NBER WORKING PAPER SERIES

QUANTITATIVE URBAN ECONOMICS

Stephen J. Redding

Working Paper 33130
http://www.nber.org/papers/w33130

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

I am grateful to Princeton University for research support. This paper was commissioned for the Handbook of Regional and Urban Economics. An accompanying Online Appendix contains the technical derivations of all results. A toolkit with the Matlab code for the baseline quantitative urban model in Section 4 is available from https://www.quantitativeurbanmodels.com/. I am grateful to Arthur Adam and Zhuokai Huang for excellent research assistance. I would like to thank Gabriel Ahlfeldt, Don Davis, Jonathan Dingel, Dave Donaldson, Gilles Duranton, Stephan Heblich, Ferdinando Monte, Esteban Rossi-Hansberg, Daniel Sturm and Felix Tintelnot for helpful comments and discussion. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Quantitative Urban Economics
Stephen J. Redding
NBER Working Paper No. 33130
November 2024
JEL No. R32, R41, R52

ABSTRACT

This paper reviews recent quantitative urban models. These models are sufficiently rich to capture observed features of the data, such as many asymmetric locations and a rich geography of the transport network. Yet these models remain sufficiently tractable as to permit an analytical characterization of their theoretical properties. With only a small number of structural parameters (elasticities) to be estimated, they lend themselves to transparent identification. As they rationalize the observed spatial distribution of economic activity within cities, they can be used to undertake counterfactuals for the impact of empirically-realistic public-policy interventions on this observed distribution. Empirical applications include estimating the strength of agglomeration economies and evaluating the impact of transport infrastructure improvements (e.g., railroads, roads, Rapid Bus Transit Systems), zoning and land use regulations, place-based policies, and new technologies such as remote working.

Stephen J. Redding
Department of Economics &
School of Public and International Affairs
Princeton University
Princeton, NJ 08544
and CEPR
and also NBER
reddings@princeton.edu

1 Introduction

Modern metropolitan areas include vast concentrations of economic activity, with Greater London and the New York metropolitan area encompassing around 8.0 and 18.9 million people in 2010, respectively. Within these metropolitan areas, we observe large differences in land prices and rich patterns of land use. Some locations are used for workplaces, while others are used for residences. Among residential locations, some are prosperous (such as the Upper East Side in New York), while others are poor (such as the South Bronx in New York). Underlying these rich patterns of economic activity are large-scale investments in building structures and transport networks. What are the economic forces that explain these concentrations of economic activity and rationalize these large-scale investments? How large are the economic gains from land use specialization? How are these concentrations of economic activity likely to evolve in the future in the face of new transportation, information and communication technologies, such as remote working, online retail, autonomous vehicles and artificial intelligence?

This chapter reviews recent quantitative urban models, which are concerned with the spatial organization of economic activity within cities, including patterns of land use by workplace and residence, and patterns of spatial sorting by heterogeneous agents. These models are designed to connect directly with spatially-disaggregated data for individual cities, including for example data on residents, employment and land values on thousands of city blocks. Through modelling the observed organization of economic activity within cities, they provide platforms for undertaking counterfactuals for realistic public-policy interventions. Perhaps the most common empirical application is to transport infrastructure improvements (e.g., railroads, roads, public transit), but they also provide frameworks to evaluate zoning and land use regulations, other place-based policy interventions, and the implications of new communication technologies.

A useful distinction can be drawn between quantitative urban and regional models. While quantitative urban models are concerned with internal city structure (the network of economic interactions within a single city), quantitative regional models are instead concerned with systems of cities or regions (the network of economic interactions between cities or regions).[1] The main distinction between these two classes of models is that their different spatial scales change the relative importance of different economic mechanisms. Between cities, goods trade and migration are typically the dominant mechanisms of interaction. Within cities, commuting (the separation of residence and workplace) and consumption travel (the separation of residence and

[1]For a survey of quantitative regional models, see Allen and Arkolakis (2024) and Allen and Arkolakis (2022a). For reviews of quantitative spatial models more broadly, see Redding and Rossi-Hansberg (2017) and Redding (2022b). Within the traditional regional and urban economics literatures, the contrast is between the Alonso-Muth-Mills model of internal city structure and the Rosen-Roback model of a system of cities. For an informal discussion of quantitative urban models, see Redding (2023).

consumption) become relatively more important. The economic issues of most relevance also differs at these two different spatial scales. At finer spatial scales within cities, many of the classic issues of urban economics rise to the fore, such as segregation, crime, neighborhood effects, gentrification, zoning, housing supply restrictions, and the funding of public schools.

Before introducing quantitative urban models, it is useful to distinguish between two broad classes of explanations for the spatial distribution of economic activity: first-nature and second-nature geography. First-nature geography corresponds to exogenous natural advantages or locational fundamentals, such as access to natural water, or a waterfront location in a 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 can reflect a self-sustaining process of agglomeration, in which the location decisions of agents mutually reinforce one another.[2]

Developing quantitative frameworks that incorporate both first and second-nature geography is challenging, because of the complex feedbacks introduced into agents' location decisions by agglomeration forces. Yet which of these two explanations drives the observed concentration of economic activity is central to a host 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 can be 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, such that even small public policy interventions can have substantial effects, by shifting the location of economic activity between multiple equilibria.[3]

Alongside these agglomeration forces, there operate a variety of dispersion forces. As eco-

[2]Sometimes the term "second-nature" is reserved for historical man-made factors, such as canals and railroads (e.g., Lin and Rauch 2022). We follow Krugman (1993) in using "second-nature" to refer to all effects of the location of economic agents relative to one another (whether current or historical). For further discussions of first and second-nature geography, see the surveys in Redding (2010, 2011).

[3]For reviews of the theoretical and empirical literatures on agglomeration, see Duranton and Puga (2004), Rosenthal and Strange (2004), Combes and Gobillon (2015), and Duranton and Puga (2020).

nomic 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.[4]

To isolate the role of second-nature geography, the traditional theoretical literature on cities typically abstracted from first-nature geography by assuming a featureless plain, and focused on stylized geographies, such as a one-dimensional line or a symmetric circle. 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 structure, 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 one-dimensional line or a symmetric circle. In these more general specifications, non-monocentric patterns of economic activity can emerge, with alternating patterns of residential and commercial land use.

Although these stylized settings reveal important mechanisms and capture some systematic tendencies in the data, real world cities are not well approximated by a one-dimensional line or a 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

[4]For a recent discussion of the implications of the spread of disease and social distancing on the future development of cities, see Glaeser and Cutler (2021).

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 theoretical analysis of their properties, such as the conditions under which there is a unique equilibrium versus multiple equilibria. Additionally, these models have a small number of structural parameters (elasticities) 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 distribution of economic activity, they can be used to undertake counterfactuals for impact of realistic public policy interventions on this observed distribution, including for example the construction of a new subway line connecting a particular set of locations within a city.

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, zoning and land use regulations, 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 this paper is structured as follows. Section 2 provides some empirical evidence on the organization of economic activity within cities to motivate the development of quantitative urban models, using New York City (NYC) as an example. Section 3 summarizes the traditional theoretical literature on the internal structure of cities. Section 4 introduces a baseline quantitative urban model following Ahlfeldt et al. (2015). Section 5 discusses some key properties of quantitative urban models that facilitate their empirical application to real-world data on cities. Section 6 discusses a range of extensions, including endogenous floor space supply, richer specifications of goods trade, dynamics, spatial sorting by heterogeneous agents, and incorporating a separation between residence and locations of consumption. Section 7 discusses some empirical applications of quantitative urban models, focusing in particular on the estimation of the strength of agglomeration forces and the impact of transport improvements. Section 8 discusses areas for further research. Section 9 summarizes our conclusions.

2 Motivating Evidence

We begin by providing some motivating evidence on the organization of economic activity within cities. We use the five boroughs of New York City (NYC) as an example (Bronx, Brooklyn, Manhattan, Queens and Staten Island), which are part of the New York - Newark - Jersey City metropolitan area. Together these five boroughs have a combined area of 469 square miles and a 2010 population of 8.2 million, compared with an area of 3,450 square miles and 2010 population of 18.9 million for the wider metropolitan area.[5]

2.1 Land Prices

Land prices are determined by competition between alternative uses of land (given the constraints of land use regulation) and hence provide summary information on the relative attractiveness of locations. 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 concepts 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.[6] 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.[7] As apparent from the figure, Midtown Manhattan has the highest land prices, with a smaller secondary peak in downtown Manhattan, and lower land values in census tracts in between these two peaks. Towards the bottom of Manhattan, the Lower East Side has lower land prices than other nearby areas. Across the East River, there is another smaller peak in land prices in central Brooklyn. Finally, the areas in the immediate vicinity of Central Park are more

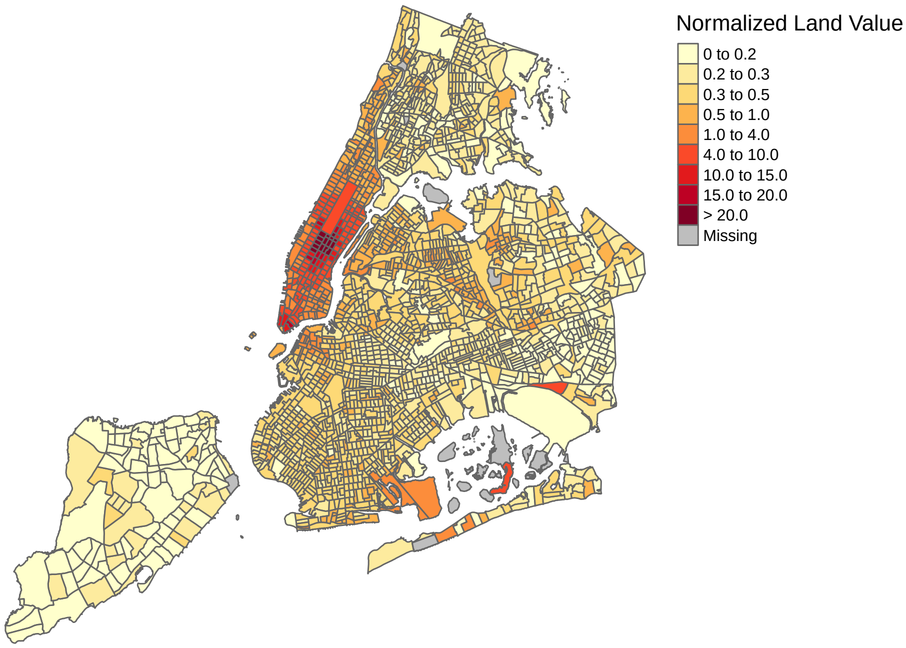
[5]While we focus on NYC as a motivating example, similar rich patterns of internal specialization hold for other urban areas, including for example Berlin (Ahlfeldt et al. 2015), Chicago (McMillen 1996), Detroit (Owens III et al. 2020), London (Heblich et al. 2020), Los Angeles (Severen 2023), and Bogota (Tsivanidis 2024), among others.

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

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

expensive than those located further away.

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



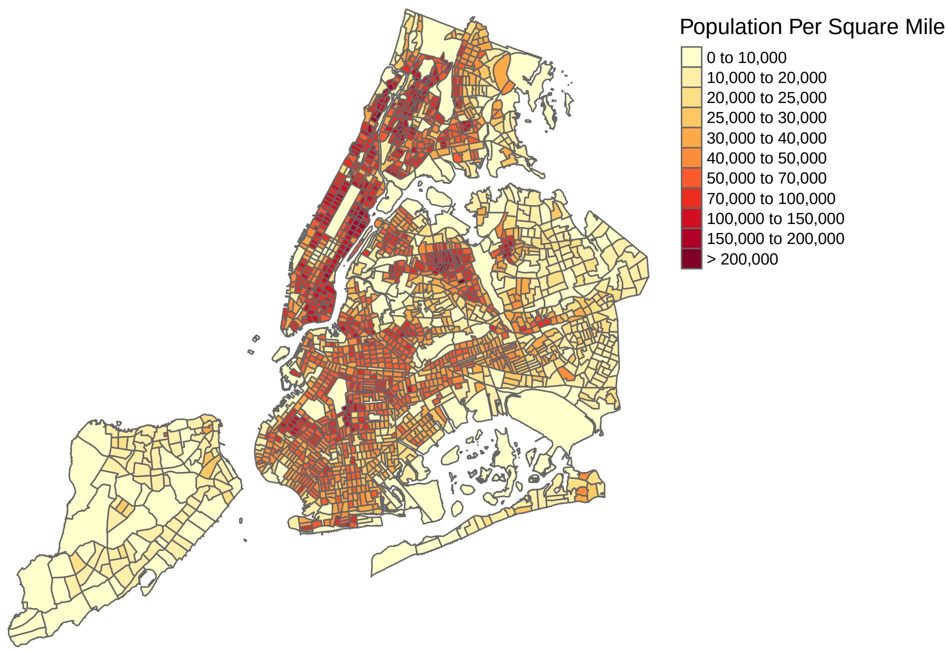
Note: Assessed land values per square mile in 2011, normalized by the mean across census tracts in New York City (the five boroughs of Bronx, Brooklyn, Manhattan, Queens and Staten Island). Therefore a value above one corresponds to an above average land value per square mile. Land values are from property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

The higher land prices observed in some locations relative to others can be explained by a combination of variation in the demand for commercial and residential land use and land use regulations. These demands for commercial and residential 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.

2.2 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 (59.1 square kilometers), with a population density of 66,579 people per square mile. In contrast, Staten Island is comparatively sparsely populated, with a population density of 7,923 people per square mile.[8] Even within Manhattan, population density fluctuates substantially, with relatively low densities in the commercial areas of midtown and downtown, and relatively high densities in the residential areas 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 productivity or amenities in a location attracts people, increasing population density, and bidding up the price of floor space. Public policy can play an important role in influencing this response in the price of floor space. 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

[8]Nevertheless, Staten Island is densely-populated compared 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.

boundaries and allowing people to commute longer distances from further away.

