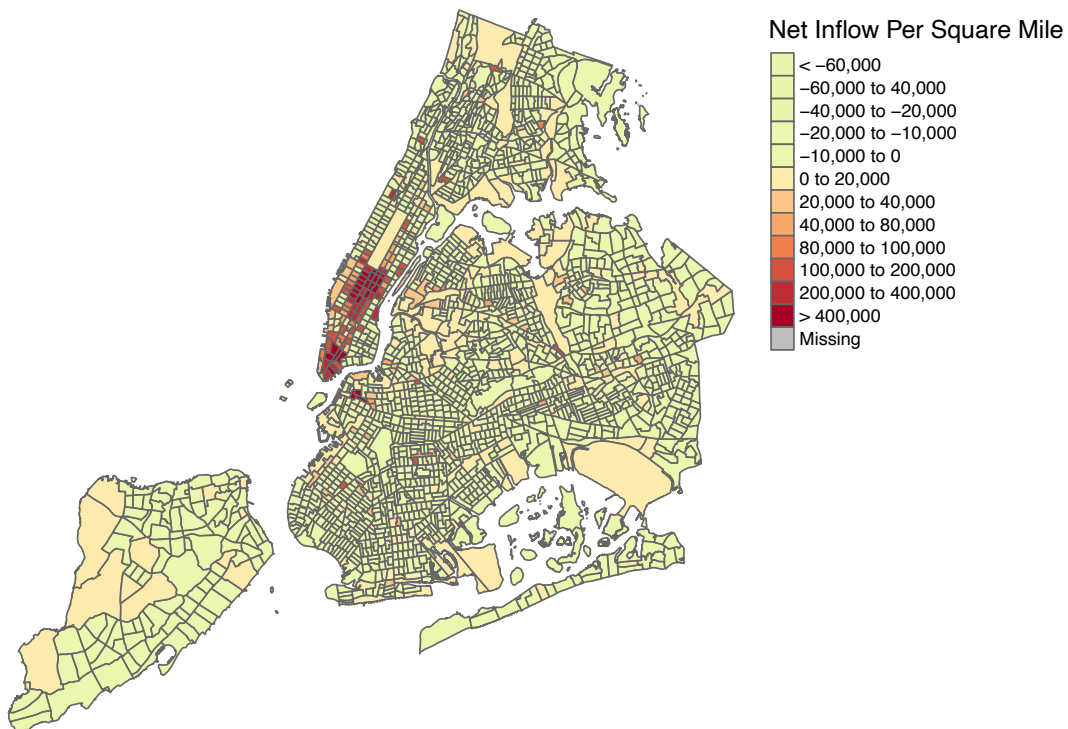
⁶See for example the discussion of Manhattan's skyscrapers in [Barr \(2016\)](#) and [Ahlfeldt and Barr \(2020\)](#).

⁷For further discussion of recent trends in working from home and their implications for the organization of economic activity within cities, see [Barrero et al. \(2021\)](#), [Ramani and Bloom \(2021\)](#) and [Gupta et al. \(2022a\)](#).

same year, Manhattan alone was a net importer of 1.4 million workers from the rest of NYC and the tri-state area (with a inflow of 1.6 million and an outflow of 0.19 million).

Figure 3 provides further evidence on this location specialization by showing net imports of commuters (workers minus residents) per square mile in 2010 for each census tract in New York City. Workers corresponds to employment by workplace, which is the sum of all in-commuting flows to a census tract (including from the census tract itself). Residents equals employment by residence, which is the sum of all out-commuting flows from a census tract (including to the census tract itself). A positive value implies that a census tract is a net importer of commuters, while a negative value implies that it is a net exporter of commuters.

Figure 3: Net Inflow of Commuters (Workers Minus Residents) Per Square Mile in New York City in 2010



Note: Net inflow of commuters (workers minus residents) per square mile; Workers equals employment by workplace, which is the sum of all inward commuting flows into a census tract from anywhere in the United States (including from the census tract itself). Residents equals employment by residence, which is the sum of all outward commuting flows from a census tract to anywhere in the United States (including to the census tract itself). Negative values represent net exports of commuters and positive values correspond to net imports of commuters. Data on commuting flows from the LEHD Origin-Destination Employment Statistics (LODES).

Location specialization is even more dramatic at this micro scale of individual census tracts. The two land price peaks of midtown and downtown Manhattan are highly-specialized commercial districts, with net imports of commuters greater than 400,000 per square mile.⁸ These values

⁸This figure is a density per land area, where census tracts can be much smaller than a square mile. Maximum and minimum net imports of commuters (without dividing by land area) are 190,292 and -7,390, respectively.

are substantially larger than the highest population densities visible in Figure 2, which reflects the fact that employment by workplace is much more spatially concentrated than population. Many of the commuters working in these commercial districts of Manhattan travel long distances from outside NYC. Although we saw above that Manhattan as a whole is a net importer of commuters, we find that parts of the island specialize as residences, with high exports of commuters per square mile in some areas on the Upper West and Upper East side. Additionally, while Manhattan is by far the largest concentration of employment in NYC, Brooklyn and the Bronx have some census tracts with high imports of commuters per square mile, again highlighting the rich and asymmetric patterns of specialization throughout the city.

If the differences in productivity and amenities that drive this location specialization are the result of exogenous natural advantages and there are no market failures, this location specialization is efficient. However, there are a number of potential sources of market failure in the allocation of land between commercial and residential use. First, if productivity and amenity differences are the result of agglomeration forces, these agglomeration forces typically feature externalities, such that each agent does not internalize the effects of their actions on other agents (e.g., knowledge spillovers). Second, these agglomeration forces typically coexist with dispersion forces (e.g., congestion), which are again characterized by externalities (e.g., each commuter does not take into account the impact of their public transit or road use on others). Third, if these agglomeration forces are sufficiently strong, there can be multiple equilibria in the spatial distribution of economic activity, which opens up a potential role for public policy in selecting between these equilibria. Fourth, land use and zoning regulations can limit the ability of agents to take advantage of differences in productivity and amenities across locations. Sometimes these regulations are intended to correct externalities (e.g., industrial pollution of a residential neighborhood). But if these regulations generate differences in the price of floor space across alternative uses that are not rationalized by externalities, they themselves introduce misallocation, and there is again scope for changes in public policy to be welfare improving.

3 Traditional Theoretical Literature

One of key challenges in theoretically modelling economic activity within cities is the complexity of the interactions between agents in the presence of agglomeration forces. Therefore, to understand the economic mechanisms shaping the internal structure of cities, the traditional theoretical literature considered stylized representations of cities. Either economic activity was monocentric by assumption, or a restrictive geography was considered, such as identical locations along the real line or a perfectly symmetric circular city.

Alonso-Muth-Mills monocentric cities In the canonical model of internal city structure following [Alonso \(1964\)](#), [Muth \(1969\)](#), [Mills \(1967\)](#), cities are monocentric by assumption. All employment is assumed to be concentrated in a central business district (CBD) and workers face commuting costs in traveling to work. As workers living further from the city center face higher commuting costs, this must be compensated in equilibrium by a lower land rent further from the city center, in order for workers to be indifferent across locations. The geographical boundary of the city is determined by the return to land in its competing use of agricultural production. Therefore, a central prediction of these traditional theories is that land rents decline monotonically with distance from the city center, consistent with the observed property of the data that central locations on average command higher land prices than outlying areas.

Non-monocentric cities Although some historical cities are well approximated by a monocentric structure, others such as Los Angeles are better described by a polycentric structure, with multiple business districts spread throughout the metropolitan area. One form that this polycentric structure can take is an edge city, consisting of a concentration of business, shopping, and entertainment outside a traditional downtown or central business district, in what had previously been a suburban residential or rural area, and typically beside a major road.

To capture these richer patterns of land use, the assumption that all employment is concentrated in the city center can be relaxed to allow for the endogenous allocation of land between commercial and residential use throughout the city. In important contributions, [Fujita and Ogawa \(1982\)](#) consider the case of a one-dimensional city along the real line, and [Lucas and Rossi-Hansberg \(2002\)](#) analyze a perfectly symmetric circular city.

In these frameworks, whether monocentric or polycentric patterns of economic activity emerge depends on the strength of agglomeration and dispersion forces. On the one hand, a non-monocentric pattern of alternating areas of commercial and residential land use reduces commuting costs, because workers typically live closer to their place of employment than in a monocentric structure. On the other hand, these alternating areas of commercial and residential land use reduce the concentration of employment, and hence diminish agglomeration economies relative to the monocentric case in which all employment is concentrated in the CBD.

In summary, key insights from this theoretical literature are the role of the trade-off between agglomeration forces and commuting costs in generating urban rent gradients, and in determining whether these rent gradients are monocentric or polycentric.

4 Quantitative Urban Model

Although this traditional theoretical literature reveals important economic mechanisms, these stylized settings abstract from empirically-relevant differences in natural advantages across locations, which limits their usefulness for empirical work. In contrast, recent quantitative urban models are designed to connect directly to observed data on economic activity within cities, and allow for differences in both natural advantages and agglomeration forces.

4.1 Baseline Quantitative Urban Model

We begin by developing a baseline quantitative version of the canonical urban model following [Ahlfeldt et al. \(2015\)](#), before discussing a number of extensions and generalizations.⁹

Assumptions We consider a city embedded in a wider economy. The city consists of a set of discrete blocks or census tracts, which are indexed by $n = 1, \dots, N$. Each block has a supply of floor space that depends on its geographical land area and the density of development (the ratio of floor space to land area). Floor space can be used either commercially or residentially. We denote the endogenous fractions of floor space used for commercial and residential use by θ_n and $1 - \theta_n$. Blocks can be either completely specialized ($\theta_n = 0$ or $\theta_n = 1$) or incompletely specialized between these two alternative uses ($0 < \theta_n < 1$). We allow a potential tax wedge between the prices of residential and commercial floor space that can differ across blocks (ξ_i), which captures the tax equivalent of zoning regulations.