2.3 Workers and Residents

Commuting allows workers to separate their residence and workplace, which enables them 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. 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.[9]

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

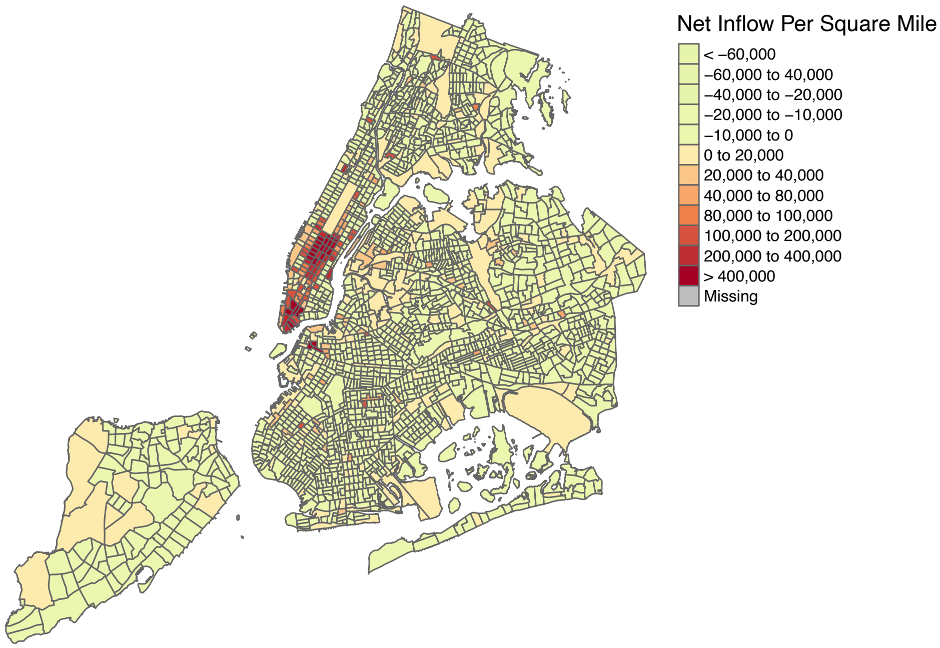
Figure 3 provides further evidence on this location specialization by showing net imports of commuters (workers minus residents) per square mile in 2010 for each census tract in NYC. Workers corresponds to employment by workplace, which equals the sum of all in-commuting flows to a census tract (including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

Location specialization is even more extreme at this micro scale of individual census tracts. The two land price peaks of midtown and downtown Manhattan are highly-specialized commercial areas, with net imports of commuters of more than 400,000 per square mile.[10] These values are substantially larger than the highest population densities visible in Figure 2, which reflects the

[9]For further discussion of working from home and its implications for cities, see Barrero et al. (2021), Ramani and Bloom (2021), Gupta et al. (2022a), Delventhal and Parkhomenko (2023) and Monte et al. (2023).

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

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. Source: Data on commuting flows from the LEHD Origin-Destination Employment Statistics (LODES).

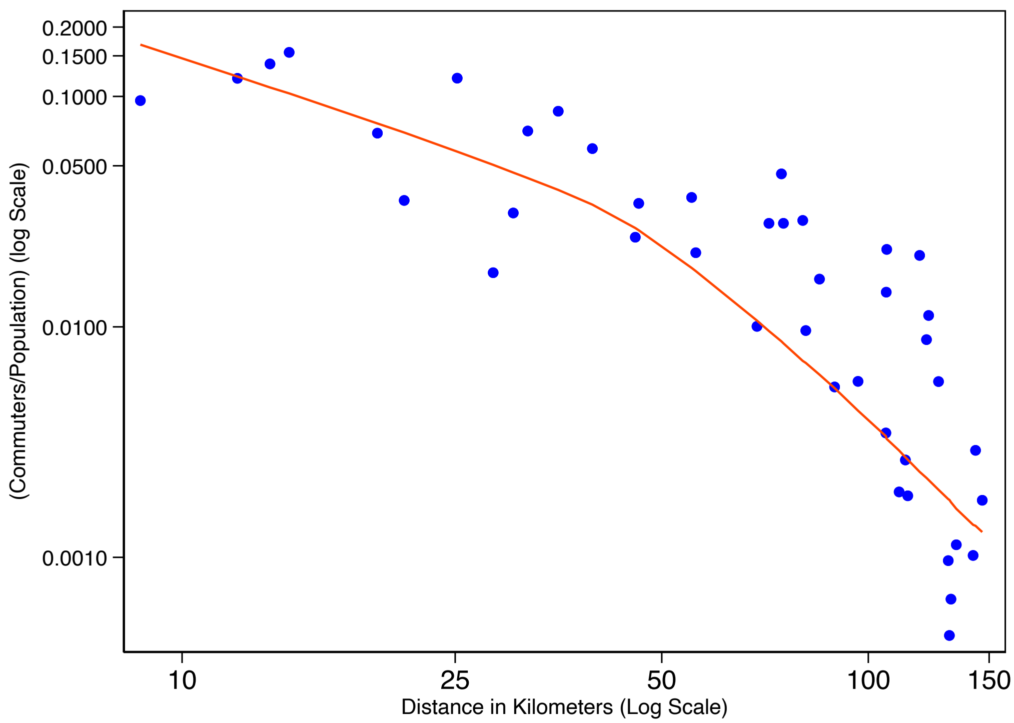
fact that employment by workplace is much more spatially concentrated than population. Many of the commuters working in these commercial districts of Manhattan travel long distances from outside NYC. Although we saw above that Manhattan as a whole is a net importer of commuters, we find that parts of the island specialize as residences, with high exports of commuters per square mile in some areas on the Upper West and Upper East sides. 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.

2.4 Gravity in Commuting

Finally, we show that the bilateral commuting flows underlying this specialization of locations by residence and workplace are well described by a so-called gravity equation. According to this relationship, bilateral commuting flows are increasing in the population size of the residence, increasing in the employment size of the workplace, and decreasing in bilateral travel frictions (as measured by for example geographical distance).

In Figure 4, we illustrate this relationship using data on commuting flows to Manhattan from the 2000 population census (the last population census to report bilateral commuting data between counties). We express commuting flows to Manhattan as a share of the residential population, where residential population includes all residents regardless of their employment status. We show results for all counties of residence in the states of New York, New Jersey, Connecticut and Pennsylvania (parts of which make up the New York metropolitan area), whose centroids lie within 150 kilometers of the centroid of Manhattan, excluding Manhattan itself from these counties of residence. We show these commuting flows to Manhattan as a share of the residential population against bilateral distance (in kilometers) using a log-log scale, where bilateral distance corresponds to the ellipsoid distance between county centroids. Each solid blue circle represents a county of residence and the red line shows the locally-weighted linear least squares relationship between the two variables.

Figure 4: Bilateral Commuting Flows to Manhattan in 2000



Note: Blue circles show counties of residence in the states of New York, New Jersey, Connecticut and Pennsylvania (parts of which make up the New York metropolitan area), whose centroids lie within 150 kilometers of the centroid of Manhattan, excluding Manhattan itself as a county of residence. Vertical axis shows commuting flows to Manhattan (NY) as a share of residential population (log scale), where residential population includes all residents regardless of their employment status. Horizontal axis shows distance to Manhattan (NY) in kilometers (log scale) from the centroid of each residence county to the centroid of Manhattan. Red line shows fitted values from locally-weighted linear least squares regression. Source: US population census.

Since this figure considers a single workplace county (Manhattan), it holds the size of the workplace constant for all counties of residence. By scaling commuting flows by residential population, this figure also controls for the size of the residence. After controlling for workplace and residence size, we find a tight and approximately log linear relationship between bilateral commuting flows and geographic distance. A natural caveat is that geographical distance is an imperfect proxy for bilateral commuting frictions, because the costs of commuting a given distance can be lowered by investing in transport infrastructure. Consistent with this, we find a somewhat concave relationship, with a lower elasticity of commuting flows to distance at shorter distances. Nevertheless, the closeness of the relationship between bilateral commuting flows and distance is striking with a R-squared in the ordinary least squares regression relationship between the two variables of 0.69. The strength of this gravity equation relationship in the data provides motivation for developing a quantitative urban model consistent with this relationship.

3 Traditional Theoretical Literature

One of the 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.

3.1 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.[11] The model considers a city on the real line. There is a single final good that can be costlessly traded. All employment is assumed to be concentrated in a central business district (CBD) and workers face commuting costs that are increasing in the distance travelled 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 this traditional theoretical literature is that land rents decline monotonically with distance from the city center, consistent with the observed property of data on cities that central locations on average command higher land prices than outlying areas.

[11] For reviews of this traditional theoretical literature on the Alonso-Muth-Mills model, see for example Brueckner (1987), Duranton and Puga (2015), and Glaeser (2008).

3.2 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.

To capture these richer patterns of land use, the assumption that all employment is concentrated in the city center can be relaxed. 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.[12] By construction, since space is symmetric, there are no differences in first-nature geography across locations, and city structure is explained solely by second-nature geography. Again there is a single final good that can be costlessly traded. But the location of both employment and residents within the city is now endogenously determined.

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, two key insights from this traditional 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. The main limitation of the Alonso-Muth-Mills is its partial equilibrium nature, with its exogenous assumption that all employment is concentrated in the CBD. Although the non-monocentric models discussed above relax this assumption, the absence of idiosyncratic worker preferences in these models implies that commuting patterns depend on a complex set of inequalities that compare equilibrium wages net of commuting costs in each possible direction of travel. The abstraction from differences in first-nature geography in both sets of models also makes it challenging to directly connect these theoretical frameworks to observed data on cities.

4 Baseline Quantitative Urban Model

We next outline a baseline quantitative urban model following Ahlfeldt et al. (2015) that develops an empirically-tractable version of this canonical theoretical model of internal city structure.

[12] For further discussion of the theoretical literature on non-monocentric cities, see Fujita and Krugman (1995), Fujita et al. (1999) and Fujita and Thisse (2002). For an analysis of optimal urban land use and zoning, see Rossi-Hansberg (2004).

Locations can differ in terms of productivity, amenities, the density of development (the ratio of floor space to land area), and access to transport infrastructure. Both productivity and amenities are influenced by natural advantages and agglomeration forces. Congestion forces take the form of an inelastic supply of land and commuting costs that are increasing in travel time, where travel time depends on the transport network.

We consider a city that is embedded in a wider economy. The city consists of a discrete set of locations (city blocks) that are indexed by $n, i \in \mathbb{N}$. Time is discrete and is indexed by t. There are two types of agents: workers and landlords. Each worker is endowed with one unit of labor that is supplied inelastically. Workers have idiosyncratic preferences for each pair of residence and workplace and are geographically mobile within the city. Landlords are geographically immobile and own local floor space.

We consider two different assumptions about worker mobility with the wider economy: (i) A "closed-city" specification, with an exogenous measure of workers ($L_{\mathbb{N}t} = L_{\mathbb{N}}$), in which worker utility is endogenous; (ii) An "open-city" specification, in which the measure of workers ($L_{\mathbb{N}t}$) is endogenously determined by population mobility with a wider economy that provides a reservation level of utility $\overline{U}_t$. In this open-city specification, workers choose whether to move to the city before observing their idiosyncratic preferences. In both specifications, workers choose their preferred pair of residence and workplace within the city after observing their realizations for idiosyncratic preferences for each pair of residence and workplace. In the baseline version of the model, we assume a continuous measure of workers $L_{\mathbb{N}t}$, which ensures that the realized values of variables equals their expected values, and abstracts from any issues of granularity or small-sample variation.

Firms produce a single final good under conditions of perfect competition and constant returns to scale. This final good is costlessly traded and chosen as the numeraire ($P_{nt} = 1$). The baseline version of the model is static, but productivity, amenities, the supply of floor space and the transport network are allowed to evolve over time. We discuss extensions and generalization of this baseline specification, including dynamics, in Section 6 below.

4.1 Workplace-Residence Choices

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

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

where we suppress the time subscript from now onwards, except where important; P_n is the price of the final consumption good; Q_n is the price of residential floor space; w_i is the wage; κ_{ni}

is an iceberg commuting cost; we model this iceberg commuting cost as depending on bilateral travel time (τ_{ni}) using the transport network: $\kappa_{ni} = e^{\kappa\tau_{ni}} \in [1, \infty)$, where $\kappa > 0$ parameterizes commuting costs;[13] B_n captures residential amenities that can be endogenous to the surrounding concentration of economic activity through agglomeration forces; and $b_{ni}(\omega)$ is an idiosyncratic preference draw that captures all the idiosyncratic factors that can cause an individual to live and work in particular locations within the city.[14]

Residential amenities (B_n) are assumed to depend on residential fundamentals ($\overline{B}_n$) and residential externalities ($\mathbb{B}_n$). Residential fundamentals capture features of physical geography that make a location a more of less attractive place to live independently of the surrounding concentration of economic activity (e.g., green areas). Residential externalities capture the interactions between residents within the city (e.g., positive externalities through local public goods and negative externalities through crime):

$$B_n = \overline{B}_n \mathbb{B}_n^{\eta^B}, \qquad \mathbb{B}_n \equiv \sum_{i \in \mathbb{N}} e^{-\delta^B \tau_{ni}} R_i, \tag{2}$$

where R_i is the measure of residents in location i; η^B governs the magnitude of these residential externalities; and δ^B controls their spatial decay with travel time.

Idiosyncratic preferences ($b_{ni}(\omega)$) are drawn from an independent extreme value (Fréchet) distribution for each residence-workplace pair and for each worker:

$$G(b) = e^{-b^{-\epsilon}}, \qquad \epsilon > 1, \tag{3}$$

where we normalize the Fréchet scale parameter in equation (3) to one, because it enters the worker choice probabilities isomorphically to residential amenities (B_n) in equation (1); the smaller the Fréchet shape parameter ϵ, the greater the heterogeneity in idiosyncratic preferences, and the less sensitive are worker location decisions to economic variables.[15]

These idiosyncratic preference shocks make solving the model's commuter market clearing condition tractable by ensuring that each residence-workplace pair faces an upward-sloping supply function for commuters in terms of wages adjusted for amenities, commuting costs and the

[13] Although commuting costs (κ_{ni}) are modelled in terms of utility, they enter the indirect utility function (1) multiplicatively with the wage, which implies that they are proportional to the opportunity cost of time. Therefore, similar results hold if commuting costs are instead modeled as a reduction in effective units of labor.

[14] A closely-related formulation assumes instead that workers have heterogeneous productivity for each pair of residence and workplace. Both specifications yield similar predictions for commuting probabilities, but imply different interpretations for observed wages. In the heterogeneous productivity specification, observed wages equal the wage per effective unit of labor multiplied by the average number of effective units of labor.

[15] Modeling idiosyncratic preferences using the extreme value distribution has a long tradition in transportation economics, dating back to McFadden (1974). Here idiosyncratic preferences are specific to a residence-workplace pair, which allows for sorting by residence and workplace. But it is straightforward to consider a generalized extreme value specification with a nesting structure, in which for example workers choose a residence and then a workplace, with potentially different dispersion parameters for each decision (e.g., Baum-Snow and Han 2024).

cost of living $\left(B_n w_i / \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)\right)$. Using the properties of the extreme value distribution (3), the probability that a worker chooses to reside in n and work in i is:

$$\lambda_{ni} = \frac{L_{ni}}{L_{\mathbb{N}}} = \frac{\left(B_n w_i\right)^{\epsilon} \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_k w_{\ell}\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}}, \tag{4}$$

where L_{ni} is the measure of commuters from n to i and $L_{\mathbb{N}}$ is the measure of workers in the city, as shown in Online Appendix B.8.

A key implication of equation (4) is that bilateral commuting flows satisfy a gravity equation, consistent with the empirical evidence presented in Section 2.4 above. As in the structural gravity equation literature in international trade, bilateral commuting flows depend not only on "bilateral resistance" (κ_{ni}) between a pair of locations n and i in the numerator, but also on "multilateral resistance" ($\kappa_{k\ell}$ for all $k, \ell \in \mathbb{N}$) in the denominator. Although individual workers have idiosyncratic preferences for each residence-workplace pair, because there is a continuous measure of these workers, there is no uncertainty in the supply of commuters for any residence-workplace pair.

Summing across workplaces in equation (4), we obtain the probability of residing in each location ($\lambda_n^R = \sum_{\ell \in \mathbb{N}} \lambda_{n\ell}$):

$$\lambda_n^R = \frac{R_n}{L_{\mathbb{N}}} = \frac{\left(B_n\right)^{\epsilon} \Phi_n^R \left(P_n^{\alpha} Q_n^{1-\alpha}\right)^{-\epsilon}}{\sum_{k \in \mathbb{N}} \left(B_k\right)^{\epsilon} \Phi_k^R \left(P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}} \qquad \Phi_n^R \equiv \sum_{\ell \in \mathbb{N}} \left(w_{\ell} / \kappa_{n\ell}\right)^{\epsilon}, \tag{5}$$

where recall that R_n is the measure of residents in location n; and we have defined Φ_n^R as a measure of residents' commuting market access; which depends on commuting costs and the wages offered in each workplace.

Summing across residences in equation (4), we obtain the probability of being employed in each location ($\lambda_i^L = \sum_{k \in \mathbb{N}} \lambda_{ki}$):

$$\lambda_i^L = \frac{L_n}{L_{\mathbb{N}}} = \frac{\left(w_i\right)^{\epsilon} \Phi_i^L}{\sum_{\ell \in \mathbb{N}} \left(w_{\ell}\right)^{\epsilon} \Phi_{\ell}^L}, \qquad \Phi_i^L \equiv \sum_{k \in \mathbb{N}} B_k^{\epsilon} \left(\kappa_{ki} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}, \tag{6}$$

where L_i is the measure of workers employed in location i; and we have defined Φ_i^L as a measure of workplace commuting market access, which depends on the cost of living adjusted for amenities and commuting costs in each residence.

Another implication of the extreme value distribution for idiosyncratic preferences (3) is that expected utility conditional on choosing a residence-workplace pair is the same across all residence and workplace pairs within the city:

$$U = \mathbb{E}\left[u\right] = \vartheta \left[\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_k w_{\ell}\right)^{\epsilon} \left(\kappa_{k\ell} P_k^{\alpha} Q_k^{1-\alpha}\right)^{-\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{7}$$

where the expectation is taken over the distribution for idiosyncratic preferences; $\vartheta \equiv \Gamma((\epsilon - 1)/\epsilon)$; $\Gamma(\cdot)$ is the Gamma function; and we report the derivation in Online Appendix B.8.

Therefore, each individual worker has a preferred residence-workplace pair based on their realizations for idiosyncratic preferences. But expected utility conditional on choosing a given residence-workplace pair is the same across all residence-workplace pairs. The intuition is as follows. On the one hand, higher amenity-adjusted real income $\left(B_i w_i / \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)\right)$ for residence n and workplace i raises the utility of a worker with a given realization for idiosyncratic preferences $b_{ni}(\omega)$, and hence increases expected utility. On the other hand, higher amenity-adjusted real income $\left(B_i w_i / \left(\kappa_{ni} P_n^{\alpha} Q_n^{1-\alpha}\right)\right)$ attracts workers with lower realizations for idiosyncratic preferences $b_{ni}(\omega)$, which reduces expected utility. With a Fréchet distribution of idiosyncratic preferences, these two effects exactly offset one another. Residence-workplace pairs with high amenity-adjusted real income attract more commuters on the extensive margin until expected utility conditional on choosing a given residence-workplace pair is the same across all residence-workplace pairs.

4.2 Production

The single final good is produced using labor and commercial floor space according to a constant returns to scale Cobb-Douglas production technology. Markets are assumed to be perfectly competitive and this final good can be costlessly traded both within the city and the wider economy. Cost minimization and zero-profits imply that price equals unit cost in each location with positive production:

$$1 = \frac{1}{A_n} w_n^{\beta} q_n^{1-\beta}, \qquad 0 < \beta < 1, \tag{8}$$

where A_n denotes productivity; q_n is the price of commercial floor space; and we have used our choice of numeraire ($P_n = 1$).

Productivity (A_n) is assumed to depend on production fundamentals ($\overline{A}_n$) and production externalities ($\mathbb{A}_n$). Production fundamentals capture features of physical geography that make a location a more or less attractive place to produce independently of the surrounding concentration of economic activity (e.g., access to natural water). Production externalities capture the interactions between workers within the city (e.g., knowledge spillovers):

$$A_n = \overline{A}_n \mathbb{A}_n^{\eta^A}, \qquad \qquad \mathbb{A}_n \equiv \sum_{i \in \mathbb{N}} e^{-\delta^A \tau_{ni}} L_i, \tag{9}$$

where η^A governs the magnitude of these production externalities and δ^A parameterizes their spatial decay with travel time.

4.3 Commuter Market Clearing

Commuter market clearing requires that the measure of workers employed in each workplace equals the measure of workers commuting to that workplace. From equations (4)-(6), we can write this commuter market clearing condition as:

$$L_i = \sum_{n \in \mathbb{N}} \lambda^R_{ni|n} R_n, \qquad \text{where} \qquad \lambda^R_{ni|n} = \frac{\lambda_{ni}}{\lambda^R_n} = \frac{\left(w_i/\kappa_{ni}\right)^{\epsilon}}{\sum_{\ell \in \mathbb{N}} \left(w_\ell/\kappa_{n\ell}\right)^{\epsilon}}. \tag{10}$$

Commuter market clearing also implies that average income per capita in each residence (v_n) is equal to the sum across workplaces of the wage in each workplace (w_i) multiplied by the probability of commuting to that workplace ($\lambda^R_{ni|n}$):

$$v_n = \sum_{i \in \mathbb{N}} \lambda^R_{ni|n} w_i. \tag{11}$$

4.4 Land Market Clearing

In our baseline specification, we assume exogenous supplies of residential floor space (H^R_n) and commercial floor space (H^L_n) in each location. Market clearing for residential floor space implies that income from the ownership of residential floor space equals payments for its use:

$$Q_n H^R_n = (1-\alpha)\, v_n R_n. \tag{12}$$

Similarly, market clearing for commercial floor space implies that income from the ownership of commercial floor space equals payments for its use:

$$q_n H^L_n = \frac{1-\beta}{\beta} w_n L_n, \tag{13}$$

where we used the property of the Cobb-Douglas production technology that payments to commercial floor space and labor are constant multiples of one another.

These supplies of residential and commercial floor space in each location reflect geographical land area, the ratio of floor space to geographical land area (as reflected in the height of buildings), and the shares of floor space allocated to each use. In reality, both the overall supply of floor space and its allocation to each use are influenced by market forces, and we discuss extensions to allow for both an endogenous supply of floor space and an endogenous allocation to each use in Section 6 below. In our baseline specification, we assume that floor space is owned by geographically immobile absentee landlords.[16]

[16]Instead of absentee landlords, one can assume that landlords consume only the final good. Since this final good is costlessly traded with the wider economy and chosen as the numeraire, this alternative assumption leaves all equilibrium conditions in the model unchanged.

4.5 General Equilibrium

The general equilibrium spatial distribution of economic activity within the city is determined by the model parameters (α, β, κ, ϵ, η^B, δ^B, η^A, δ^A) and the following exogenous location characteristics: residential fundamentals ($\overline{B}_n$), production fundamentals ($\overline{A}_n$), the supplies of residential and commercial floor space (H_n^R, H_n^L), and the transport network (τ_{ni}). Given these parameters and exogenous location characteristics, the closed-city general equilibrium of the model is referenced by residents (R_n), employment (L_n), wages (w_n), average residential income (v_n), the prices of residential and commercial floor space (Q_n, q_n), and expected utility (U), given exogenous total city population ($L_{\mathbb{N}}$). The open-city general equilibrium is analogous, except that total city population is endogenously determined by population mobility with the wider economy and its exogenous reservation level of utility ($\overline{U}$). Given these equilibrium objects, all the other endogenous variables of the model can be determined.

We now show that the general equilibrium of the model admits a tractable theoretical characterization. In particular, the conditions for general equilibrium can be written in the form required to apply Theorem 1 from Allen et al. (2024a) for uniqueness:

$$x_{nh} = \sum_{i \in \mathbb{N}} \mathcal{K}_{nih} \prod_{h' \in \mathbb{H}} x_{ih'}^{\gamma_{nhh'}}, \tag{14}$$

where x_{nh} is an endogenous variable; $\mathcal{K}_{nih}$ is a kernel that characterizes bilateral frictions; $n, i \in \mathbb{N}$ denote locations; and $h \in \mathbb{H}$ denote economic interactions, which here include residents, employment, and the prices of residential and commercial floor space. A sufficient condition for the existence of a unique equilibrium is that the spectral radius of a coefficient matrix of model parameters is less than or equal to one, as shown in Online Appendix B.5.

One special case of the model that is particularly tractable (although not entirely realistic) is the case with no residential floor space use ($\alpha = 1$) and no commercial floor space use ($\beta = 1$), as analyzed in Allen et al. (2024a). In this special case, the sufficient condition for the existence of a unique equilibrium is that the spectral radius of the following matrix is less than one:

$$\boldsymbol{\Upsilon} = \begin{bmatrix} 0 & 0 & 0 & \left|\eta^B \epsilon\right| \\ 0 & 0 & \left|\eta^A \epsilon\right| & 0 \\ 1 & 0 & \left|\eta^A \epsilon\right| & 0 \\ 0 & 1 & 0 & \left|\eta^B \epsilon\right| \end{bmatrix}.$$

From the Collatz-Wielandt Formula, a sufficient condition for the spectral radius of this matrix to be less than one is $\left|\eta^A\right| \epsilon \leq \frac{1}{2}$ and $\left|\eta^B\right| \epsilon \leq \frac{1}{2}$, which has an intuitive interpretation as the requirement that the agglomeration forces (production externalities (η^A) and residential externalities (η^B)) are sufficiently weak relative to the dispersion forces from the heterogeneity in idiosyncratic preferences (ϵ).

This approach to characterizing the existence and uniqueness of the general equilibrium is extremely powerful, in the sense that it only depends on model parameters and hence holds for any set of location characteristics, including production fundamentals ($\overline{A}_n$), residential fundamentals ($\overline{B}_n$), supplies of residential and commercial floor space (H_n^R, H_n^L), and the transport network (τ_{ni}). Nevertheless, this condition is sufficient but not necessary, in the sense that there could exist sets of location characteristics for which the equilibrium is unique, even if this condition on parameters does not hold.

In the presence of a unique equilibrium, counterfactuals for changes in natural advantages or public-policy interventions have determinate predictions for the spatial distribution of economic activity. Therefore, uniqueness is a convenient property of quantitative urban models for evaluating the impact of these public-policy interventions. But a central feature of the earlier theoretical literature on new economic geography was the presence of multiple equilibria, as in the original core–periphery model of Krugman (1991). Furthermore, at small spatial scales within cities, it is plausible that there is likely to be scope for multiple equilibria, because the differences in natural advantages are small (e.g., whether a coffee shop is on the North or South side of the street within a neighborhood).[17]

At a conceptual level, whether the equilibrium is unique or there are multiple equilibria may depend on the model's level of abstraction. All models are by necessity abstractions (otherwise they would not be models and would be literal descriptions of reality). At a high level of abstraction, the model may feature too few natural advantages to uniquely determine the spatial distribution of economic activity, such that there are multiple equilibrium. At a lower level of abstraction, the model may feature sufficiently many natural advantages (productivity, amenities, supply of floor space, etc.) that it ultimately uniquely determines the spatial distribution of economic activity.

However, even for the rich quantitative setting considered here, if agglomeration forces are sufficiently strong relative to dispersion forces, there remains the potential for multiple equilibria. We discuss the separate implications of multiple equilibria for parameter estimation and counterfactuals further below.

5 Properties of Quantitative Urban Models

In this section, we discuss a number of properties of quantitative urban models that lend themselves to empirical applications. Subsection 5.1 shows how these models can be inverted to quantify the contributions of first and second-nature geography to the observed spatial distribution of

[17]For empirical evidence on the extent to which the spatial distribution of economic activity is unique or characterized by multiple steady-states, see Davis and Weinstein (2002, 2008), Redding et al. (2011), Bleakley and Lin (2012), Michaels and Rauch (2018) and Takeda and Yamagishi (2023).

economic activity for known parameter values. Subsection 5.2 discusses parameter estimation. Subsection 5.3 shows how these models can be used to undertake counterfactuals. Subsection 5.4 uses the general equilibrium conditions of a class of these quantitative urban models to derive sufficient statistics for the spatial distribution of economic activity.

5.1 First Versus Second-Nature Geography

In this section, we show how quantitative urban models can be used to assess the role of first versus second-nature geography in explaining the spatial distribution of economic activity within cities. We take the parameters of the model as known, before discussing parameter estimation in the next subsection. A key feature of quantitative urban models is that they are sufficiently rich so as to be able to rationalize the observed data as an equilibrium of the model. They include structural residuals that vary by location, such as the production and residential fundamentals in our baseline model above. These structural residuals can adjust by location, such that the observed data are consistent with the structural equations of the model.

This property of quantitative urban models typically implies that they are invertible, in the sense that given known parameters and the observed endogenous variables, we can back out unique values of the structural residuals such that the model is consistent with the observed data. Furthermore, this invertibility property can hold even in the presence of multiple equilibria, because these models condition on the observed equilibrium in the data. Given known model parameters, the observed endogenous variables and the equilibrium conditions of the model (e.g., cost minimization, zero profits and population mobility) can together contain enough information to uniquely determine these structural residuals, even though there could have been another equilibrium for the same value of the model parameters.