The city is populated by an endogenous measure of \bar{L} workers, who are perfectly mobile between the city and the larger economy, which provides a reservation level of utility \bar{U} . Workers decide whether or not to move to the city before observing idiosyncratic preference shocks for each possible pair of residence and workplace blocks within the city. We assume that these idiosyncratic shocks are drawn from an extreme value distribution.¹⁰ They capture all the idiosyncratic reasons why an individual worker can choose to live in one place and work in another.¹¹ If a worker decides to move to the city, she observes these realizations for idiosyncratic preferences, and picks the residence-workplace pair that yields the highest utility.

Worker utility also depends on consumption of a single final good, residential floor space use, commuting costs and residential amenities. Commuting costs increase with the travel time

⁹An accompanying Online Appendix provides a more detailed development of the model and a formal characterization of its theoretical properties. See [Redding and Rossi-Hansberg \(2017\)](#) for a review of quantitative spatial models, and [Redding \(2022b\)](#) and [Allen and Arkolakis \(2022a\)](#) for surveys of the wider literature on spatial economics.

¹⁰A long line research in economics uses this extreme value assumption following [McFadden \(1974\)](#).

¹¹For example, the worker's mother may live in one neighborhood and she wants to live near her mother, while her aunt may own a factory in another neighborhood, and this aunt may offer her a job.

between the worker's residence and workplace, and reduce her utility, either because of the unpleasantness of a long commute, or the opportunity cost of time spent commuting. Residential amenities capture characteristics of a block that make it a more or less attractive place to live and depend on both natural advantages (residential fundamentals) and agglomeration forces (residential externalities). Residential fundamentals capture exogenous characteristics that make a location more or less appealing independently of surrounding economic activity (e.g. leafy streets and scenic views). Residential externalities capture agglomeration forces that depend on the travel time weighted sum of the density of residents in surrounding locations (including positive externalities from non-traded goods and negative externalities from crime).

The final good is assumed to be costlessly traded within the city and with the wider economy, and is chosen as the numeraire ($p_n = p = 1$). Markets are assumed to be perfectly competitive. This final good is produced using inputs of labor and commercial floor space according to a constant returns to scale technology. Productivity can differ across locations within the city and depends on both natural advantages (production fundamentals) and agglomeration forces (production externalities). Production fundamentals capture exogenous characteristics that determine the productivity of a location (e.g., access to natural water). Production externalities capture agglomeration forces that depend on the travel time weighted sum of employment density in surrounding locations (e.g., knowledge spillovers).

The resulting quantitative urban model allows for rich differences in characteristics across locations in order to connect with the observed data. Blocks can differ from one another in terms of their productivity, amenities, supply of floor space, and access to the transport network, which determines bilateral travel times between locations. Productivity and amenities have exogenous components that capture differences in natural advantages and endogenous components that capture agglomeration forces. Congestion forces are governed by the elasticity of supply of floor with respect to its price, and commuting costs that increase with the travel time between a worker's residence and workplace.

Equilibrium City Structure Internal city structure is shaped by nine parameters of the model: the share of worker expenditure on the tradeable final good versus residential floor space; the share of firm costs on labor versus commercial floor space; the share of land versus capital in construction costs; the dispersion of worker idiosyncratic preferences; the elasticity of commuting costs with respect to travel times; the relative importance of residential fundamentals and externalities for amenities; the decay of residential externalities with travel time; the relative importance of production fundamentals and externalities for productivity; and the decay of production externalities with travel time.

Given these nine parameters, the spatial distribution of economic activity within the city is

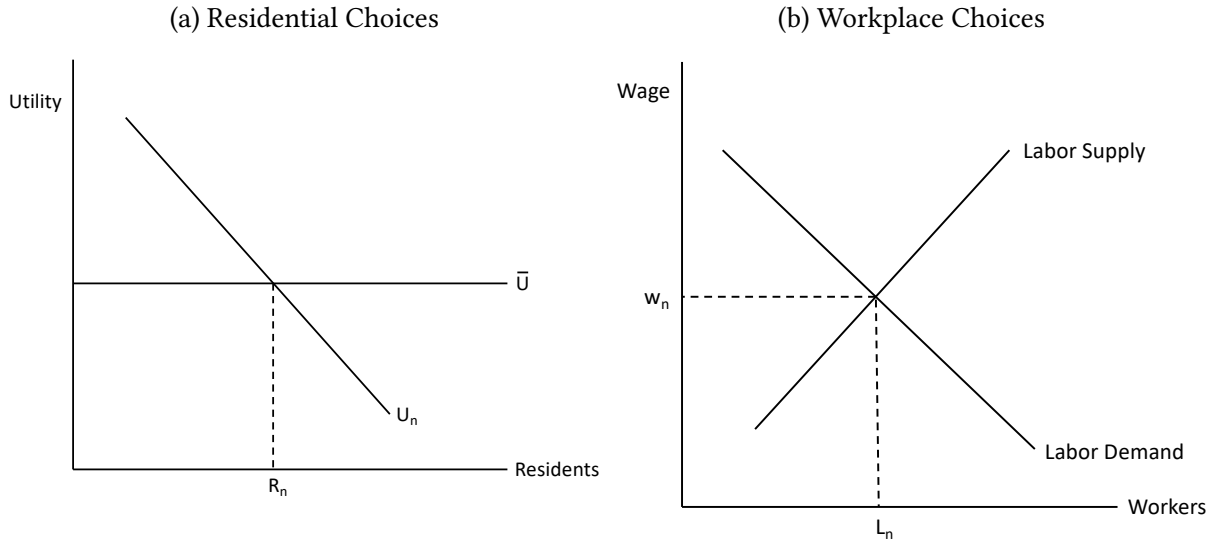
determined by six exogenous location characteristics: travel time as determined by the transport network; geographical land area; the productivity of the construction technology; the tax equivalent of land use regulations; production fundamentals; and residential fundamentals. Given these six exogenous location characteristics, the equilibrium of the model can be fully characterized using seven endogenous variables: the share of people who choose to live in each block; the share of people who choose to work in each block; the price of residential floor space; the price of commercial floor space; the wage; the share of floor space allocated to commercial versus residential use within each block; the total supply of floor space in each block; and the total number of people that choose to live in the city.

We now provide intuition for how internal city structure is determined by considering residential and workplace choices in partial equilibrium, where these two sets of decisions are linked in general equilibrium. In the left panel of Figure 4, we illustrate the determination of the number of residents (R_n), where we derive this diagrammatic representation formally in the Online Appendix. The horizontal line shows the reservation level of utility in the wider economy (\bar{U}). The downward-sloping line shows expected utility from residence n (U_n). Expected utility is decreasing in the number of residents for two reasons. First, as we increase the number of residents for a given supply of residential floor space, this bids up the price for residential floor space, and reduces expected utility. Second, as we increase the number of residents in a given location, we attract workers with lower realizations for idiosyncratic preferences for that location, which reduces expected utility through a composition or batting-average effect.

The equilibrium number of residents is determined by the intersection of the reservation level of utility in the wider economy (\bar{U}) and the expected utility from living in block n . The position of the expected utility line (U_n), and hence the equilibrium number of residents, depends on both exogenous and endogenous variables. The key exogenous variables are residential fundamentals and geographical land area. Increases in both of these exogenous variables shift the expected utility line outwards and increase the equilibrium number of residents. The key endogenous variables are commuting access to surrounding workplaces (as determined by travel times and wages), the ratio of residential floor space to geographical land area, and residential externalities. Increases in each of these endogenous variables shift the expected utility line outwards and increase the equilibrium number of residents, where these endogenous variables are determined together with workplace choices in the general equilibrium of the model.

In the right panel of Figure 4, we illustrate the determination of the number of workers in each workplace (L_n), where again we derive this diagrammatic representation formally in the Online Appendix. The downward-sloping line shows labor demand in workplace n , as determined by the equality between the wage and the value marginal product of labor. An increase in the number of workers employed in a location leads to a decrease in the wage, because of diminishing marginal

Figure 4: Residential and Workplace Choices



Note: Left panel shows the partial equilibrium determination of the number of residents in location n (R_n) by the reservation utility in the wider economy (\bar{U}) and the expected utility of living in that location (U_n). Right panel shows the partial equilibrium determination of workers in location n (L_n) by labor demand and supply as a function of the wage (w_n) in that location. See the Online Appendix for a formal derivation of this diagram.

physical productivity of labor in the production technology. The upward-sloping line shows labor supply for workplace n , as determined by worker choices of residence and workplace. In order to increase labor supply, firms must offer a higher wage, in order to attract workers with lower realizations for idiosyncratic preferences for that workplace.

The equilibrium number of workers (L_n) is determined by the intersection of labor demand and labor supply. The position of the labor demand and labor supply lines is also shaped by both exogenous and endogenous variables. Increases in the exogenous variables of production fundamentals and geographical land area shift the labor demand line outwards, as do increases in the endogenous variables of production externalities and the ratio of commercial floor space to geographical land area. An increase in the endogenous variable of commuting access to surrounding residences (as determined by travel times and wages) shifts the labor supply line outwards. Again these endogenous variables are jointly determined with residence choices in the general equilibrium of the model.