We now illustrate this invertibility property using our baseline model from Section 4. We suppose that the parameters of the model are known (α, β, κ, ϵ, η^B, δ^B, η^A, δ^A). We suppose that a researcher observes employment by residence (R_n), employment by workplace (L_i), residential and commercial floor prices (Q_n, q_n), and bilateral commuting times (τ_{ni}). First, given these observed variables, the researcher can compute bilateral commuting costs ($\kappa_{ni}^{\epsilon} = e^{\kappa\epsilon\tau_{ni}}$), and solve for a unique vector of equilibrium wages (w_i) from the commuter market clearing condition (10) (up to a choice of units). Second, using these solutions for wages (w_i), the researcher can compute residents' commuting market access (Φ_n^R) from equation (5) and average residential income (v_n) from equation (11).

Third, given the known model parameters, and observed residents (R_n) and travel times (τ_{ni}), the researcher can compute residential externalities ($\mathbb{B}_n$) from equation (2). Fourth, using the observed shares of residents in each location (λ_n^R) and observed prices of residential floor space (Q_n), together with the solutions for residents' commuting market access (Φ_n^R) and residential externalities ($\mathbb{B}_n$), the researcher can recover unobserved residential fundamentals ($\overline{B}_n$) from the residential choice probabilities (5):

$$\overline{B}_n = \left(\frac{U}{\vartheta}\right) \mathbb{B}_n^{-\eta^B} \left(\lambda_n^R\right)^{\frac{1}{\epsilon}} \left(\Phi_n^R\right)^{-\frac{1}{\epsilon}} Q_n^{1-\alpha}, \tag{15}$$

where residential fundamentals ($\overline{B}_n$) are determined up to a choice of units in which to measure the common level of utility across all locations (U) and the above choice of units in which to measure wages (w_i). Intuitively, if a location has a high share of residents (λ_n^R) and a high price of residential floor space (Q_n), despite low commuting market access (Φ_n^R) and low residential externalities ($\mathbb{B}_n$), it must have unobserved residential fundamentals ($\overline{B}_n$) that make it an attractive place to live.

Fifth, given the known model parameters, and observed workers (L_n) and travel times (τ_{ni}), the researcher can compute production externalities ($\mathbb{A}_n$) from equation (9). Sixth, using the observed commercial floor prices (q_n), together with the solutions for wages (w_n) and production externalities ($\mathbb{A}_n$), the researcher can recover unobserved production fundamentals ($\overline{A}_n$) from the zero-profit condition (8):

$$\overline{A}_n = \mathbb{A}_n^{-\eta^A} w_n^{\beta} q_n^{1-\beta}, \tag{16}$$

where production fundamentals ($\overline{A}_n$) are determined up to the above choice of units in which to measure wages (w_i). Intuitively, if a location offers high wages (w_n) despite a high price of commercial floor space (q_n) and low production externalities ($\mathbb{A}_n$), it must have high unobserved production fundamentals ($\overline{A}_n$) that make it an attractive place to produce.

Seventh, given the known model parameters, the observed values of residents (R_n), workers (L_n) and prices of floor space (Q_n, q_n) and the solutions for wages (w_n) and average residential income (v_n), the researcher can recover the unobserved supplies of residential floor space (H_n^R) and commercial floor space (H_n^L) from the floor space market clearing conditions (12) and (13). Analogous intuitions hold. If a location has a low price of residential floor space (Q_n), despite many residents (R_n) and a high average income (v_n), it must have a large supply of residential floor space (H_n^R). Similarly, if a location has a low price of commercial floor space (q_n), despite high employment (L_n) and high wages (w_n), it must have a large supply of commercial floor space (H_n^L). Eighth and finally, given observed geographical land area (G_n), the researcher can then recover the ratios of residential floor space to land area (H_n^R/G_n) and commercial floor space to land area (H_n^L/G_n).

This invertibility property of quantitative urban models implies that they can be used to decompose the observed variation in the endogenous variables across locations into the contributions of first-nature and second-nature geography. First-nature geography here corresponds to production fundamentals ($\overline{A}_n$), residential fundamentals ($\overline{B}_n$) and geographical land area (G_n).

Second-nature geography here corresponds to residence and workplace commuting market access (Φ_n^R, Φ_n^L), residential externalities ($\mathbb{B}_n$), production externalities ($\mathbb{A}_n$), the density of residential development (H_n^R/G_n) and the density of commercial development (H_n^L/G_n). Therefore, second-nature geography incorporates agglomeration forces (residential and production externalities), the location of firms and workers relative to one another in commuting markets (as captured in commuting market access), and the decision to invest in costly building structures (through the densities of residential and commercial development).

The ability of quantitative urban models to rationalize the observed data in an initial equilibrium is important for at least two reasons. First, this property ensures that these models provide an internally-consistent way of thinking about the data, in which the structural residuals have clear economic interpretation within the model. Otherwise, a researcher would need to include an additional *ad hoc* residual to capture the discrepancy between the model and the data, and the researcher would have limited ways of thinking about this residual within the model. Second, since the observed spatial distribution of economic activity is an initial equilibrium of these models, they can be used to undertake counterfactuals to solve for how public-policy interventions would change this observed spatial distribution of economic activity.

Nevertheless, the presence of these structural residuals has two important implications for the analysis of these models. First, the ability of quantitative urban models to explain the data in the observed equilibrium should not be interpreted as "success" in terms of model fit, because the model has enough degrees of freedom in the form of the structural residuals to exactly match the observed data for any parameter values. Therefore, the model fit of quantitative urban models must be assessed in other ways. For example, the model's predictions can be compared to separate data not used in its calibration, such as separate data on the ratio of floor space to land area. One would not expect the predictive power of the model to be perfect, because it is an abstraction and the ratio of floor space to land area in the model is implicitly quality-adjusted, whereas the observed data on this variable need not perfectly adjust for quality. Or the structural residuals of residential fundamentals ($\overline{B}_n$) and production fundamentals ($\overline{A}_n$) can be compared to plausible empirical proxies for these variables, such as scenic views and access to natural water. Alternatively, in the presence of an exogenous shock that is uncorrelated with the structural residuals, one can examine the quantitative success of the model in explaining the observed changes in the spatial distribution of economic activity in response to this shock, holding constant the structural residuals at their values in the initial equilibrium before the exogenous shock. We return to discuss these points in our empirical applications below.

Second, since the model has enough degrees of freedom to match the observed data for any vector of parameters, additional information is required in order to estimate these parameters. Typically, this additional information comes in the form of orthogonality conditions, such that the structural residuals or changes in these structural residuals are uncorrelated with an exogenous shock to the spatial distribution of economic activity. Again we provide examples of these exogenous shocks in our empirical applications below.

5.2 Parameter Estimation

A key difference between quantitative urban models and earlier computable general equilibrium (CGE) models is that they have a relatively small number of structural parameters (elasticities) to be estimated. This parsimonious parameterization has leant itself to transparent identification using quasi-experimental sources of variation in empirical applications.[18]

In this section, we focus on estimating the model's commuting parameters. We begin by re-writing the commuting gravity equation (4) as follows:

$$L_{ni} = \mathcal{R}_n \mathcal{W}_i e^{-\phi \tau_{ni}} u_{ni}, \tag{17}$$

where we have used our parameterization of commuting costs (κ_{ni}) as a function of travel time (τ_{ni}): $\kappa_{ni}^{-\epsilon} = e^{-\kappa \epsilon \tau_{ni}}$; $\phi \equiv \kappa \epsilon$ is a composite parameter that depends on the response of commuting costs to travel time (κ) and the dispersion of idiosyncratic preferences (ϵ); the residence fixed effects ($\mathcal{R}_n$) summarize the attractiveness of each residence, as measured by the prices of consumption goods (P_n) and residential floor space (Q_n) in the numerator of equation (4); the workplace fixed effects ($\mathcal{W}_i$) summarize the attractiveness of the workplace, as measured by wages (w_i) in the numerator of equation (4); and we have absorbed the common values of total city population ($L_{\mathbb{N}}$) and expected utility (U) from the denominator of equation (4) into the fixed effects; u_{ni} is a stochastic error that can be interpreted as unobserved components of bilateral commuting costs not captured in travel times, measurement error or small-sample variation.

A reduced-form empirical literature finds that this gravity equation provides a good approximation to observed bilateral commuting flows.[19] The specification in equation (17) assumes an exponential relationship between commuting costs and travel time, which has the advantage that the commuting cost iceberg lies in the unit interval ($\kappa_{ni}^{-\epsilon} = e^{-\phi \tau_{ni}} \in (0, 1]$ for $\phi > 0$), and implies a semi-log relationship between commuting flows and travel time. But a geometric relationship between commuting costs and travel time also provides a good approximation to the data ($\kappa_{ni}^{-\epsilon} = \tau_{ni}^{-\phi}$), and implies a log-log relationship between commuting flows and travel time. In comparing these two specifications, the semi-log relationship in equation (17) tends to fit the data marginally better at small spatial scales within cities.

[18] For a review of the earlier CGE literature, see Shoven and Whalley (1992). For a wider conceptual discussion of the connection between models and data, see Donaldson (2022).

[19] For surveys of this reduced-form empirical literature, see for example Fortheringham and O'Kelly (1989) and McDonald and McMillen (2010).

Despite this explanatory power, there are several empirical challenges in estimating the commuting gravity equation. A first challenge is the measurement of bilateral travel times. Perhaps the most common approach involves specifying a transport network, including railway lines, bus schedules and driving time by road. Given assumptions about the travel speed ("weight" or "cost") for each mode of transport, a researcher can construct least-cost path travel times between the centroids of locations. The researcher must decide whether agents can connect to a transport mode anywhere along its route, or whether they must connect to it at designated stations, and must choose how much of a travel time cost is incurred in changing between modes of transport. Given these choices, travel times between locations can be computed using commonly-available geographical information systems (GIS) software, such as ArcGIS, QGIS or R.[20] A related alternative is to compute bilateral travel times between locations using API-based routing software, such as Google Maps or OpenStreetMap.[21]

A related approach is to allow agents to choose between alternative transport modes according to a discrete choice model, such as a nested logit specification. In this extension, the choice of transport mode can depend on a number of observed characteristics besides travel time. This extension involves estimating the elasticities of agents' choice of transport mode with respect to travel time. Typically, this elasticity with respect to travel time is finite, such that transport modes are imperfect substitutes for one another, in which case agents do not necessarily choose the option with the lowest travel time.[22] A related approach allows agents to choose between alternative routes between a given origin and destination according to a discrete choice model, in which agents have idiosyncratic preferences over these routes, as in Allen and Arkolakis (2022b) and Allen et al. (2024b). Assuming that these idiosyncratic preferences have an extreme value distribution, agents have a finite elasticity of substitution across routes, and hence again need not choose the option with the lowest travel time.

One limitation of most travel time specifications in quantitative urban models is that they abstract from congestion, even though this is a key feature of transportation within large metropolitan areas. Modelling congestion is challenging because of the two-way interaction between the transport network and the location of economic activity. On the one hand, improvements in transport infrastructure cause agents to make different routing choices, which alters traffic and congestion throughout the network. On the other hand, changes in congestion affect travel costs, and hence in turn influence the spatial distribution of economic activity, as agents re-optimize their location decisions. But this change in the spatial distribution of economic activity in turn

[20] Examples of this approach include Donaldson (2018), Donaldson and Hornbeck (2016), and the calculation of travel time using public transport in Ahlfeldt et al. (2015).

[21] See for example as well as Miyauchi et al. (2022) , Akbar et al. (2023) and Akbar et al. (2024).

[22] Examples include the estimation of the choice between private and public transport in Ahlfeldt et al. (2015) and Tsivanidis (2024).

feeds back to influence traffic and congestion. Recent analyses of the spatial implications of congestion include Duranton and Turner (2011), Couture et al. (2018), Allen and Arkolakis (2022b), Akbar et al. (2023) and Kreindler (2024).

Besides the measurement of travel time, a second empirical challenge in estimating the commuting gravity equation is that the placement of transport infrastructure is unlikely to be randomly assigned. Instead transport infrastructure could be targeted toward locations that would have experienced different patterns of economic activity, even in the absence of this transport infrastructure. On the one hand, transportation networks are often operated by private-sector companies, whose search for profits could lead them to target locations that otherwise would have grown more rapidly. On the other hand, local governments provide incentives for the provision of transport infrastructure in locations that are unattractive to private-sector companies, which could target locations that otherwise would have grown less rapidly. A substantial reduced-form empirical literature uses a variety of quasi-experimental methods to address this concern, including the use of instrumental variables, such as straight-line distance, planned routes, or historical routes; and the use of variation from inconsequential places, which are only connected to the transport network because they happen to lie along the way between another pair of more consequential locations.[23]

A third empirical challenge in estimating the commuting gravity equation is the presence of zero bilateral commuting flows in the data. In our baseline model with a continuous measure of workers, all bilateral pairs of locations have positive bilateral commuting flows for positive and finite values of commuting costs, productivity and amenities. The reason is that the Fréchet distribution has a support that is unbounded from above, which implies that there is always a small measure of workers who draw a realization for idiosyncratic preferences for which any given residence-workplace pair is the preferred pair. One potential way of rationalizing zero bilateral commuting flows in the data is to assume prohibitive commuting costs ($\kappa_{ni}^{-\epsilon} \to 0$) for residence-workplace pairs with zero flows. While this assumption provides a theory-consistent rationalization of the observed data, it is not especially compelling conceptually, and it has the unattractive implication for counterfactuals of ruling out zero flows becoming positive. An alternative approach recognizes that these zero bilateral flows are typically observed using data on small spatial units, where the number of possible bilateral residence-workplace pairs is large relative to the number of agents, as examined in Dingel and Tintelnot (2020) based on the approach in the international trade literature from Eaton et al. (2013). Therefore, a natural explanation for zero observed bilateral flows is small sample variation or granularity, where for any finite sample of workers we may not observe a realization of idiosyncratic preferences for which there is positive commuting for a given residence-workplace pair, even though with a continuous measure

[23]For surveys of this empirical literature, see Donaldson (2015), Redding and Turner (2015), and Donaldson (2024).

of workers, positive commuting for that pair would be observed.

From the perspective of this second approach based on granularity, the observed data on residence-workplace choices can be interpreted as a finite sample from the continuum model, in which case the parameters of the model can be estimated using the canonical conditional-logit likelihood of McFadden (1974) applied to location choices. The maximum of this likelihood function is numerically equivalent to the Poisson Pseudo Maximum Likelihood (PPML) estimator, as shown in Guimarães et al. (2003) and Dingel and Tintelnot (2020). This PPML estimator is commonly used to estimated the gravity equation for goods trade in the international trade literature following Santos Silva and Tenreyro (2006) and Head and Mayer (2014). Therefore, in the presence of granularity, the predicted commuting flows from the PPML gravity estimation become key model inputs in the place of the observed bilateral commuting flows in the data.

In this subsection, we have focused on estimating the response of bilateral commuting flows to bilateral travel time (ϕ). Some of the other parameters of quantitative urban models are relatively straightforward to determine. The residential floor space parameter (α) can be disciplined using data on the share of housing in household expenditure. The commercial floor space parameter (β) can be pinned down using data on the share of land, buildings and structures in firm costs. In contrast, the dispersion of idiosyncratic preferences (ϵ), the production agglomeration parameters (η^A, δ^A) and the residential agglomeration parameters (η^B, δ^B) are more challenging to estimate. Existing empirical research has largely sought to estimate these parameters using quasi-experimental variation, as discussed further in the empirical applications in Section 7.

The possibility of multiple equilibria in quantitative urban models for a range of parameter values presents challenges for some estimation approaches. In particular, simulation methods such as indirect inference or simulated method of moments (SMM) involve simulating the model, solving for equilibrium, and comparing moments in the model and in the data. When making this comparison for a given set of parameters, the discrepancy between moments in the model and data could depend on which equilibrium has been selected. One approach to overcoming this challenge is to use classical estimation methods, such as maximum likelihood or generalized method of moments (GMM), in which the estimation can be undertaken conditional on the observed equilibrium in the data. By conditioning on this observed equilibrium, the model's parameters can be estimated regardless of whether or not another equilibrium could have occurred.[24]

5.3 Model Counterfactuals

Since quantitative urban models are able to rationalize the observed spatial distribution of economic activity within an urban area, they provide a suitable platform for undertaking counterfac-

[24]See for example Ahlfeldt et al. (2015), as discussed further in Section 7 below.

tuals for how realistic public-policy interventions would change this observed spatial distribution of economic activity. One common empirical application is a transport infrastructure improvement between a particular set of locations.

In general, two main approaches to undertaking counterfactuals have been taken. The first "covariates" approach estimates the model, recovers the predicted values of model variables such as commuting costs, and undertakes counterfactuals using these model predictions. The second "calibrated shares" approach undertakes counterfactuals conditioning on observed bilateral commuting flows in the initial equilibrium in the data. Following the international trade literature (Dekle et al. 2007 and Costinot and Rodríguez-Clare 2014), the calibrated shares approach is often referred to as "exact-hat algebra," because it rewrites the model's counterfactual equilibrium conditions in terms of the observed values of variables in the initial equilibrium in the data and the relative changes of variables ("hats") between the initial equilibrium and the counterfactual equilibrium. The approach is exact, in the sense that it solves for a counterfactual equilibrium using the full structure of the non-linear model, without any approximation. We discuss the strengths and limitations of these two approaches below.

We begin by illustrating the calibrated shares approach, before then discussing how to undertake counterfactuals using the covariates approach. We consider a counterfactual in our baseline model from Section 4 for the construction of a subway line that reduces bilateral travel times (τ_{ni}) for some pairs of locations by more than for other pairs of locations. We consider a closed-city and assume exogenous supplies of residential and commercial floor space (H_n^R, H_n^L). We allow both amenities (B_n) and productivity (A_n) to respond endogenously to changes in the spatial distribution of economic activity through residential and production externalities ($\mathbb{B}_n$, $\mathbb{A}_n$). We denote the value of variables in the counterfactual equilibrium by a prime (x'), the value of variables in the observed equilibrium without a prime (x), and the relative changes of variables between the two equilibria by a hat ($\widehat{x} = x'/x$).

Given an exogenous change in travel times ($\widehat{\tau}_{ni}$), and hence commuting costs, and initial guesses for changes in wages ($\widehat{w}_n^0$) and the price of residential floor space ($\widehat{Q}_n^0$), we can re-write the system of equations for a counterfactual equilibrium in the closed-city as follows:

$$\widehat{\lambda}_n^R \lambda_n^R = \frac{\sum_{\ell \in \mathbb{N}} \lambda_{n\ell} \left(\widehat{B}_n \widehat{w}_\ell^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{n\ell}} \right) \left(\widehat{Q}_n^0 \right)^{-\epsilon(1-\alpha)}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \lambda_{k\ell} \left(\widehat{B}_k \widehat{w}_\ell^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{k\ell}} \right) \left(\widehat{Q}_k^0 \right)^{-\epsilon(1-\alpha)}}, \tag{18}$$

$$\widehat{\lambda}_n^L \lambda_n^L = \frac{\sum_{k \in \mathbb{N}} \lambda_{kn} \left(\widehat{B}_n \widehat{w}_n^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{kn}} \right) \left(\widehat{Q}_k^0 \right)^{-\epsilon(1-\alpha)}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \lambda_{k\ell} \left(\widehat{B}_\ell \widehat{w}_\ell^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{k\ell}} \right) \left(\widehat{Q}_k^0 \right)^{-\epsilon(1-\alpha)}}, \tag{19}$$

$$\widehat{\lambda}_{ni|n}^R \lambda_{ni|n}^R = \frac{\lambda_{ni|n}^R \left(\widehat{w}_i^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{ni}} \right)}{\sum_{\ell \in \mathbb{N}} \lambda_{n\ell|n}^R \left(\widehat{w}_\ell^0 \right)^{\epsilon} \left(e^{-\phi \widehat{\tau}_{n\ell}} \right)}, \tag{20}$$

$$\widehat{R}_n R_n = \widehat{\lambda}_n^R \lambda_n^R L_{\mathbb{N}}, \qquad \widehat{L}_n L_n = \widehat{\lambda}_n^L \lambda_n^L L_{\mathbb{N}}, \tag{21}$$

$$\widehat{q}_n = \widehat{A}_n^{\frac{1}{1-\beta}} \left(\widehat{w}_n^0\right)^{-\frac{\beta}{1-\beta}}, \qquad \widehat{v}_n v_n = \sum_{i \in \mathbb{N}} \widehat{\lambda}_{ni|n} \lambda_{ni|n} \widehat{w}_i^0 w_i, \tag{22}$$

$$\widehat{B}_n = \left(\frac{\sum_{i \in \mathbb{N}} e^{-\delta^B (\widehat{\tau}_{ni} + \tau_{ni})} \widehat{R}_i R_i}{\sum_{i \in \mathbb{N}} e^{-\delta^B \tau_{ni}} R_i} \right)^{\eta^B}, \qquad \widehat{A}_n = \left(\frac{\sum_{i \in \mathbb{N}} e^{-\delta^A (\widehat{\tau}_{ni} + \tau_{ni})} \widehat{L}_i L_i}{\sum_{i \in \mathbb{N}} e^{-\delta^A \tau_{ni}} L_i} \right)^{\eta^A}, \tag{23}$$

where, with a slight abuse of notation, we denote $\widehat{\tau}_{ni} = \tau'_{ni} - \tau_{ni}$; we have used $\widehat{\overline{A}}_n = \widehat{\overline{B}}_n = \widehat{H}_n^R = \widehat{H}_n^L = 1$; and the derivations are reported in Online Appendix B.7.

From this system of equations, we obtain implied changes in labor demand ($\widehat{L}_n^D$) and labor supply ($\widehat{L}_n^S$) in each location:

$$\widehat{L}_n^D = \frac{\widehat{q}_n}{\widehat{w}_n^0}, \qquad \widehat{L}_n^S = \widehat{\lambda}_n^L, \tag{24}$$

and the implied changes in income from residential floor space (Ω_n^S) and expenditure (Ω_n^D) on residential floor space:

$$\widehat{\Omega}_n^S = \widehat{Q}_n^0, \qquad \widehat{\Omega}_n^D = \widehat{v}_n \widehat{R}_n. \tag{25}$$

To solve for the counterfactual equilibrium, we update our initial guesses for changes in wages ($\widehat{w}_n^0$) and the price of residential floor space ($\widehat{Q}_n^0$) until the changes in the demand and supply of labor are equal to one another ($\widehat{L}_n^D = \widehat{L}_n^S$) in equation (24) and the changes in the demand and supply for residential floor space are equal to one another ($\widehat{\Omega}_n^D = \widehat{\Omega}_n^S$) in equation (25), and all the other equilibrium conditions of the model are satisfied.

In this system of general equilibrium conditions, the direct effect of the construction of the new subway line is to reduce bilateral travel times for some residence-workplace pairs by more than for others ($\widehat{\tau}_{ni} < \widehat{\tau}_{k\ell}$ for some n, i, k, ℓ). In response, workers adjust their residence and workplace decisions. The resulting changes in employment by workplace ($\widehat{L}_n$) affect productivity in each location through agglomeration forces (production externalities ($\widehat{\mathbb{A}}_n$)). The resulting changes in employment by residence ($\widehat{R}_n$) affect amenities in each location through agglomeration forces (residential externalities ($\widehat{\mathbb{B}}_n$)). The mechanisms that restore equilibrium are changes in wages ($\widehat{w}_n$) and the prices of residential and commercial floor space ($\widehat{Q}_n$, $\widehat{q}_n$) until (i) the demand for workers equals the supply of commuters in each location with positive production; (ii) firms make zero profits in each location with positive production; (iii) each worker chooses her preferred residence-workplace pair, and expected utility conditional on choosing a residence-workplace is the same across all residence-workplace pairs; (iv) demand equals supply for residential and commercial floor space.

Counterfactuals in the covariates approach can be written in the same form as for the calibrated shares approach above, except that instead of using the observed endogenous variables in the initial equilibrium in the data, one instead uses the estimated model's predictions for these endogenous variables in the initial equilibrium. Given these model predictions, one can solve for the counterfactual relative changes in the endogenous variables using the general equilibrium conditions of the model following the same approach as for the calibrated shares approach above.

The calibrated shares approach has its advantages and disadvantages. An advantage is that it avoids having to estimate unobserved location characteristics in the initial equilibrium, because it uses the observed values of the endogenous variables in the initial equilibrium to control for these unobserved location characteristics. In our example here, observed initial commuting shares (λ_{ni}) are used to capture unobserved initial commuting costs. A disadvantage is that the calibrated shares approach conditions on the realized values of the endogenous variables in the initial equilibrium. In empirical settings with small spatial units in which granularity is a concern, these realized values are subject to small sample variation, introducing a source of error into the model's counterfactual predictions.

More broadly, when undertaking counterfactuals, a key question for the researcher is what to hold constant. Both the calibrated shares and covariates approaches hold unobserved location characteristics constant, except for the assumed counterfactual change (here in bilateral travel times). The model allows for endogenous changes in amenities and productivity (through residential and production externalities) and endogenous changes in commuting market access. But it assumes that residential fundamentals ($\overline{B}_n$), production fundamentals ($\overline{A}_n$) and the supplies of residential and commercial floor space (H_n^R, H_n^L) are invariant with respect to the policy counterfactual. If this assumption were violated, a form of the Lucas Critique would apply, and the instability of these location characteristics in response to the policy change would introduce a source of error into the model's counterfactual predictions. In our example here, this assumption would be violated if the supplies of floor space (H_n^R, H_n^L) responded to changes in their price (Q_n, q_n). We discuss below how to generalize the model to allow an endogenous floor space response. An approach to assessing the sensitivity of results to this concern is to undertake counterfactuals under different assumptions about what location characteristics are held constant (e.g., with and without agglomeration forces, and with and without endogenous responses in the supply of floor space), and examine the robustness of quantitative conclusions across these different scenarios.

As discussed above, parameter estimation of quantitative urban models can be robust to the existence of multiple equilibria, to the extent that this estimation can be undertaken conditional on the observed equilibrium in the data, regardless of whether or not another possible equilibrium exists. But multiple equilibria poses more of a serious challenge for undertaking counterfactuals. In the region of the parameter space for which there is a unique equilibrium, quantitative urban models yield determinate predictions for the impact of a counterfactual policy change on the spatial distribution of economic activity. In contrast, in the region of the parameter space for which there are multiple equilibria, a given counterfactual can have different effects on the spatial distribution of economic activity depending on which equilibrium is selected. Therefore, the researcher must specify an equilibrium selection rule in order to obtain determinate predictions, and it can be unclear what the appropriate equilibrium selection rule is.

One common approach is to solve for a counterfactual equilibrium using the initial values of the endogenous variables from the observed equilibrium in the data. A related approach is to search for alternative equilibria by perturbing these initial values used to solve for a counterfactual equilibrium. However, in this search, a researcher never can be sure of having enumerated all possible equilibria. Further work is needed on equilibrium selection rules and methods for undertaking counterfactuals in models with multiple equilibria, as discussed further below. One promising approach is to undertake Monte Carlo simulations over different initial conditions to obtain a distribution of potential counterfactual outcomes, as in Allen and Donaldson (2020).

5.4 Sufficient Statistics

Solving for counterfactuals using the full non-linear model structure has a number of important advantages. One uses an internally-consistent equilibrium framework to evaluate the impact of counterfactual public policies. This approach makes clear what assumptions are made, what is held constant, and what parameter values are used. Budget constraints are imposed and market clearing conditions hold. Furthermore, the solution for the counterfactual equilibrium is exact, because no approximation is invoked.

Nevertheless, there are some potential disadvantages to solving for counterfactuals using the full non-linear model structure. It can be unclear which predictions depend on the entire model structure versus which predictions would hold in a broader class of models. It is also unclear how sensitive counterfactual predictions are to perturbations in model assumptions.

An alternative approach that addresses some of these disadvantages is to evaluate counterfactual policies using sufficient statistics. According to this approach, counterfactual predictions are derived from a smaller number of reduced-form equations that hold in a larger class of models. These counterfactual predictions only depend on observed variables and reduced-form parameters that are combinations of the structural parameters of each model in that class.[25]

In Online Appendix C, we show that the commuting gravity equation (4) holds in an entire class of quantitative urban models that make different assumptions about preferences, trade costs and market structure in the goods market. This class of models includes: (i) a new economic geography model with monopolistic competition, increasing returns to scale and trade costs, as in Helpman (1998), Redding and Sturm (2008) and Monte et al. (2018); (ii) a Ricardian spatial model

[25]For a review of the public finance literature on sufficient statistics, see Chetty (2009).

with perfect competition, constant returns to scale and trade costs, as in Eaton and Kortum (2002) and Redding (2016); (iii) an Armington spatial model with a neoclassical production structure and trade costs, as in Armington (1969), Allen and Arkolakis (2014) and Allen et al. (2017).