Theoretical Predictions We now discuss some key theoretical predictions of this quantitative urban model, which allow it connect directly with observed data on cities, such as the data for NYC examined above. First, the extreme value assumption for idiosyncratic preferences implies a gravity equation for bilateral commuting flows. Therefore, the probability of commuting between residence n and workplace i depends on the characteristics of the residence n (e.g., amenities and

the price of residential floor space), the attributes of the workplace i (in particular, wages) and bilateral commuting costs as determined by bilateral travel time (“bilateral resistance”). Additionally, this probability depends on the characteristics of all residences k , all workplaces ℓ and all bilateral commuting costs (“multilateral resistance”), as in the structural gravity equation in international trade. A large reduced-form empirical literature finds that the gravity equation is a strong property of observed commuting flows.¹² Given the strength of this relationship in the data, arguably any reasonable model of city structure should be consistent with this property of observed commuting flows.

Second, each workplace and residence faces an upward-sloping supply function in real income, as shown above. To obtain additional workers, a workplace must pay higher wages in order to attract workers with lower idiosyncratic preferences for that workplace. Similarly, to acquire additional residents, a residence must offer a lower cost of living net of amenities in order to attract residents with lower idiosyncratic preferences for that residence. These upward-sloping functions allow the model to accommodate the large differences in productivity and amenities across locations that are implied by the observed data, while generating realistic differences in wages across locations, and remaining well-behaved in response to counterfactual policy interventions, such as transport infrastructure improvements.

Third, the extreme value specification for idiosyncratic preferences implies that expected utility is equalized across all pairs of residences and workplaces within the city and is equal to the reservation level of utility in the wider economy. The intuition for this third result is that bilateral commutes with attractive economic characteristics (high workplace wages and low residence cost of living) attract additional commuters with lower idiosyncratic amenities, until expected utility (taking into account idiosyncratic amenities) is the same across all bilateral commutes and equal to the reservation utility. This property ensures that workers are *ex ante* indifferent between living in the city and the wider economy, although after observing their idiosyncratic preference shocks, each worker *ex post* has a preferred workplace-residence pair.

Connecting the Model and Data Given values for the model’s nine parameters, we now discuss how it can be used to rationalize the observed data on land prices, workers and residents across blocks within cities, as in the example of NYC above. Having rationalized these observed data as an equilibrium outcome, the model then can be used to undertake counterfactuals for realistic policy interventions, such as the construction of a new subway line between specific locations within the city.

First, commuter market clearing requires that the number of workers in each workplace

¹²See for example [Fortheringham and O’Kelly \(1989\)](#), [McDonald and McMillen \(2010\)](#) and [Redding \(2022a\)](#). For further discussion of structural gravity equations in international trade, see [Head and Mayer \(2014\)](#).

equals the sum across locations of the number of residents choosing to commute to that workplace, where these commuting flows satisfy the gravity equation discussed above. Given observed workers and residents in each location, and a parameterization of commuting costs as a function of bilateral travel times, the model can be solved for a unique vector of wages for which this commuter market clearing condition is satisfied. If separate data on wages and commuting flows between locations are also available, then these separate data can be used as an overidentification check on the model's predictions.

Second, using profit maximization and zero profits in the construction sector, we can solve for the price of floor space, given the price of land observed in the data and an assumed common cost of capital across all locations. Third, profit maximization and zero profits in production imply that the price of the final good must equal unit cost in each location with positive employment. The model thus yields a zero profit condition in production, which links wages, the price of floor space, productivity and the price of the final good ($p_i = p = 1$). Using wages from commuter market clearing and the price of floor space from the construction sector, we can use this zero-profit condition in production to solve for a unique value of productivity that is consistent with the observed data. Finally, given the assumption that production externalities are a constant elasticity function of the travel time weighted sum of employment density, we can further decompose productivity into production fundamentals and externalities. Production fundamentals correspond to a structural residual that enables the model to exactly rationalize the observed data. Zero employment in a location in the data is explained within the model by zero production fundamentals for that location.

Fourth, utility maximization and population mobility determine the probability that a worker chooses to live in each location within the city. This residential choice probability links wages, commuting travel times, the price of floor space, amenities and the price of the final good ($p_i = p = 1$). Therefore, using wages from commuter market clearing, the price of floor space from the construction sector and observed travel times, we can solve for a unique value of amenities that is consistent with the observed data. Finally, given the assumption that residential externalities are a constant elasticity function of the travel time weighted sum of residents density, we can further decompose amenities into residential fundamentals and externalities. Residential fundamentals again correspond to a structural residual that enables the model to exactly rationalize the observed data. Zero residents in a location in the data is explained within the model by zero residential fundamentals for that location.

Fifth, using the observed data on workers and residents, together with our solutions for wages, the price of floor space, productivity and amenities, we can solve for the demand for residential and commercial floor space in each location. Equating the demand and supply for floor space, and using observed geographical land area, we can recover the implied density of development (the

ratio of floor space to geographical land area). This density of development again corresponds to a structural residual that enables the model to exactly rationalize the observed data.

Therefore, using the equilibrium structure of the model, we can recover unobserved natural advantages in the form of production fundamentals, residential fundamentals and the density of development from the observed data on the endogenous variables of the model. We can thus quantify how much of the observed variation in land prices, workers and residents is explained by these natural advantages relative to the model's endogenous forces of commuting market access and agglomeration forces in production and residence. Undertaking this analysis requires knowledge of the model's parameters, and we discuss below how they can be estimated using credible sources of exogenous variation.

Extensions and Generalizations Although the quantitative urban model outlined above is sufficiently rich to connect directly with the observed data, and to incorporate both natural advantages and agglomeration forces, it remains stylized in several respects. We now discuss a number of extensions, which allow quantitative urban models to incorporate additional mechanisms or connect with additional observed data. The flexibility of the specification of worker commuting choices using idiosyncratic preference shocks allows a wide range of these extensions to be accommodated, while preserving the quantitative model's tractability.

First, this class of models can accommodate non-traded goods, as in [Heblich et al. \(2020\)](#) and [Fajgelbaum and Redding \(2022\)](#). Second, it can accommodate other reasons for travel apart from commuting, and capture trip chains, as in [Miyauchi et al. \(2022\)](#). Third, it can allow for multiple final goods with costly trade and technology differences, as in [Eaton and Kortum \(2002\)](#) and [Redding \(2016\)](#). Fourth, it can encompass final goods that are differentiated by origin and costly trade, as in [Armington \(1969\)](#), [Allen and Arkolakis \(2014\)](#) and [Allen et al. \(2017\)](#). Fifth, it can encapsulate horizontally-differentiated firm varieties with costly trade, as in [Helpman \(1998\)](#), [Redding and Sturm \(2008\)](#) and [Monte et al. \(2018\)](#). Sixth, it can be used to quantify the impact of zoning regulations on internal city structure, as in [Allen et al. \(2017\)](#). Seventh, it can be used as a platform for evaluating neighborhood development programs, as in the analysis of the redevelopment of Detroit in [Owens III et al. \(2020\)](#).

Eighth, it can incorporate forward-looking investments in capital accumulation, as in [Kleinman et al. \(2023\)](#). Ninth, it can allow for multiple groups of workers that are ex ante heterogeneous, and where workers from each group draw idiosyncratic preference shocks for each residence-workplace pair, as in [Redding and Sturm \(2016\)](#) and [Tsivanidis \(2018\)](#). Tenth, whereas travel time was treated as exogenous and independent of commuting flows above, congestion can be introduced, as in [Allen and Arkolakis \(2022b\)](#).¹³

¹³For a wider discussion of congestion pricing, see [Small et al. \(1989\)](#)

Which of these specifications is most useful for empirical work may depend on the application and what data are available, such as whether bilateral trade in goods between locations is observed (as in [Monte et al. 2018](#)), or whether commuting data are available for different ethnic, racial or demographic groups (as in [Tsivanidis 2018](#)). With data for such multiple groups of workers, it becomes possible to analyze the distributional consequences across these worker groups of public policy interventions, such as transport infrastructure improvements.

Mechanisms Although the quantitative urban model outlined above allows for agglomeration forces in both production and residential decisions, these agglomeration forces are assumed to be reduced-form functions of travel time weighted employment and residents density, respectively. An exciting area for further research is opening this black box and distinguishing between different underlying economic mechanisms for agglomeration.

Following [Marshall \(1920\)](#), three main sets of forces for agglomeration are traditionally distinguished, which reflect the costs of moving goods, people and ideas. First, firms may locate near suppliers or customers in order to save on transportation costs. Second, workers and firms may cluster together to pool specialization skills. Third, physical proximity may facilitate knowledge spillovers, as “the mysteries of the trade become no mystery, but are, as it were, in the air.” Another line of research dating back to [Smith \(1776\)](#) emphasizes a greater division of labor in the larger markets, as examined empirically in [Duranton and Jayet \(2011\)](#). More recently, [Duranton and Puga \(2004\)](#) distinguish between sharing, matching and learning as alternative mechanisms for the agglomeration of economic activity.

Although these mechanisms are well understood conceptually, there is relatively little evidence on their empirical importance, with a few exceptions such as [Ellison et al. \(2010\)](#). Over time, the nature of economic activity undertaken within cities has changed dramatically, from the market places and ports of pre-industrial Europe, through the centers of manufacturing of the industrial revolution, through the concentrations of office space of the mid-twentieth century, to an increasing focus on the consumption of non-traded goods and services in the twenty-first century. Using the verbs from occupational descriptions, [Michaels et al. \(2019\)](#) quantify the change in the tasks undertaken by workers in cities over time. Whereas the tasks most concentrated in cities in 1880 involved the manipulation of the physical world, such as Thread and Sew, those most concentrated in cities in 2000 involve human interaction, such as Advise and Confer.