This class of quantitative urban models characterized by a gravity equation for bilateral commuting flows lends itself to a sufficient statistics representation in terms of commuting market access, as shown in Tsivanidis (2024). Using the residential choice probability (5) and the workplace choice probability (6), we can rewrite residents and workplace market access (Φ_n^R, Φ_i^L) as follows:

$$\Phi_n^R = \frac{1}{\xi} \sum_{i \in \mathbb{N}} \kappa_{ni}^{-\epsilon} \frac{L_i}{\Phi_i^L},$$
$$\Phi_i^L = \frac{1}{\xi} \sum_{n \in \mathbb{N}} \kappa_{ni}^{-\epsilon} \frac{R_n}{\Phi_n^R},$$

as shown in Section B.5 of the online appendix.

Given data on employment (L_i) and residents (R_n), and estimates of bilateral commuting costs ($\kappa_{ni}^{-\epsilon}$) from gravity equation estimation, residents and workplace market access (Φ_n^R and Φ_i^L, respectively) can be recovered (up to scale) from this system of equations.[26] Under the assumption of no agglomeration forces in production or residential decisions ($\eta^B = \eta^A = 0$ in equations (2) and (9)), the impact of a change in commuting costs ($\kappa_{ni}^{-\epsilon}$) on the spatial distribution of economic activity within the city can be represented solely in terms of changes in commuting market access:

$$\begin{aligned} \ln \widehat{y}_n^R &\approx \rho^R \ln \widehat{\Phi}_n^R + e_n^R, \\ \ln \widehat{y}_n^L &= \rho^L \ln \widehat{\Phi}_n^L + e_n^R, \end{aligned} \tag{26}$$

where the hat above a variable again denotes a relative change between an initial and counterfactual equilibrium; the outcome variables $\widehat{y}_n^R = \left[\widehat{R}_n, \widehat{Q}_n\right]$ and $y_n^L = \left[\widehat{L}_n, \widehat{q}_n\right]$ are changes in residents, residential floor prices, employment, and commercial floor prices; the reduced-form coefficients (ρ^R, ρ^L) depend on structural parameters of the model; the residuals (e_n^R, e_n^L) capture changes in location characteristics (productivity, amenities, and the supplies of residential and commercial floor space); in the specification model considered here, in which workers' idiosyncratic shocks are to preferences (rather than productivity), the first equation involves an approximation around an equilibrium with prohibitive commuting costs between locations (such that $v_n \approx \left(\Phi_n^R\right)^{1/\epsilon}$); and the derivations are reported in Online Appendix B.6.

Under the assumption of no agglomeration forces, this system of equations (26) implies that changes in the transport network only matter for equilibrium outcomes through commuting market access. Indeed, the change in the entire spatial distribution of economic activity within the

[26] An alternative approach is to estimate residential and workplace commuting market access using the fixed effects from the gravity equation (17), $\Phi_n^R = \sum_{i \in \mathbb{N}} e^{-\kappa \epsilon \tau_{ni}} \mathcal{W}_i$ and $\Phi_i^L = \sum_{n \in \mathbb{N}} e^{-\kappa \epsilon \tau_{ni}} \mathcal{R}_n$, as in Redding (2022a).

city depends only on the change in commuting market access and changes in structural residuals that capture changes in location characteristics (productivity, amenities, and the supplies of residential and commercial floor space). This role of commuting market access in urban models with a commuting gravity equation is analogous to the role of goods market access in models of trade between regions that feature a gravity equation for trade in goods. One limitation here is the assumption of no agglomeration forces. If production externalities depend on own-location employment alone, and residential externalities depend on own-location residents alone, commuting market access remains a sufficient statistic. However, in the presence of spatial spillovers of agglomeration forces across locations, counterfactual changes in model outcomes also depend on the structure of these spatial spillovers across locations. Another limitation is that the system (26) only holds approximately around an equilibrium with prohibitive commuting costs, which may not be a good approximation in settings with finely-disaggregated spatial units and large flows of commuters between locations.

6 Extensions and Generalizations

In this section, we consider a number of extensions and generalizations of the baseline quantitative urban model from Section 4. Subsection 6.1 considers open versus closed-city specifications and factor mobility with the wider economy. Subsection 6.2 discusses an endogenous supply of floor space and an endogenous allocation of floor space between residential and commercial use. Subsection 6.3 examines richer specifications of goods trade between locations. Subsection 6.4 considers spatial sorting by *ex ante* heterogeneous groups of workers. Subsection 6.5 introduces dynamics from mobility frictions or the gradual accumulation of immobile structures. Subsection 6.6 examines the role of consumption access in addition to commuter market access.

6.1 Closed versus Open Cities

In the baseline model in Section 4 above, we discussed the distinction between a closed-city (complete population immobility) and an open-city (perfect population mobility). This distinction is central to evaluating the welfare effects of public policy interventions. In the closed-city specification, public-policy interventions (e.g., transport improvements) directly affect worker welfare. In contrast, in the open-city specification, these public-policy interventions have no effects on worker welfare, because the total population of the city adjusts, until changes in wages and the price of residential floor space are such that worker welfare is equal to the unchanged reservation level of utility in the wider economy. Therefore, in the open-city specification, all welfare effects are experienced by landlords through changes in the value of land and buildings, as in the classical approach to valuing public goods following George (1879).

An intermediate case allows for an upward-sloping labor supply function for the city as a whole, with a positive but finite labor supply elasticity, as in Heblich et al. (2020). This intermediate case arises when workers make residence-workplace decisions across locations within and outside the city's boundaries. In this intermediate case, expected utility conditional on choosing a residence-workplace pair can be written as:

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

where $L_{\mathbb{N}}$ is the total population of the city and $L_{\mathbb{M}}$ is the total population of the entire economy.

Intuitively, for a given common level of expected utility in the wider economy ($\bar{U}$), locations within the city must offer higher real wages adjusted for amenities and commuting costs ($B_k w_\ell / \left(\kappa_{k\ell}P_k^{\alpha}Q_k^{1-\alpha}\right)$) to attract workers with lower idiosyncratic draws (thereby raising $L_{\mathbb{N}}/L_{\mathbb{M}}$), with an elasticity governed by the parameter ϵ.

6.2 Floor Space Supply Elasticities

In the baseline model in Section 4, we assumed exogenous supplies of residential and commercial floor space (H_n^R and H_n^L, respectively). But it is relatively straightforward to relax this assumption. We begin by allowing for an endogenous allocation of floor space between residential and commercial floor space, holding constant the total supply of floor space (H_n). In this case, the fractions of floor space allocated to commercial and residential use (θ_n, $1-\theta_n$, respectively) are determined by the following no-arbitrage condition:

$$\begin{array}{llll}\theta_n & =1 & \text{if} & q_n > \xi_n Q_n \\ \theta_n & \in[0,1] & \text{if} & q_n = \xi_n Q_n \\ \theta_n & =0 & \text{if} & q_n < \xi_n Q_n\end{array}, \tag{27}$$

where $\theta_n \equiv H_n^L/H_n$ and $\xi_n \geq 1$ is the tax equivalent of land use regulations.

We next allow the overall supply of floor space to be endogenously determined by a construction sector, as considered in Epple et al. (2010), Ahlfeldt et al. (2015) and Combes et al. (2021). We assume a competitive construction sector that produces floor space (H_n) with land (G_n) and capital (K_n) according to the following Cobb-Douglas production technology:

$$H_n = K_n^{\mu}G_n^{1-\mu}, \qquad 0<\mu<1, \tag{28}$$

where capital is assumed to be in perfect elastic supply from the wider economy at a constant price p_K.

From cost minimization in construction, equilibrium payments for capital are a constant share of payments for floor space ($\mathbb{Q}_n H_n$):

$$p_K K_n = \mu \mathbb{Q}_n H_n, \tag{29}$$

where $\mathbb{Q}_n = q_n \geq Q_n$ for $\theta_n \in (0,1]$ and $\xi_n \geq 1$; and $\mathbb{Q}_n = Q_n$ for $\theta_n = 0$.

Using this equilibrium condition (29) to substitute for capital (K_n) in the construction technology (28), we obtain a constant elasticity supply function for residential floor space:

$$H_n = \left(\frac{\mu}{p_K}\right)^{\frac{\mu}{1-\mu}} \mathbb{Q}_n^{\frac{\mu}{1-\mu}} G_n. \tag{30}$$

Therefore, this extension of the baseline quantitative urban model allows for an endogenous density of development. Those locations with high productivity, high amenities and low travel costs have high prices of floor space ($\mathbb{Q}_n$), other things equal, which induces the endogenous construction of taller buildings with higher ratios of floor space to geographical land area (H_n/G_n) in equation (30).[27] The Cobb-Douglas functional form for the construction technology (28) has been found to provide a good approximation to property transactions data, as shown in Ahlfeldt et al. (2015), and has been widely used in empirical work. Glaeser et al. (2005), Saiz (2010) and Harari (2020) provide evidence that these housing supply elasticities are heterogeneous and influenced by constraints of both physical geography and local regulations.

Although the assumption of a constant share of land in construction costs provides a good approximation to property transaction data, an important limitation of this model of the construction sector is that it is static. Glaeser and Gyourko (2005) provide empirical evidence that city growth and decline have asymmetric effects on the price of housing and that durable housing drives this asymmetry. As cities decline, the fall in the price of the existing stock of durable housing dampens the drop in population. In contrast, as cities expand, new construction of housing facilitates rapid population growth and mitigates the rise in the price of housing. We discuss further the implications of dynamics for quantitative urban models below.[28]

6.3 Goods Trade Assumptions

In the baseline model in Section 4, we considered a version of the canonical urban model, with a single homogeneous final good that is costlessly traded across locations. In reality, the goods produced in different locations can be imperfect substitutes for one another and can be subject to trade costs, such that the location of agents relative to one another matters in both commuting and goods markets.

Monte et al. (2018) develop a spatial general equilibrium model that features three-way interactions between locations through (i) goods trade, (ii) commuting, and (iii) migration. As the spatial scale of these three sets of interactions can differ from one another (e.g., commuting can

[27] A recent empirical literature examines the economics of skyscrapers, including Ahlfeldt and McMillen (2018), Ahlfeldt and Barr (2022) and Ahlfeldt et al. (2023).

[28] See also Baum Snow and Duranton (2024) for a more detailed discussion of the dynamics of housing supply.

be concentrated at small spatial scales, whereas goods trade extends over longer distances), this framework simultaneously models internal city structure and a system of cities.

The specification of worker indirect utility and commuting remains the same as in equation (1) in the baseline model in Section 4 above. But the consumption goods sector is modelled as in the new economic geography literature following Krugman (1991), with love of variety, increasing returns to scale and goods trade costs. The consumption goods price index in location n (C_n) is assumed to be a constant elasticity of substitution (CES) function of consumption of a continuum of tradable varieties sourced from each location i:

$$C_n = \left[\sum_{i \in \mathbb{N}} \int_0^{M_i} c_{ni} \left(j \right)^{\rho} dj \right]^{\frac{1}{\rho}}, \qquad \sigma = \frac{1}{1-\rho} > 1,$$

where $c_{ni}(j)$ is consumption in location n of variety j produced in location i; M_i is the endogenous measure of varieties produced in location i; and σ is the elasticity of substitution between varieties.

Varieties are produced under conditions of monopolistic competition and increasing returns to scale. Labor is the only factor of production. To produce a variety, a firm must incur a fixed cost (F) in terms of labor and a constant variable cost that depends on a location's productivity (A_i). Therefore, the total amount of labor ($l_i(j)$) required to produce $x_i(j)$ units of a variety j in location i is $l_i(j) = F + x_i(j)/A_i$. Varieties face iceberg variable costs of trade, such that $\tau_{ni} \geq 1$ units of a variety must be shipped from location i in order for one unit to arrive in location n. Profit maximization implies that equilibrium prices are a constant mark-up over marginal cost: $p_{ni}(j) = \left(\frac{\sigma}{\sigma-1} \right) \frac{\tau_{ni} w_i}{A_i}$, where w_i denotes the wage. Zero profits implies that the equilibrium output of each variety is a constant: $x_i(j) = A_i F (\sigma - 1)$. Labor market clearing implies that the endogenous measure of varieties produced in each location is proportional to the measure of employed workers (L_i): $M_i = L_i / (\sigma F)$.

With imperfect substitutability and trade costs, the model features a gravity equation for bilateral goods trade as well as for bilateral commuting flows. The share of location n's expenditure on goods produced in location i is:

$$\pi_{ni} = \frac{M_i p_{ni}^{1-\sigma}}{\sum_{k \in \mathbb{N}} M_k p_{nk}^{1-\sigma}} = \frac{L_i \left(\tau_{ni} w_i / A_i \right)^{1-\sigma}}{\sum_{k \in \mathbb{N}} L_k \left(\tau_{nk} w_k / A_k \right)^{1-\sigma}}. \tag{31}$$

Therefore, bilateral trade between locations depends on bilateral trade costs (τ_{ni}) in the numerator ("bilateral resistance") and on trade costs to all possible sources of supply in the denominator ("multilateral resistance").

With a gravity equation in both goods and commuting markets, the uniqueness of the general equilibrium can be characterized using the same techniques as for our baseline model in Section

4 above following Allen et al. (2024a).[29] Market access now matters for the spatial distribution of economic activity in both goods and commuting markets.[30] Labor market clearing again implies that employment in each workplace is equal to the number of people commuting to that workplace in equation (10), where the measure of residents in each location is endogenously determined through population mobility across locations.

One of the key implications of the gravity equations in goods and commuting markets in equations (4) and (31) is that elasticity of local economic activity with respect to local shocks (e..g., productivity shocks) is heterogeneous across locations, depending on the network of connections to other locations in goods and commuting markets. Therefore, local labor market shocks induced by globalization and technological change can have heterogeneous effects across local labor market areas depending on these networks of connections, which in turn implies heterogeneous effects of policy interventions in response to these shocks.[31]

6.4 Spatial Sorting

In the baseline model in Section 4, we considered a single worker type, such that workers are *ex ante* homogenous (before the realization of idiosyncratic preferences) and only *ex post* heterogenous (after the realization of idiosyncratic preferences). Another important extension is to consider multiple worker types and allow for *ex ante* heterogeneity. In the presence of this *ex ante* heterogeneity, public-policy interventions have income distributional consequences across worker types. Furthermore, spatial sorting across locations by these heterogeneous groups of workers is an important mechanism through which these income distribution consequences occur. Workers sort into and out of locations in response to place-based policy interventions, such that the set of people living in each location can be quite different before and after these interventions, as in the literature concerned with gentrification.[32]

A generalization of the baseline model from Section 4 is to assume multiple types of workers (e.g., high and low-education) with different structural parameters, as in Fajgelbaum and Gaubert (2020), Tsivanidis (2024), Redding and Sturm (2024), and Weiwu (2024).[33] Worker indirect utility (1) and the distribution of idiosyncratic preferences (3) now vary by worker type (e.g., education, occupation, income, race or ethnicity), such that the probability that a worker ω from type $o \in$

[29]See also Monte et al. (2018) and Allen and Arkolakis (2024).