Given these large-scale changes in the types of economic activities performed in urban areas over time, it is reasonable to think that the nature and scope of agglomeration economies could have evolved over time. Consistent with this idea, [Autor \(2019\)](#) finds substantial changes in the urban wage premium for workers with different levels of skills over time. At the beginning of the sample period in the 1970s, average wages were sharply increasing in population density for

both low-skill workers (high-school or less) and high-skill workers (some college or greater). By the end of the sample period in 2015, this wage premium to population density had increased for high-skill workers but almost disappeared for low-skill workers.

Looking ahead, the wealth of newly-available sources of Geographical Information Systems (GIS) data promising to offer new opportunities to distinguish between different mechanisms for agglomeration, including ride-hailing data (e.g. Uber and Lyft), Smartphone data with Global Positioning System (GPS) information, firm-to-firm data from sales (VAT) tax records, credit card data with consumer and firm location, barcode scanner data with consumer and firm location, public transportation commuting data (e.g. Oyster card), and satellite imaging data.

5 Applications

We now provide two illustrations of the insights from quantitative urban models from recent applications to estimate the strength of agglomeration forces (Section 5.1) and evaluate the impact of transport infrastructure improvements (Section 5.2).

5.1 Estimating Agglomeration Forces

Although there is a long literature on economic geography and urban economics, it has been empirically challenging to distinguish agglomeration and dispersion forces (second-nature geography) from variation in natural advantages (first-nature geography). Although high land prices and levels of economic activity in a group of neighboring locations are consistent with strong agglomeration forces, they are also consistent with shared amenities that make these locations attractive places to live (e.g., leafy streets and scenic views) or common natural advantages that make these locations attractive for production (e.g., access to natural water).¹⁴ To empirically disentangle these two alternative explanations for location choices, one requires a source of exogenous variation in the surrounding concentration of economic activity.

Division of Berlin [Ahlfeldt et al. \(2015\)](#) uses the division of Berlin in the aftermath of the Second World War and its reunification following the fall of the Iron Curtain as such a source of exogenous variation. The political process that ultimately led to the construction of the Berlin Wall dates back to war-time planning during the Second World War. A protocol signed in London in September 1944 designated separate occupation sectors in Berlin for the American, British and Soviet armies. The boundaries between these occupation sectors were chosen based on pre-war administrative districts that had little prior significance, such that the three sectors were of

¹⁴This is an example of the broader challenge in the social sciences of distinguishing spillovers from correlated individual effects, as discussed in [Manski \(1995\)](#).

roughly equal population, with the Americans and British in the West, and the Soviets in the East. Later a French sector was created from part of the British sector.

The original plan was for Berlin to be administered jointly by a central committee (“Kommandatura”). However, following the onset of the Cold War, East and West Germany were founded as separate states, and separate city governments emerged in East and West Berlin in 1949. From this point onwards, the adoption of Soviet-style policies of command and control in East Berlin limited economic interactions with West Berlin. For a while travel between the different sectors of Berlin remained possible, until to stop civilians leaving for West Germany, the East German authorities constructed the Berlin Wall in 1961.

Economic Mechanisms Within the quantitative urban model developed about, there are a number of different economic mechanisms through which the division of Berlin affects the spatial distribution of economic activity within West Berlin.¹⁵ First, firms in West Berlin cease to benefit from production externalities from East Berlin, which reduces productivity, land prices and employment by workplace. Second, firms in West Berlin lose access to flows of commuters from East Berlin, which increases the wage required to achieve a given employment, reducing land prices and employment by workplace. Third, residents in West Berlin no longer benefit from residential externalities from East Berlin, which reduces amenities, land prices and employment by residence. Fourth, residents in West Berlin lose access to employment opportunities in East Berlin, which decreases expected income, land prices and employment by residence.

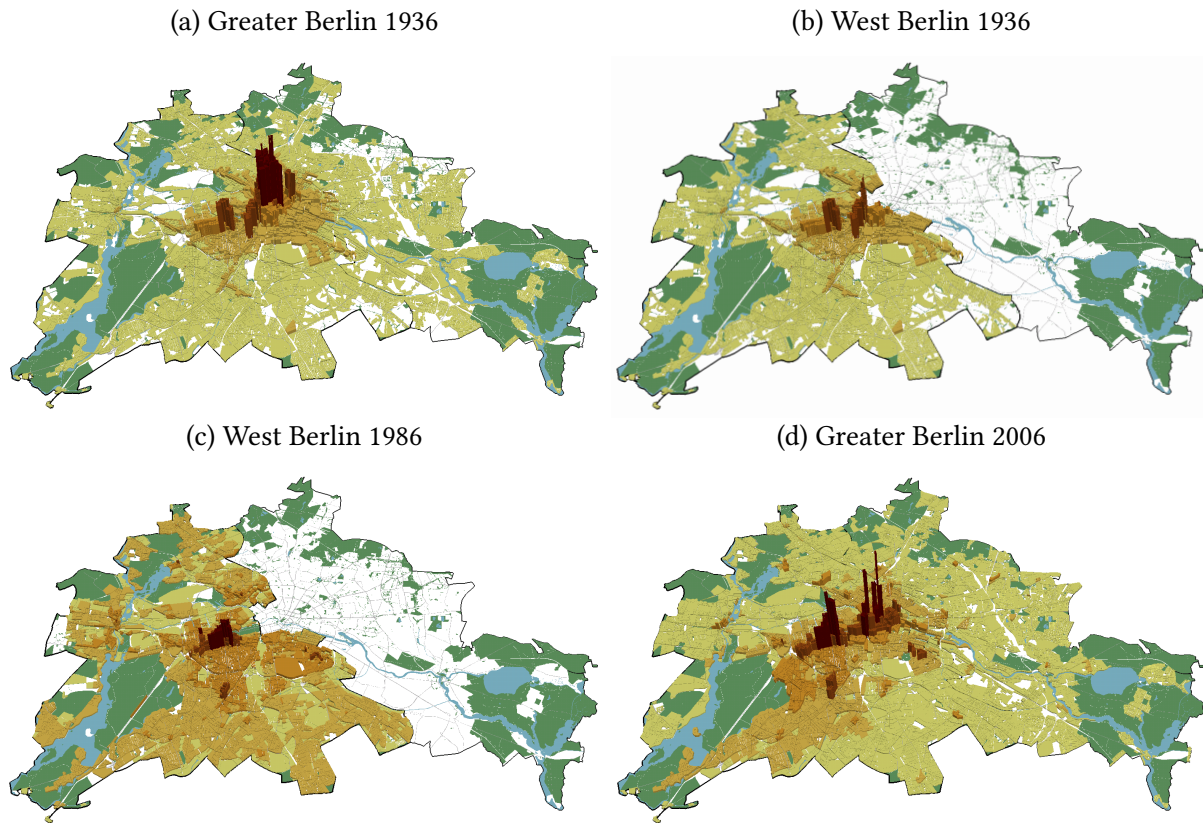
Since commuting costs rise with travel time, and production and residential externalities fall with travel time, each of these effects is greater for the parts of West Berlin closer to employment and residential concentrations in East Berlin. To restore equilibrium in the model, employment and residents reallocate within West Berlin and the wider economy of West Germany, until wages and land prices adjust, such that firms market zero profits in each location with positive production, workers are indifferent across all locations with positive residents, and there is no-arbitrage between commercial and residential land use.

Qualitative Evidence To examine the extent to which these theoretical predictions are borne out empirically, Figure 5 displays the evolution of Berlin’s land price gradient over time. In each year, land prices are normalized to have a mean of one, such that the levels of the land price surfaces are comparable over time. As shown in Panel (a), Berlin’s land price gradient in 1936 was approximately monocentric, with the highest values concentrated in the pre-war CBD of Mitte, East of the future line of the Berlin Wall. Around this central point, there are concentric

¹⁵The analysis focuses on West Berlin, because it remained a market economy, and hence one would expect the mechanisms in the model to apply. In contrast, allocations in East Berlin during the period of division were determined by central planning, which is unlikely to mimic market forces.

rings of progressively lower land prices surrounding the pre-war CBD. In Panel (b), we display land prices in 1936 for the areas of the city that were to become part of the future West Berlin, which highlights how division cut the Western part of the city off from the pre-war CBD. The two parts of the future West Berlin that had the highest land prices in 1936 were an area just West of the pre-war CBD and the future line of the Berlin Wall, surrounding the “Anhalter Bahnhof” station, and the Kudamm (“Kurfürstendamm”) in Charlottenburg and Wilmersdorf, which had developed into a fashionable shopping area in the decades leading up to the Second World War.

Figure 5: Evolution of Berlin’s Land Price Gradient over Time



Note: Land prices in Berlin over time; Land prices normalized to have a mean of one in each panel; the main public parks and forests are shown in green; the main bodies of water are shown in blue; white areas correspond to other undeveloped areas including railways. Source [Ahlfeldt et al. \(2015\)](#).

In Panel (c), we display land prices in 1986 in West Berlin. Following division, we find that the first pre-war price peak just West of the pre-war CBD is entirely eliminated, as this area ceased to be an important center of commercial and retail activity. Instead, the second pre-war price peak in the Kudamm develops into West Berlin’s CBD during the period of division. In Panel (d), we show land prices in 2006 for Berlin as a whole. We find that the area just West of the pre-war CBD and the former line of the Berlin Wall reemerges as an area of high land values, with a concentration of office and retail development. We also find a resurgence of the pre-war

CBD in Mitte in the former East Berlin as a land price peak.