[30]For further analysis of the role of market access in goods markets in shaping location choices see, Redding and Venables (2004), Hanson (2005), Redding and Sturm (2008), Donaldson and Hornbeck (2016), Redding (2016), Redding (2022b) and Allen and Arkolakis (2024).

[31]For further empirical evidence on local labor market responses to productivity and trade shocks, see Adão et al. (2019), Borusyak et al. (2024), and Hornbeck and Moretti (2024).

[32]For further discussion of research on spatial sorting and segregation, see Diamond and Suarez Serrato (2024).

[33]Other research in the broader literature concerned with neighborhood segregation includes Bayer et al. (2016), Fogli and Guerrieri (2019), and Chetty et al. (2016).

$\{1, \ldots, \mathbb{O}\}$ chooses residence n and workplace i is:

$$\lambda_{ni}^{o} = \frac{L_{ni}^{o}}{L_{\mathbb{N}}^{o}} = \frac{\left(B_{n}^{o} w_{i}^{o}\right)^{\epsilon^{o}} \left(\kappa_{ni}^{o} P_{n}^{\alpha^{o}} Q_{n}^{1-\alpha^{o}}\right)^{-\epsilon^{o}}}{\sum_{k \in \mathbb{N}} \sum_{\ell \in \mathbb{N}} \left(B_{k}^{o} w_{\ell}^{o}\right)^{\epsilon} \left(\kappa_{k\ell}^{o} P_{k}^{\alpha^{o}} Q_{k}^{1-\alpha^{o}}\right)^{-\epsilon^{o}}}. \tag{32}$$

Within this literature on spatial sorting, one strand of research simply assumes that observable worker types differ in terms of their underlying structural parameters, as in equation (32). In contrast, another line of work derives these differences in parameters from more primitive assumptions, such as non-homothetic preferences (e.g., Couture et al. 2024b, Tsivanidis 2024). In each case, public policy interventions, such as the construction of a new subway line along a given route, have distributional consequences across worker types through several channels. First, the differences in parameters across groups will induce different initial patterns of spatial sorting, leaving some groups more exposed to the reduction in travel times, because they live in neighborhoods geographically closer to the new subway line. Second, these differences in parameters imply that some groups may be more sensitive to the reductions in travel times from the transport improvement. Third, the change in travel times induces a general equilibrium reorganization of economic activity, including residence and workplace choices, wages and prices of residential floor space. What ultimately matters for the incidence of the policy is which groups gain improved access to their preferred locations of residence and workplace in the new equilibrium after the transport improvement.

Using a quantitative urban model with multiple groups of workers in terms of income, Tsivanidis (2024) evaluates the impact of the construction of the world's most-used Bus Rapid Transit (BRT) system - TransMilenio - in Bogotá, Colombia. Opened in the year 2000, TransMilenio handles over 2.2 million passenger journeys per day through 147 stations in its network. Lower-income workers are substantially more likely to use public transit and hence are initially more exposed to this transport improvement. However, once the geography of the commuting network and general equilibrium effects are taken into account, high-income workers ultimately end up benefiting more from TransMilenio. This finding is partly driven by the fact that the residences and workplaces of the rich are concentrated in North and Center of the city, which experience the largest travel time reductions, whereas the residence and workplace choices of the poor are more dispersed. These findings reinforce the insights to be gained from quantitative urban models relative to more basic empirical exposure measures. Whereas one might naively expect an improvement in public transit to reduce real income inequality, this policy intervention has the unintended consequence of increasing real income inequality.[34]

The role of spatial sorting in shaping the incidence of public policies is central to debates about gentrification. Couture et al. (2024b) develops a quantitative urban model with heteroge-

[34]Other quantitative evidence on the role of heterogeneity across worker groups in shaping the impact of public policy interventions includes Almagro and Domínguez-lino (2024), Lee and Tan (2024) and Warnes (2024).

neous agents and non-homothetic preferences for neighborhoods that supply non-traded amenities with endogenous quality. A secular economy-wide rise in income inequality leads to an increased demand from the rich for the high-quality amenities available in downtown neighborhoods. This rise in demand for high-quality amenities drives up house prices and further stimulates the development of higher-quality neighborhoods downtown. This gentrification of downtowns makes poor incumbents worse off, as they are either displaced to the suburbs or pay higher rents for amenities that they do not value as highly.

Using data from Dar es Salaam's Bus Rapid Transit system, Balboni et al. (2020) provides further evidence that spatial sorting shapes the impact of transport infrastructure investments. Although these interventions are place-based, in the sense that they are built in specific neighborhoods, they induce changes in patterns of spatial sorting throughout the city, and hence need not benefit the residents originally living in those neighborhoods.

6.5 Dynamics

The baseline model in Section 4 is static, such that internal city structure responds immediately to changes in productivity, amenities and the transport network. In reality, there are a number of reasons why the organization of economic activity is likely to only respond gradually to changes in these location characteristics. From this perspective, the static model could be viewed as an approximation to the steady-state of a dynamic model, around which there are transition dynamics as the economy converges towards this steady-state.

Two main sources of gradual adjustment have been emphasized in existing research.[35] First, workers may face mobility frictions in adjusting their choice of residence and workplace in response to changes in location characteristics. One approach to modelling these mobility frictions adopts a dynamic discrete choice specification. Although this specification is most commonly used for migration across regions, as in Artuç et al. (2010) and Caliendo et al. (2019), it also can be applied to residence-workplace commuting decisions, as in Warnes (2024).[36] Suppose that at the beginning of each period t, the economy inherits a mass of workers for each residence-workplace pair. After supplying labor and consuming in each period t, workers observe idiosyncratic mobility shocks for each residence-workplace pair. The value function for a worker residing in n and working in i in period t (V_{nit}) is equal to the current flow of utility from that residence-workplace pair, plus the continuation value from the optimal choice of residence-workplace pair:

$$V_{nit} = \ln u_{nit} + \max_{\{rs\}_{r=1,s=1}^{r=N,s=N}} \left\{ \beta V_{rst+1} - \mu_{ni,rs} + \rho \epsilon_{rst} \right\}, \tag{33}$$

[35]For a review of the wider literature on spatial dynamics, see Desmet and Parro (2024).

[36]An alternative specification is a Calvo model of mobility, in which agents receive moving opportunities with a constant Poisson probability, and then draw idiosyncratic preferences across locations conditional on receiving such an opportunity, as in Heblich et al. (2021) and Takeda and Yamagishi (2023).

where β is the discount rate; u_{nit} takes the same form as in equation (1); $\mu_{ni,rs}$ is the cost of moving from residence-workplace pair ni to residence-workplace pair rs; ϵ_{rst} is the idiosyncratic mobility shock; and ρ regulates the relative importance of these idiosyncratic shocks.

We make the conventional assumption that idiosyncratic mobility shocks (ϵ_{rst}) are drawn from an extreme value distribution with CDF $F(\epsilon) = e^{-e^{(-\epsilon-\overline{\gamma})}}$, where $\overline{\gamma}$ is the Euler-Mascheroni constant. Using standard properties of this extreme value distribution, the expected value of choosing residence-workplace pair ni is:

$$v_{nit} = \ln u_{nit} + \rho \ln \left(\sum_{r=1}^{N} \sum_{s=1}^{N} \exp\left(\beta v_{rst} - \mu_{ni,rs}\right)^{1/\rho} \right) \tag{34}$$

where v_{nit} is the expected value over the distribution for idiosyncratic utility. The corresponding probability that a worker moves from residence-workplace pair ni to pair rs is:

$$\xi_{ni,rs} = \frac{\exp\left(\beta v_{rst+1} - \mu_{ni,rs}\right)^{1/\rho}}{\sum_{r'=1}^{N} \sum_{s'=1}^{N} \exp\left(\beta v_{r's't+1} - \mu_{ni,r's'}\right)^{1/\rho}}. \tag{35}$$

According to this formulation, a change in productivity, amenities or the transport network that alters the distribution of expected values (v_{nit}) across residence-workplace pairs will lead to a gradual change in the network of commuters between each residence-pair. Intuitively, workers wait to incur the mobility costs ($\mu_{rs,ni}$) until they receive a favorable idiosyncratic shock (ϵ_{rst}), which induces gradual adjustment to changes in location characteristics.

Second, another source of transition dynamics is the accumulation of immobile capital in the form of building structures.[37] Two potential reasons why this immobile capital is accumulated gradually are consumption smoothing and adjustment costs. Modelling forward-looking capital investments in spatial models with population mobility is challenging, because investment and migration decisions in each location depend on one another, and on these decisions in all locations in all future time periods. Kleinman et al. (2023) develops a tractable framework that overcomes this challenge by distinguishing between workers, who make migration decisions, and landlords, who are geographically immobile and make investment decisions. Landlords have constant intertemporal elasticity of substitution preferences. The solution to their intertemporal consumption-saving decision implies that consumption and saving are linear functions of current-period wealth. Therefore, in response to shocks to location characteristics, landlords gradually accumulate or decumulate immobile capital over time. Kleinman et al. (2023) shows that the interaction between these two sources of dynamics (from worker mobility and capital accumulation) can lead to persistent and heterogeneous effects of local shocks.[38]

[37] Other research on dynamics considers endogenous innovation, as in Desmet and Rossi-Hansberg (2014), Desmet et al. (2018) and Peters (2022).

[38] For empirical evidence on local labor market shocks from globalization and technological change, see Topalova (2010), Moretti (2011), Autor et al. (2013), Autor et al. (2016) and Dix-Carneiro and Kovak (2017).

One important issue of debate about spatial dynamics is the extent to which the spatial distribution of economic activity is characterized by path dependence, in the sense that temporary shocks can have permanent effects on the spatial distribution of economic activity. Allen and Donaldson (2020) develop a framework, in which depending on parameter values, the spatial distribution of economic activity can be either (i) uniquely determined by location fundamentals; (ii) exhibit multiple steady-states, such that the location of economic activity is uniquely determined given initial conditions, but different initial conditions can lead to different steady-states; (iii) exhibit multiple equilibria, such that neither location fundamentals nor initial conditions uniquely determine the spatial distribution of economic activity. For the estimated parameter values, small and temporary shocks are found to have permanent effects on the location of economic activity and substantial impacts on welfare. Although this framework is implemented across U.S. counties, exploring path dependence at fine spatial scales within cities is an exciting area of ongoing research, as in Takeda and Yamagishi (2023).

6.6 Consumption Access

Research on quantitative urban models to date has mainly focused on residence and workplace choices within cities. However, commuting to work is only one of a large number of motives for travel within urban areas. In the 2017 National Household Transportation Survey (NHTS) for the United States, more than 67 percent of all trips by privately-owned vehicles were taken for purposes like shopping, errands and recreation.[39]

Recognizing the importance of consumption trips yields a number of insights for the determinants of the location of economic activity. First, focusing solely on commuting trips undercounts the volume of travel within urban areas, and hence underestimates the welfare gains from transport improvements. Second, incorporating consumption trips implies that location decisions within urban area are influenced not only by commuting market access, but also by consumption market access. Third, commuting and consumption travel are frequently undertaken as part of a travel itinerary, defined as a journey starting and ending at home on a given day and including more than one stop along the way. Fourth, the existence of these travel itineraries gives rise to consumption externalities, which generate rich patterns of complementarity and substitutability between locations. As one location becomes more attractive as a destination, this makes other locations that are nearby or along the way more attractive destinations (complementarity and a positive pecuniary externality), and reduces the attractiveness of other locations that are neither nearby nor along the way (substitutability and a negative pecuniary externality).

Modelling travel itineraries is computationally challenging, because agents choose both

[39] For further evidence on the importance of consumption travel, see Couture (2016), Couture et al. (2018), Agarwal et al. (2023) and Vitali (2024).

the number of destinations and the sequence in which to visit them, which results in an extremely high-dimensional state space in empirical settings with finely-detailed spatial units. Miyauchi et al. (2022) develop a quantitative urban model that overcomes this challenge of a high-dimensional state space through the use of importance sampling.[40] This method uses a Monte Carlo simulation from an auxiliary distribution and adjusts the sampling rate based on the likelihood ratio between the true distribution and the auxiliary distribution. Using this method, travel itinerary choice probabilities can be simulated, while avoiding the curse of dimensionality. The consumption externalities generated by travel itineraries are shown to be quantitatively relevant for understanding the impact of large retail store closures and the collapse in foot traffic in central cities following the shift to working from home (WFH). In the aftermath of this shift to WFH, workers commute into offices in the central city fewer days each week. In the presence of travel itineraries, this reduction in work trips leads to an accompanying decline in non-work trips in the central city, as workers no longer stop off to consume non-traded services along the way to and from work.[41] In contrast, in a conventional urban model in which all consumption occurs from home, there is no accompanying collapse in non-work trips, because workers are assumed to travel to consume non-traded services from home, regardless of where they work.

7 Empirical Applications

We now provide examples of two empirical applications of quantitative urban models. In Section 7.1, we examine the estimation of agglomeration forces. In Section 7.2, we evaluate the impact of transport infrastructure improvements. These are only two of the rich range of issues that can be addressed with quantitative urban models. Other empirical applications include urban revival (Owens III et al. 2020, Gechter and Tsivanidis 2022), congestion pricing (Herzog 2023), Bus Rapid Transit Systems (Tsivanidis 2024, Balboni et al. 2020), rail lines (Lee and Tan 2024), telecommuting (Miyauchi et al. 2022, Delventhal and Parkhomenko 2023, Monte et al. (2023)), new towns (Loumeau 2024), neighborhood effects (Redding and Sturm 2024), tourism and short-term rentals (Almagro and Domínguez-lino 2024, Fuchs et al. 2024), wartime destruction (Takeda and Yamagishi 2023), and zoning (Allen et al. 2017), among many others.

[40] For an alternative approach to modelling trip chains, in which agents do not plan travel itineraries in advance but rather solve a sequential decision problem, and in which returning home is costless, see Oh and Seo (2023).