The observed reorientation of the land price gradient in Figure 5 is consistent with the qualitative predictions of the model discussed above. We find the biggest decline in land prices in the parts of West Berlin closest to the pre-war CBD, which lose access to nearby concentrations of employment and residents in East Berlin. We also find little evidence of an impact on land prices along other sections of the Berlin Wall following division. This pattern of results supports the idea that it is not proximity to the Berlin Wall per se that matters, but rather the loss of access to nearby concentrations of employment and residents in East Berlin. Finally, the changes in the land price gradient shown in Figure 5 are accompanied by a similar reorientation of employment and residents within West Berlin, again consistent with the predictions of the model.

Quantitative Evidence To examine whether the quantitative urban model developed above can account quantitatively for the observed changes in the spatial distribution of land prices, employment and residents, [Ahlfeldt et al. \(2015\)](#) structurally estimate the model's parameters using the generalized method of moments (GMM). For any given set of parameters, we saw above that we can use the structure of the model to solve for the unobserved values for production fundamentals, residential fundamentals and the density of development, such that the model exactly rationalizes the observed distribution of land prices, employment and residents in the data in each year before and after division and reunification.

The model's parameters are estimated using the identifying assumption that the log changes in production and residential fundamentals in each block are uncorrelated with the change in the surrounding concentration of economic activity induced by Berlin's division and reunification. Since the city's division stemmed from military considerations during the Second World War and its reunification originated in the wider collapse of Communism, the resulting changes in the surrounding concentration of economic activity are plausibly exogenous to changes in production and residential fundamentals in each block. In particular, the log changes in production and residential fundamentals in West Berlin are assumed to be orthogonal to indicator variables for distance grid cells to the pre-war CBD. This identifying assumption requires that the systematic change in the gradient of economic activity in West Berlin relative to the pre-war CBD following division is explained by the mechanisms of the model (the changes in commuting access and production and residential externalities) rather than by systematic changes in the structural residuals of production and residential fundamentals.

The parameters are estimated for both division and reunification separately, and pooling both experiments together. All three specifications yield a similar pattern of estimated coefficients, with evidence of substantial agglomeration forces for both production and residential decisions. In the specification pooling both sources of variation, the estimated elasticity of productivity

with respect to travel time weighted employment density is 0.07, while the estimated elasticity of amenities with respect to travel time weighted residents density is 0.15. Both production and residential externalities are highly localized, with exponential rates of decay with travel time of 0.36 and 0.76, respectively. These estimates imply that production and residential externalities fall to close to zero after around 10 minutes of travel time, which corresponds to around 0.83 kilometers by foot (at an average speed of 5 kilometers per hour) and about 4 kilometers by underground and suburban railway (at an average speed of 25 kilometers per hour).

Other Evidence These parameter estimates are broadly consistent with the findings of other empirical research. The estimate of the elasticity of productivity with respect to production externalities of 0.07 is towards the high end of the 3-8 percent range stated in the survey by [Rosenthal and Strange \(2004\)](#), but less than the elasticities from some quasi-experimental studies (e.g., [Greenstone et al. 2010](#) and [Kline and Moretti 2014](#)). This elasticity of 0.07 also captures the effect on productivity of doubling employment density holding constant travel times. In reality, a doubling in total city population is typically achieved by a combination of an increase in the density of employment and an expansion in geographical land area, with the accompanying increase in average travel times within the city. Therefore, the elasticity of productivity with respect to such a doubling of total city population is less than 0.07, because the increase in average travel times reduces production externalities according to the estimated exponential rate of decay.

The finding of highly localized production externalities is also consistent with other research using within-city data. Using data on the location of advertising agencies in Manhattan, [Arzaghi and Henderson \(2008\)](#) find little evidence of knowledge spillovers beyond 500 meters straight-line distance. In comparison, a straight-line distance of 450-550 meters in Berlin corresponds to around 9 minutes of travel time, after which production externalities are estimated to have declined to around 4 percent. Finally, the finding of substantial residential externalities is in line with recent empirical findings that urban amenities are endogenous to the surrounding concentration of economic activity, as in [Glaeser et al. \(2001\)](#), [Diamond \(2016\)](#) and [Almagro and Domínguez-Iino \(2019\)](#). The finding that these residential externalities are highly localized is also consistent with other evidence. Using data on an urban revitalization program in Richmond, Virginia, [Rossi-Hansberg et al. \(2010\)](#) finds that housing externalities fall by approximately one half every 1,000 feet.

Counterfactuals Since the quantitative urban model developed above connects directly with the observed data on land prices, workers and residents, it can be used to undertake counterfactuals, both for the estimated parameters and for alternative parameter values. Starting from the observed data before either division or reunification, one can undertake a counterfactual for each

of these changes in the surrounding concentration of economic activity. In the special case of the model with no agglomeration forces, productivity and amenities are determined solely by exogenous natural advantages, and the only impact of division and reunification is through changes in commuting market access. A key finding from these counterfactuals is that this special case of the model with no agglomeration forces generates predicted impacts of division and reunification that are much smaller than in the data. In contrast, for the estimated agglomeration parameters, the model generates predicted impacts of division and reunification close to those in the data. Therefore, these findings provide evidence that substantial agglomeration forces are required to explain the observed concentration of economic activity within cities.

Taken together, the empirical evidence in this section suggest that the quantitative urban model developed above can not only rationalize the observed spatial variation in land prices, employment and residents in the data, but for the estimated parameter values is quantitatively successful in predicting the change in the internal organization of economic activity within cities in response to large-scale shocks such as Berlin's division and reunification.

5.2 Transport Infrastructure Improvements

We now turn to a second application of quantitative urban models to examine the impact of transport infrastructure improvements. The dense concentrations of economic activity observed in modern metropolitan areas involve transporting millions of people each day between their home and place of work. For example, the London Underground today handles around 3.5 million passenger journeys a day, and its trains travel around 3.5 million passenger journeys a day, and its trains travel around 76 million kilometers each year, about 200 times the distance between the Earth and the Moon. Therefore, a key public policy question is the economic impact of the transport network, and its role in sustaining these dense concentrations of economic activity.

Nineteenth-Century Transport Revolution To provide evidence on the effect of transport infrastructure improvements on the spatial distribution of economic activity, [Heblich et al. \(2020\)](#) use the mid-nineteenth-century invention of steam railways as a natural experiment. The key idea behind this 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, steam railways dramatically reduced travel time for a given distance, thereby permitting the first large-scale separation of workplace and residence.

Greater London provides an attractive empirical setting for this analysis, because of the availability of spatially-disaggregated data on economic activity over a long time horizon from 1801-1921, before and after this transport innovation. Data are available for a number of different

geographical definitions of London. First, there is Greater London, as defined by the boundaries of the modern Greater London Authority (GLA), which includes a 1921 population of 7.39 million and an area of 1,595 square kilometers. Second, there is the historical County of London, which has a 1921 population of 4.48 million and an area of 314 square kilometers. Third, there is the City of London, which has a 1921 population of 13,709 and an area of around 3 square kilometers, and whose boundaries correspond approximately to the Roman city wall.

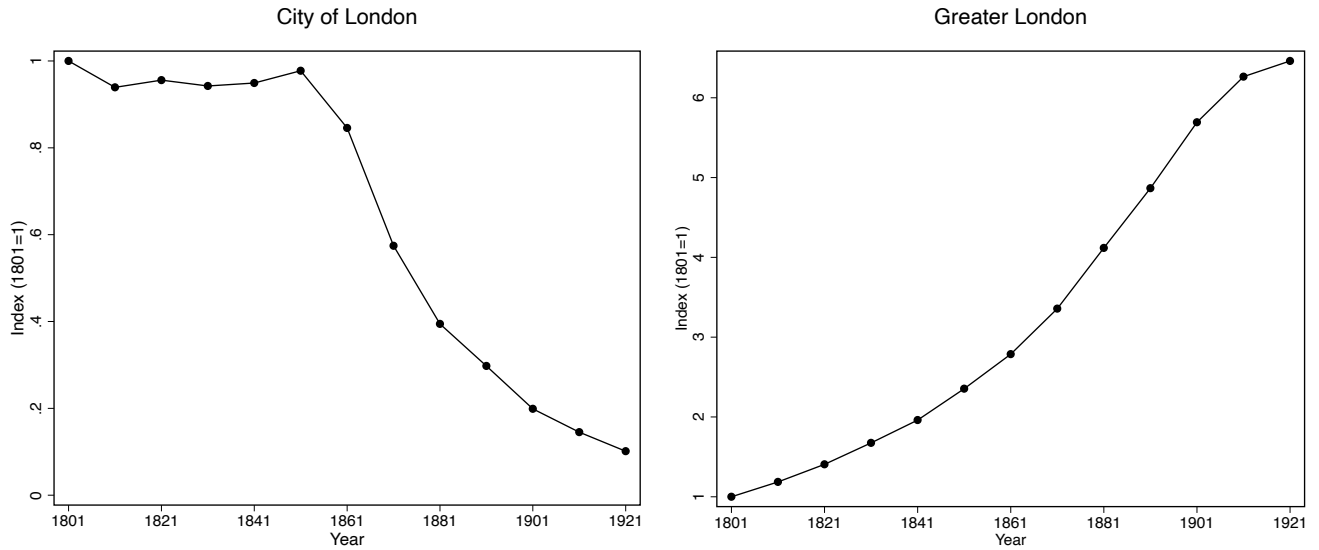
At the beginning of the nineteenth century, the most common mode of transport was walking, with average travel speeds in good road conditions of around 3mph. With the growth of urban populations, attempts to improve existing modes of transport led to the introduction of the horse omnibus from Paris to London in the 1820s. However, the limitations of horse power and road conditions ensured that average travel speeds remained low, at around 6 mph. Against this backdrop, the invention of the steam locomotive in 1825 was a major transport innovation, with the London and Greenwich railway opening in 1836 as the first steam railway to be built specifically for passengers. This innovation transformed the relationship between travel time and distance, with average travel speeds of around 21 mph.

Economic Activity in Greater London Following the invention of the steam passenger railway, there is a large-scale change in the organization of economic activity within Greater London. In Figure 6, 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 nineteenth century, population in the City of London was relatively constant (at around 130,000), while population in Greater London grew substantially (from 1.14 to 2.69 million). From 1851 onwards shortly after the first steam passenger railways, there is a sharp drop in population in the City of London, which falls by around 90 percent to 13,709 in 1921. In contrast, the population of Greater London as a whole continues to grow rapidly from 2.69 million in 1851 to 7.39 million in 1921.

Figure 7 provides some first evidence on what is driving for these starkly different patterns of population growth in Central London and the metropolitan area as a whole. The solid black line with circle markers again shows residential population (but as thousands of people), which is sometimes referred to a night population, because it is measured based on the location where a person sleeps on census night. The gray line with triangle markers shows day population, as measured by the City of London day censuses for 1866, 1881, 1891 and 1911, and employment by workplace for 1921 from the population census.¹⁶ In the decades following the first steam passenger railways, the sharp decline in night population is combined with a steep rise in day

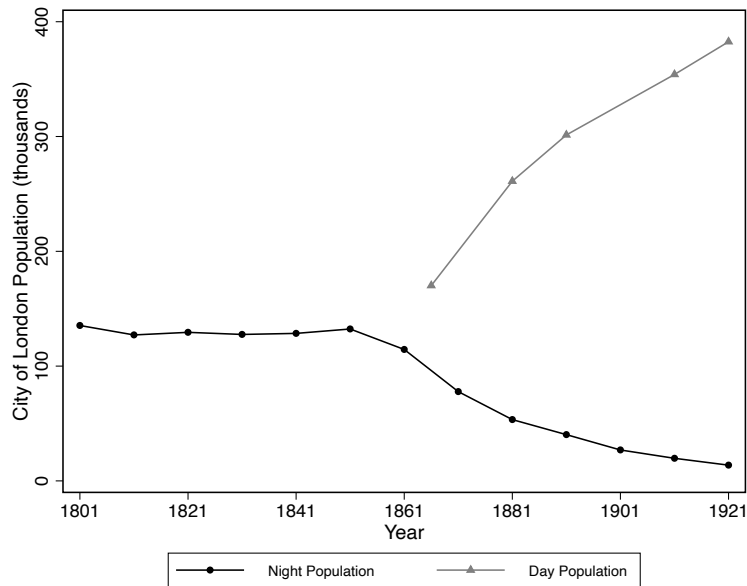
¹⁶Day population is defined as “every person, male or female, of all ages, residing, engaged, occupied, or employed in each and every house, warehouse, shop, manufactory, workshop, counting house, office, chambers, stable, wharf, etc ... during the working hours of the day, whether they sleep or do not sleep there (Salmon 1891, page 97).”

Figure 6: Population Indexes Over Time (City of London and Greater London, 1801 equals 1)



Note: Indexes of residence (night) population from the population censuses of England and Wales. Greater London corresponds to the modern-day Greater London Authority (GLA). City of London is the historical commercial center of London, whose boundaries correspond approximately to the Roman city wall. Source: [Heblich et al. \(2020\)](#).

Figure 7: Night and Day Population in the City of London



Note: “Night Population” is residential population from the population census; “Day Population” is day population from the City of London Day Censuses for 1866, 1881, 1891 and 1911 and workplace employment from the population census for 1921. City of London is the historical commercial center of London, whose boundaries correspond approximately to the Roman city wall. Source: [Heblich et al. \(2020\)](#).

population, consistent with a change in the City of London’s pattern of specialization as a workplace rather than as a residence.

Quantitative Evidence To rationalize these empirical findings, [Heblich et al. \(2020\)](#) develop an estimation procedure that holds in an entire class of quantitative urban models, including the framework developed in Section 4 above. This estimation procedure uses a combined land and commuter market-clearing condition, which uses the gravity equation for commuting flows and the equality between the income of owners of floorspace and payments for the use of residential and commercial floor space. Although bilateral commuting data are only observed for 1921 at the end of the sample period, this framework can be used to estimate the construction of the railway network going back to the early-nineteenth century.

In a first step, these bilateral commuting data for 1921 are used to estimate the parameters that determine commuting costs as a function of travel times using the transport network. In a second step, these parameter estimates are combined with historical data on population, land values and the evolution of the transport network back the beginning of the nineteenth century. Conditioning on the historical data on population by residence and land values, the model's combined land and commuter market-clearing condition is used to generate predictions for historical employment by workplace and bilateral commuting flows. An advantage of this approach is that the historical data on population by residence and land values control for a range of other potential determinants of economic activity, such as changes in productivity, amenities, the costs of trading goods, the floor space supply elasticity, and expected utility in the wider economy. Within the structure of the model, population by residence and land values are sufficient statistics for these other unobserved determinants of economic activity, thereby isolating the impact of the change in commuting costs from the new transport technology.

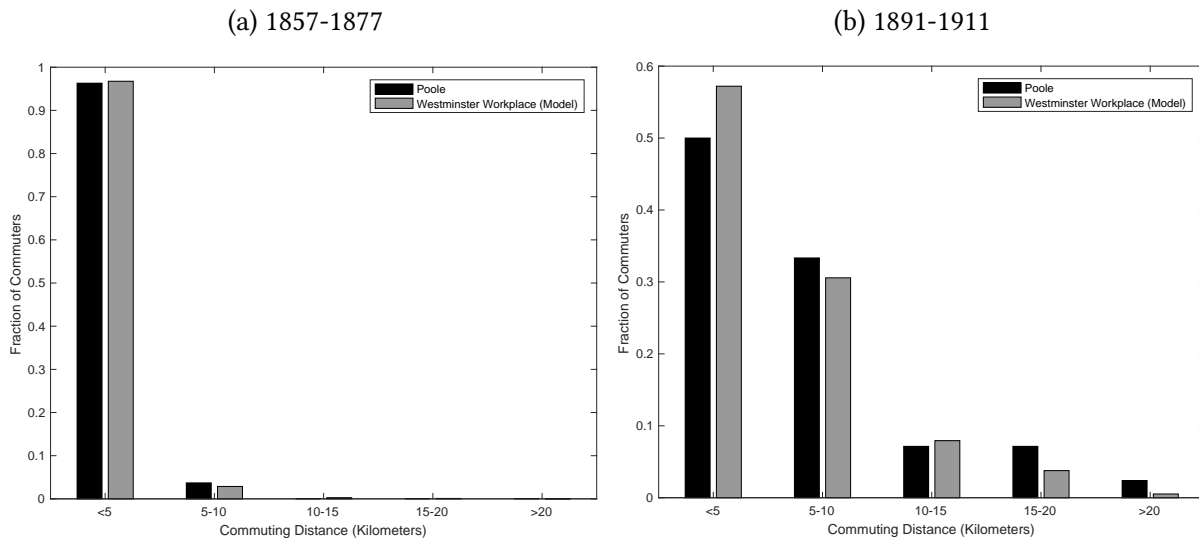
Although the impact of the railway network on commuting costs is estimated using 1921 information alone, the model successfully captures the observed sharp divergence between nighttime and daytime population in the City of London from the mid-nineteenth century onwards. As the improvement in transport technology reduces commuting costs, workers become able to separate their residence and workplace to take advantage of high wages in locations with high productivity relative to amenities (so that these locations specialize as workplaces) and the lower cost of living in locations with high amenities relative to productivity (so that these locations specialize as residences). Therefore, the finding that Central London specializes as a workplace is consistent with it having high productivity relative to amenities compared to the suburbs of Greater London, 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 workers and residents, 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.

Although the City of London experiences by far the largest absolute increase in employment,

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 of log employment density respect to distance from the Guildhall in the center of the City of London declines between 1831 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 about 48 percent in 1921. This pattern of results is in with a long line of empirical research that finds evidence of employment (and population) decentralization in response to transport improvements, as reviewed in [Redding and Turner \(2015\)](#).

As a specification check, the model’s predictions for commuting flows are compared to historical data on commuting distances from the personnel ledgers of Henry Poole Tailors, a high-end bespoke tailoring firm, which was founded in 1802.¹⁷ Figure 8 compares the model’s predictions for commuting into the workplace of Westminster, in which this company is located, with the commuting distances observed in the data based on workers’ residential addresses at the time they joined Henry Poole. The left panel compares the model’s predictions for 1861 with the commuting distances of workers who joined Henry Poole between 1857 and 1877. The right panel compares the model’s predictions for 1901 with commuting distances of workers who joined Henry Poole between 1891 and 1911.

Figure 8: Commuting Distances in the Model and 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. Source: [Heblich et al. \(2020\)](#).

In making this comparison, there are a number of possible sources of discrepancies between the model’s predictions and the data, including the fact that this company is located in a specific

¹⁷For further discussion of the Henry Poole data, see [Green \(1988\)](#).

site in Westminster, where as the model covers all of that borough. Nevertheless, the model is remarkably successful in capturing the change in the distribution of commuting distances between these time periods. In the opening decades of the railway age, most workers in Westminster in both the model and data lived within 5km of their workplace. In contrast, by the turn of the twentieth century, commuting distances up to 20km are observed 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.

Counterfactuals An advantage of the empirical methodology developed above, which conditions on observed historical data on population and land values, is that these observed variables control for changes in other potential determinants of economic activity, such as productivity and amenities. However, in evaluating the economic rationale for transport infrastructure improvements, a key counterfactual question of interest is how the spatial distribution of economic activity would have evolved if the only thing that changed were the transport network, holding all else constant. As discussed above, since quantitative urban models connect directly with the observed data, they are well suited for such counterfactual analysis. Table 1 reports the results of undertaking counterfactuals for the removal of the railway network, starting at the observed equilibrium in the data in 1921 at the end of the sample period. These counterfactuals are undertaken under a range of alternative assumptions about the floor space supply elasticity (the elasticity of the supply of floor space with respect to changes in its price) and the strength of agglomeration and dispersion forces (the elasticity of productivity and amenities with respect to changes in the density of workers and residents, respectively).

Each counterfactual removes the entire overground and underground railway network, holding constant the rest of the transport network (including omnibus and tram routes) at its 1921 structure. Each specification holds expected utility and total population in the wider economy of Great Britain constant, and allows the share of the economy's workers that choose residence-workplace pairs in Greater London to adjust, until expected utility in Greater London is equal to its unchanged value in the wider economy.

Column (1) holds constant the supply of floor space, productivity and amenities in each location in Greater London ("Inelastic No Agglom"). In this specification, removing the 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, which compares with the observed 1831 value of 1.92 million. Column (2) assumes a calibrated floor space supply elasticity of 1.83, but holds productivity and amenities in each location in Greater London constant. Allowing for an endogenous response in the supply of floor space to the new transport technology magnifies its effect on the spatial distribution of economic activity, with the removal of the railway network decreasing the total

population of Greater London by 22.30 percent to 5.74 million in 1831. This pattern of results highlights a complementarity between the development of the built environment and changes in land use and improvements in transport infrastructure.

Table 1: Counterfactuals for Removing the Entire Overground and Underground Railway Network, Starting from the Equilibrium in the 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	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0.172$
<i>Economic Impact</i>				
Population	6,376,698	5,742,937	5,462,344	3,587,947
Change in Rateable Value (RV)	−£8.24m	−£15.55m	−£20.78m	−£35.07m
NPV Change in RV (3 percent)	−£274.55m	−£518.26m	−£692.76m	−£1,169.05m
NPV Change in RV (5 percent)	−£164.73m	−£310.96m	−£415.66m	−£701.43m
<i>Construction Costs</i>				
Cut-and-Cover Underground		−£9.96m		
Bored-tube Underground		−£22.90m		
Overground Railway		−£33.19m		
Total All Railways		−£66.05m		
<i>Ratio Economic Impact / Construction Cost</i>				
NPV Change in RV (3 percent) Construction Cost	4.16	7.85	10.49	17.70
NPV Change in RV (5 percent) Construction Cost	2.49	4.71	6.29	10.62

Note: Counterfactuals start in 1921 and remove the entire overground and underground railway network; omnibus and tram network held 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 the 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 the estimated production agglomeration force; $\eta^R = 0.172$ corresponds to the estimated residential agglomeration force; all specifications assume population mobility between Greater London and the wider economy, with the elasticity of population supply regulated by the commuting elasticity $\epsilon = 5.25$; population is the counterfactual value after removing the railway network; actual population for Greater London in 1831 and 1921 is 1,915,824 and 7,390,721, respectively; rateable value (RV) is the rental value of land and buildings; 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. Source: [Heblich et al. \(2020\)](#).

Column (3) assumes the same calibrated floor space supply elasticity of 1.83 and an estimated production agglomeration elasticity of 0.086, but holds constant amenities in each location in Greater London. Introducing production agglomeration forces further enhances the impact of the new transport technology, with the removal of the railway network reducing the total population of Greater London by 26.09 percent to 5.46 million in 1831. Column (4) assumes the same calibrated floor space supply elasticity of 1.83 and estimated production and residential agglomeration elasticities of 0.086 and 0.172, respectively. Incorporating both production and residential agglomeration forces again magnifies the counterfactual changes in population, with the removal

of the railway network reducing the total population of Greater London by 51.45 percent to 3.59 million in 1831, compared again with the observed 1831 value of 1.92 million.

Economic Impact and Construction Cost These counterfactual predictions can be used to evaluate the welfare effects of the construction of the railway network. Under the 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 is unaffected by the construction of the railway network. Therefore, as in the classical approach to valuing public goods following [George \(1879\)](#), the welfare gains from the new transport technology are experienced by landlords through changes in the rental value of land and buildings (measured by the rateable value in the data). The magnitude of these welfare gains can be assessed by comparing the counterfactual changes in the net present value of land and buildings from the removal of the railway network with its construction costs. Construction costs can be measured using historical estimates of the authorized capital per mile for the private-sector companies that built these railway lines, which yields estimates of £555,000 per mile for bored-tube underground railways, £355,000 per mile for cut-and-cover underground railways, and £60,000 per mile for overground railways (all in 1921 prices).

As shown in [Table 1](#), the ratio of changes in the net present value of land and buildings to construction costs is substantially greater than one, regardless of whether a 3 or 5 percent discount rate is used. This pattern of results suggests that the large-scale investments in the 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), allowing for a positive floor space supply elasticity substantially increases the economic impact of the railway network, again highlighting the role of complementary responses in the built environment and land use. Comparing Columns (2), (3) and (4), incorporating agglomeration forces further magnifies the economic effects of the railway network, illustrating the relevance of taking into account agglomeration forces in cost-benefit evaluations of transport infrastructure investments.

Alternative Transport Infrastructure Investments Often policy makers are also interested in comparing many alternative possible investments, such as the decision of which of many links in a railway or highway network to improve. To develop a framework to address this question, [Allen and Arkolakis \(2022b\)](#) embed a specification of endogenous route choice in a quantitative spatial model. Individuals' experience idiosyncratic shocks to travel costs for each route and choose the least-cost route taking into account these idiosyncratic shocks. A key implication of this framework is that the welfare effects of a small improvement in a transport link is equal to the percentage cost saving multiplied by the initial value of travel along that link. Although this

result is derived for particular function forms, this implication is closely related to the celebrated result of [Hulten \(1978\)](#) that a sufficient statistic for the welfare effect of a small productivity shock in an efficient economy can be summarized by the appropriate Domar weight, as used for an analysis of China’s High Speed Rail Network in [Barwick et al. \(2020\)](#).¹⁸

More generally, [Fajgelbaum and Schaal \(2020\)](#) develop a framework for characterizing optimal transport networks in spatial equilibrium. This characterization is challenging, because the problem is high dimensional and can be non-convex. The paper shows that the problem of finding the optimal transport network can be transformed into the problem of finding the optimal flow in a network, which has been studied in the operations research literature. The planner chooses the optimal amount to invest in each link in the transport link, where the trade costs for each link are assuming to be increasing in the volume of traffic, such that the problems remains convex. While so far this approach has been applied to trade in goods between cities, incorporating commuting within cities is an exciting avenue for further research.

6 Conclusions

Cities are arguably one of the ultimate achievements of human civilization. Modern metropolitan areas have economies as large as entire countries, with for example the New York–Newark–Jersey City metropolitan area estimated to have a gross domestic product (GDP) comparable to Canada.¹⁹ These cities feature complex internal structures, with a rich specialization by residential and commercial land use, and an intricate division of labor.

The traditional theoretical literature on cities focused on stylized settings, such as a one-dimensional line or a perfectly symmetric circular city, in order to highlight key economic mechanisms. In the canonical Alonso-Muth-Mills model, cities have a monocentric structure, with all employment concentrated in a central business district (CBD), and a land price gradient that declines monotonically in distance from the CBD. However, real-world cities are not well-approximated by a one-dimensional line or a perfectly symmetric circle, which has limited the usefulness of these stylized settings for empirical work.

The real-world cities in which people live can be monocentric or polycentric, and can exhibit dramatic changes in land prices and land use, both across neighborhoods and across blocks within neighborhoods. A key breakthrough in recent research has been the development of quantitative urban models that are able to rationalize these observed features of the data. These frameworks can accommodate many locations that differ in productivity, amenities, land area, the supply of floor space and transport connections. Nevertheless, these models remain tractable and amenable

¹⁸Relating to our NYC example in Section 2 above, see [Gupta et al. \(2022b\)](#) for evidence on the quantitative impact of the latest expansion to New York’s subway network, the Second Avenue Subway.

¹⁹See <https://www.aei.org/carpe-diem/understanding-americas-enormous-20-6t-economy-by-comparing-us-metro-area-gdps-to-entire-countries/>.

to theoretical analysis with a small number of parameters to be estimated.

One key insight from these quantitative urban models is that the observed concentration of economic activity within cities cannot be explained by natural advantages alone, but instead requires substantial agglomeration forces. Using the division of Berlin by the Berlin Wall as a natural experiment, a model with exogenous productivity and amenities generates estimated impacts substantially smaller than those observed in the data. In contrast, a model with the estimated value of agglomeration forces in production and residential decisions successfully accounts for the observed reorientation of economic activity.

A second key insight is the role of advanced transport networks in sustaining dense concentrations of economic activity in modern metropolitan areas. Before the steam passenger railway and the automobile, the slow travel times achievable by human or horse power meant that most people lived close to where they worked. In contrast, the dramatic reduction in commuting costs from these advances in transport technology enabled a large-scale separation of workplace and residence. Using the invention of the steam passenger railway in the mid-nineteenth century as a natural experiment, quantitative urban models are able to rationalize the sharp divergence between workers and residents observed in Central London in the second half of the nineteenth century, and the observed increase in commuting distances.

Looking ahead there remain many exciting areas for further research. In the aftermath of the Covid-19 pandemic, will fears about disease transmission and preferences for social distancing have a lasting impact on city structure? Or will changes in social norms and advances in remote-working technology be more influential? What will be the consequences of these developments for public transit systems? What will be effects of further innovations in transport technology, such as ride hailing and sharing and autonomous vehicles?

Over the centuries, cities have changed drastically from market places, to the locus of manufacturing industry, to clusters of office and retail development, and to centers of consumption. Looking ahead, perhaps the only thing that is certain is that they will continue to change. Nevertheless, as long as there are benefits to eliminating physical space, through reductions in the costs of moving people, goods and ideas, cities are likely to thrive and prosper.

References

- AHLFELDT, G., S. REDDING, D. STURM, AND N. WOLF (2015): “The Economics of Density: Evidence from the Berlin Wall,” *Econometrica*, 83, 2127–2189.
- AHLFELDT, G. M. AND J. BARR (2020): “The Economics of Skyscrapers: A Synthesis,” *CEPR Discussion Paper*, 14987.
- ALLEN, T. AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129, 1085–1140.
- (2022a): “Supply and Demand in Space,” *NBER Working Paper*, 30598.
- (2022b): “The Welfare Effects of Transportation Infrastructure Improvements,” *Review of Economic Studies*, forthcoming.
- ALLEN, T., C. ARKOLAKIS, AND X. LI (2017): “Optimal City Structure,” Yale University, mimeograph.
- ALMAGRO, M. AND T. DOMÍNGUEZ-IINO (2019): “Location Sorting and Endogenous Amenities: Evidence from Amsterdam,” New York University, mimeograph.
- ALONSO, W. (1964): *Location and Land Use*, Cambridge MA: Harvard.
- ARMINGTON, P. S. (1969): “A Theory of Demand for Products Distinguished by Place of Production,” *IMF Staff Papers*, 16, 159–178.
- ARZAGHI, M. AND J. V. HENDERSON (2008): “Networking Off Madison Avenue,” *Review of Economic Studies*, 75, 1011–1038.
- AUTOR, D. (2019): “Work of the Past, Work of the Future,” *American Economic Review*, 109, 1–32, richard Ely Lecture.
- BARR, JASON, F. H. S. AND S. J. KULKARNI (2018): “What’s Manhattan Worth? A Land Values Index from 1950 to 2014,” *Regional Science and Urban Economics*, 70, 1–19.
- BARR, J. M. (2016): *Building the Skyline: The Birth and Growth of Manhattan’s Skyscrapers*, Oxford: Oxford University Press.
- BARRERO, J., N. BLOOM, AND S. DAVIS (2021): “Why Working From Home Will Stick,” *NBER Working Paper*, 28731.

- BARWICK, P. J., D. DONALDSON, S. LI, AND Y. LIN (2020): “The Welfare Effects of Passenger Transportation Infrastructure: Evidence from China,” MIT, mimeograph.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106, 479–524.
- DURANTON, G. AND H. JAYET (2011): “Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities,” *Journal of Urban Economics*, 69, 56–71.
- 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, Amsterdam: Elsevier, vol. 4, 2063–2117.
- EATON, J. AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70, 1741–1779.
- ELLISON, G., E. L. GLAESER, AND W. R. KERR (2010): “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns,” *American Economic Review*, 100, 1195–1213.
- FAJGELBAUM, P. AND S. REDDING (2022): “Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914,” *Journal of Political Economy*, 130, 1249–1318.
- FAJGELBAUM, P. D. AND E. SCHAAL (2020): “Optimal Transport Networks in Spatial Equilibrium,” *Econometrica*, 88, 1411–1452.
- FORTHERINGHAM, S. AND M. O’KELLY (1989): *Spatial Interaction Models: Formulations and Applications*, Dordrecht: Kluwer.
- FUJITA, M. AND H. OGAWA (1982): “Multiple Equilibria and Structural Transformation of Non-monocentric Urban Configurations,” *Regional Science and Urban Economics*, 12, 161–196.
- GEORGE, H. (1879): *Progress and Poverty*, New York: Doubleday, Page & Company.
- GLAESER, E. AND D. CUTLER (2021): *Survival of the City: Living and Thriving in an Age of Isolation*, New York: Penguin Press.
- GLAESER, E. L., J. KOLKO, AND A. SAIZ (2001): “Consumer City,” *Journal of Economic Geography*, 1, 27–50.
- GREEN, D. R. (1988): “Distance to Work in Victorian London: A Case Study of Henry Poole, Bespoke Tailors,” *Business History*, 30, 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, 536–598.
- GUPTA, A., V. MITTAL, AND S. VAN NIEUWERBURGH (2022a): “Work from Home and the Office Real Estate Apocalypse,” *NBER Working Paper*, 30526.
- GUPTA, A., S. VAN NIEUWERBURGH, AND C. KONTOKOSTA (2022b): “Take the Q Train: Value Capture of Public Infrastructure Projects,” *Journal of Urban Economics*, 129, 103422.
- HAUGHWOUT, ANDREW, J. O. AND D. BEDOLL (2008): “The Price of Land in the New York Metropolitan Area,” *Current Issues in Economics and Finance, Federal Reserve Bank of New York*, 14, 1–7.
- HEAD, K. AND T. MAYER (2014): “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in *Handbook of International Economics*, ed. by E. Helpman, G. Gopinath, and K. Rogoff, Amsterdam: Elsevier, North Holland, vol. 4, 131–196.
- HEBLICH, S., S. REDDING, AND D. STURM (2020): “The Making of the Modern Metropolis: Evidence from London,” *Quarterly Journal of Economics*, 135, 2059–2133.
- 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, Cambridge: Cambridge University Press, 33–54.
- HULTEN, C. R. (1978): “Growth Accounting with Intermediate Inputs,” *Review of Economic Studies*, 511–518.
- KLEINMAN, B., E. LIU, AND S. REDDING (2023): “Dynamic Spatial General Equilibrium,” *Econometrica*, forthcoming.
- 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, 275–331.
- LUCAS, R. E. AND E. ROSSI-HANSBERG (2002): “On the Internal Structure of Cities,” *Econometrica*, 70, 1445–1476.
- MANSKI, C. F. (1995): *Identification Problems in the Social Sciences*, Cambridge MA: Harvard University Press.
- MARSHALL, A. (1920): *Principles of Economics*, London: Macmillan.

- McDONALD, J. AND D. McMILLEN (2010): *Urban Economics and Real Estate: Theory and Policy*, Hoboken, NJ: John Wiley & Sons.
- McFADDEN, D. (1974): “The Measurement of Urban Travel Demand,” *Journal of Public Economics*, 3, 303–328.
- MICHAELS, G., F. RAUCH, AND S. REDDING (2019): “Task Specialization in U.S. Cities from 1880–2000,” *Journal of the European Economic Association*, 17, 754–798.
- MILLS, E. S. (1967): “An Aggregative Model of Resource Allocation in a Metropolitan Centre,” *American Economic Review*, 57, 197–210.
- MIYAUCHI, Y., K. NAKAJIMA, AND S. REDDING (2022): “The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data,” *NBER Working Paper*, 28497.
- MONTE, F., S. REDDING, AND E. ROSSI-HANSBERG (2018): “Commuting, Migration and Local Employment Elasticities,” *American Economic Review*, 108, 3855–3890.
- MUTH, R. (1969): *Cities and Housing*, Chicago: University of Chicago Press.
- OWENS III, R., E. ROSSI-HANSBERG, AND P.-D. SARTE (2020): “Rethinking Detroit,” *American Economic Journal*, 12, 258–305.
- RAMANI, A. AND N. BLOOM (2021): “The Donut Effect of Covid-19 on Cities,” *NBER Working Paper*, 28876.
- REDDING, S. J. (2016): “Goods Trade, Factor Mobility and Welfare,” *Journal of International Economics*, 101, 148–167.
- (2022a): “Suburbanization in the United States 1970–2000,” *Economica*, 89, S110–S136, 100th Anniversary Special Issue.
- (2022b): “Trade and Geography,” in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Amsterdam: Elsevier, chap. 3, 147–217.
- 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, 1766–1797.
- (2016): “Neighborhood Effects: Evidence from the Streets of London,” Princeton University, mimeograph.

- REDDING, S. J. AND M. A. TURNER (2015): “Transportation Costs and the Spatial Organization of Economic Activity,” in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. Strange, Amsterdam: Elsevier, chap. 20, 1339–1398.
- 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, Amsterdam: Elsevier, vol. 4.
- ROSSI-HANSBERG, E., P.-D. SARTE, AND R. I. OWENS (2010): “Housing Externalities,” *Journal of Political Economy*, 118, 485–535.
- SALMON, J. (1891): *Ten Years’ Growth of the City of London*, London: Simkin, Marshall, Hamilton Kent & Company.
- SMALL, K. A., C. WINSTON, AND C. A. EVANS (1989): *Road Work: A New Highway Pricing and Investment Policy*, Washington DC: Brookings Institution.
- SMITH, A. (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations*, Dublin: Messrs. Whitestone, Chamberlaine, Potts and Watson.
- TSIVANIDIS, N. (2018): “The Aggregate And Distributional Effects Of Urban Transit Infrastructure: Evidence From Bogotá’s TransMilenio,” University of California, Berkeley, mimeograph.
- UNITED NATIONS (2018): *World Urbanization Prospects: The 2018 Revision*, New York: United Nations, Department of Economic and Social Affairs.