[41] These findings are consistent with the evidence in Althoff et al. (2022) that the centers of large cities were most exposed to the shift to remote working, because they had the largest concentrations of business services jobs that can be done remotely, and the largest drops in visits to establishments in non-tradable services sectors.

7.1 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).[42] To empirically disentangle these two alternative explanations for location choices, one requires a source of exogenous variation in the surrounding concentration of economic activity.

7.1.1 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 wartime 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 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.

7.1.2 Economic Mechanisms

Within the baseline quantitative urban model from Section 4, there are a number of different economic mechanisms through which the division of Berlin affects the spatial distribution of

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

economic activity within West Berlin.[43] 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 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, expected utility conditional on choosing a residence-workplace pair is the same across all residence-workplace pairs; and there is no-arbitrage between commercial and residential land use (in the extension to allow an endogenous land use allocation).

7.1.3 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 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.

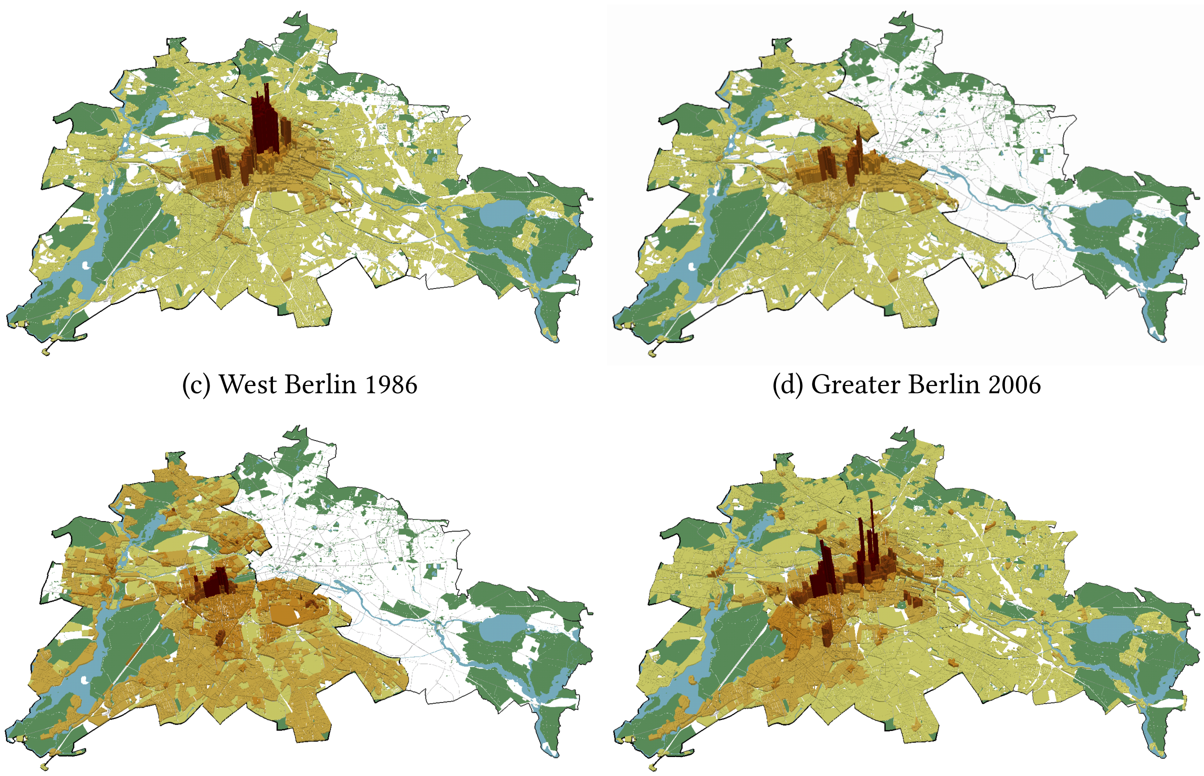
In Panel (c), we display land prices in 1986 in West Berlin. Following division, we find that the first pre-war land price peak just West of the pre-war CBD is entirely eliminated, as this area

[43] The 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.

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

(a) Greater Berlin 1936 (b) West Berlin 1936

(c) West Berlin 1986 (d) Greater Berlin 2006



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).

ceased to be an important center of commercial and retail activity. Instead, the second pre-war land 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. The biggest declines in land prices are observed 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. Additionally, there is 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.

7.1.4 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, the model can be inverted to recover 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).

7.1.5 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 discussed in Rosenthal and Strange (2004) and Combes and Gobillon (2015), 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-lino (2024). 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.

7.1.6 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.[44] Therefore, these findings provide evidence that substantial agglomeration forces are required to explain the observed concentration of economic activity within cities.

7.1.7 Summary

Taken together, the empirical evidence in this section suggests 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 it is quantitatively successful in predicting the change in the internal organization of economic activity within Berlin in response to the large-scale shock of the city's division and reunification.

7.2 Transport Infrastructure Improvements

We next 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 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.

7.2.1 The 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

[44] See Adão et al. (2024) for a formal methodology for assessing the validity of quantitative models in economics.

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.

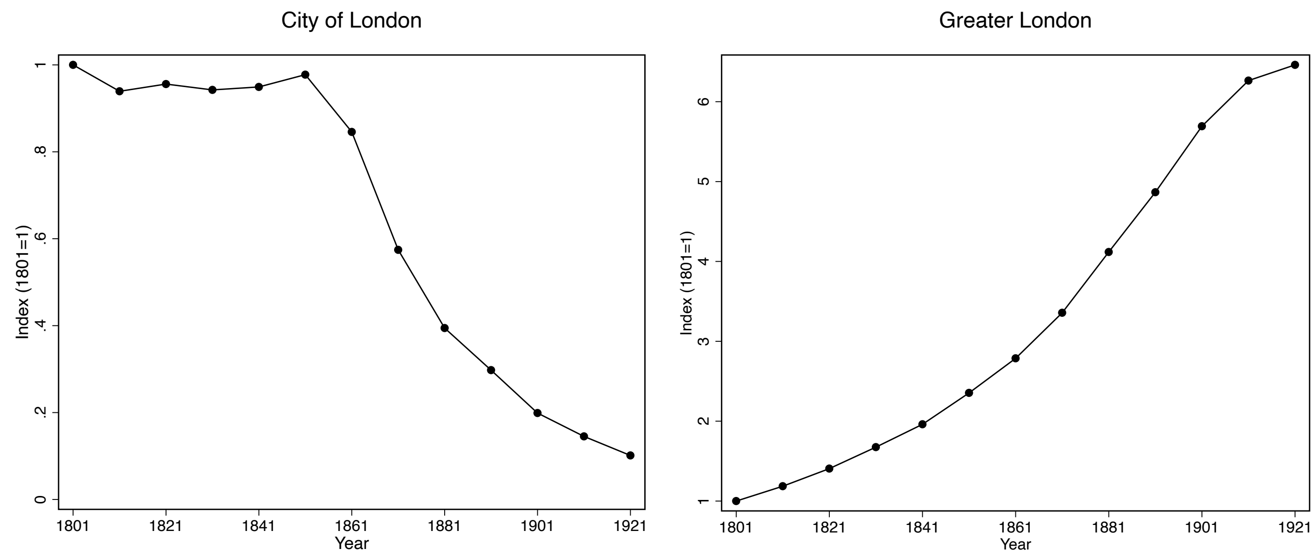
7.2.2 Economic Activity in Greater London

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 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 as 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.[45] In the decades following the first steam

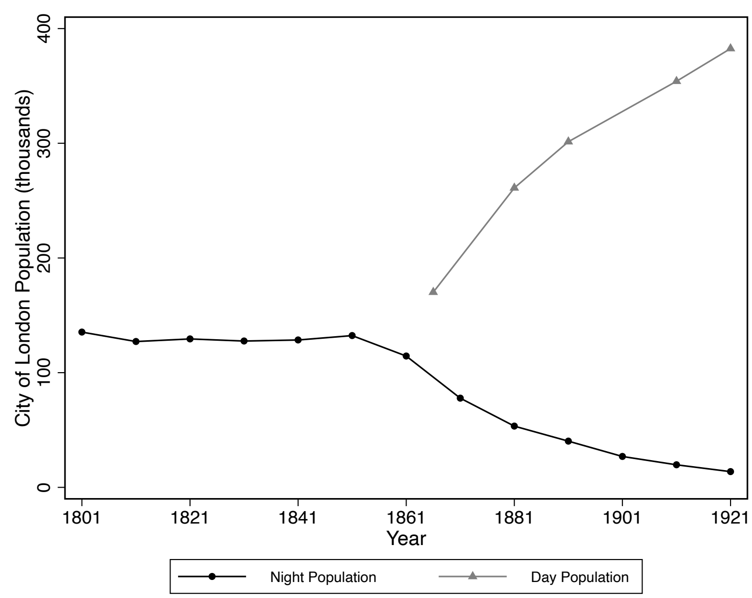
[45]Day population is defined as "every person, male or female, of all ages, residing, engaged, occupied, or employed in each and every house, warehouse, shop, manufactory, workshop, counting house, office, chambers, stable, wharf, etc ... during the working hours of the day, whether they sleep or do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

passenger railways, the sharp decline in night population is combined with a steep rise in day population, consistent with a change in the City of London's pattern of specialization as a workplace rather than as a residence.

7.2.3 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 floor space 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 impact of 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 going back to the beginning of the nineteenth century. Conditioning on the historical data for 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.

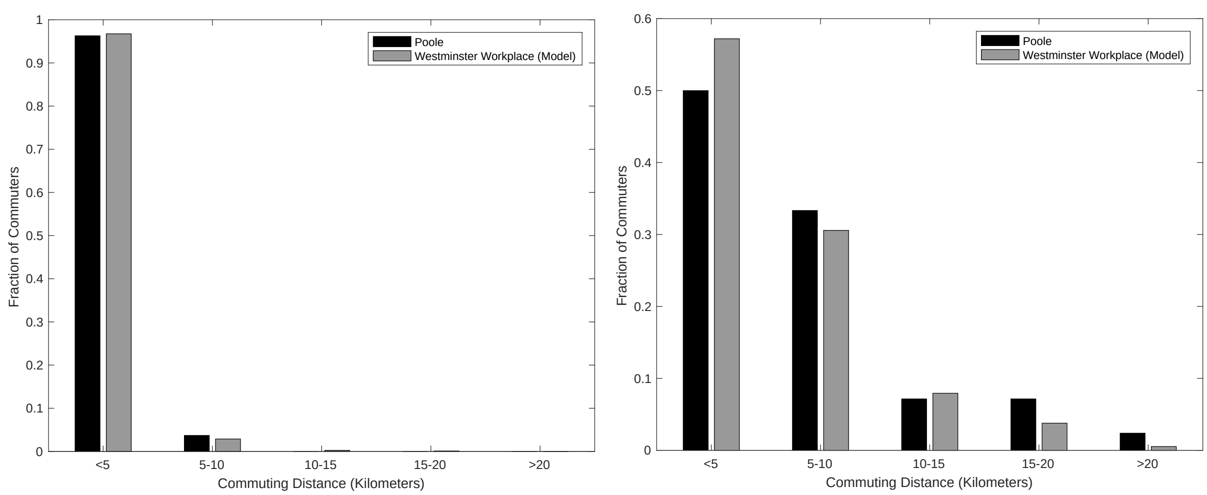
Although the impact of the railway network on commuting costs is estimated using 1921 information alone, the model successfully captures the observed sharp divergence between night-time and daytime population in the City of London from the mid-nineteenth century onwards. As the improvement in transport 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 with 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 line with a body of empirical research that finds evidence of employment (and population) decentralization in response to transport improvements, as reviewed in Redding and Turner (2015).[46]

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.[47] 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

(a) 1857-1877 (b) 1891-1911



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

[46]See also Warner (1978), Jackson (1987), Baum-Snow (2007) and Baum-Snow (2020).

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

the model's predictions and the data, including the fact that this company is located in a specific site in Westminster, whereas 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 two 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.

7.2.4 Counterfactuals

An advantage of the above empirical methodology conditioning on the 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. Table 1 reports the results of undertaking such 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 changes in transport infrastructure and changes in land use and building structures.

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	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0$	$\eta^R = 0.172$
Economic Impact				
Population	6,376,698	5,742,937	5,462,344	3,587,947
Change in Rateable Value (RV)	$-£8.24$m	$-£15.55$m	$-£20.78$m	$-£35.07$m
NPV Change in RV (3 percent)	$-£274.55$m	$-£518.26$m	$-£692.76$m	$-£1,169.05$m
NPV Change in RV (5 percent)	$-£164.73$m	$-£310.96$m	$-£415.66$m	$-£701.43$m
Construction Costs				
Cut-and-Cover Underground	$-£9.96$m			
Bored-tube Underground	$-£22.90$m			
Overground Railway	$-£33.19$m			
Total All Railways	$-£66.05$m			
Ratio Economic Impact / Construction Cost				
$\frac{\text{NPV Change in RV (3 percent)}}{\text{Construction Cost}}$	4.16	7.85	10.49	17.70
$\frac{\text{NPV Change in RV (5 percent)}}{\text{Construction Cost}}$	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.

7.2.5 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 (rateable value). 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 improvements.

7.3 Optimal Transport Infrastructure

Often policy makers are also interested in comparing alternative transport infrastructure improvements, such as how much to invest in each link in a railway or highway network. To develop a framework to address this question, Allen and Arkolakis (2022b) embed a specification of endogenous route choice in a quantitative spatial model. Agents 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).[48]

Most research that has used quantitative urban models to evaluate transport improvements has focused on evaluating the impact of a given transport improvement or comparing alternative possible transport improvements. A more challenging problem is solving for the optimal transport network.[49] This problem is extremely high-dimensional, because it involves choosing a path in geographic space, and is potentially non-convex. Fajgelbaum and Schaal (2020) show that this problem of solving for 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. A social planner chooses the optimal amount to invest in each link in the transport network, where the trade costs for each link are assumed to be increasing in the volume of traffic on that link and decreasing in the level of the investment for that link. The paper provides theoretical conditions under which this problem is globally convex, thereby guaranteeing its numerical tractability, and shows how to implement this framework using data on the European transport network.[50]

While Fajgelbaum and Schaal (2020) studies the optimal transport network for a single global planner, Bordeu (2023) considers the decentralized choice of transport infrastructure by multiple municipalities within a single urban area. These municipalities compete for residents and workers by investing in commuting infrastructure to maximize net land value in their jurisdictions. Compared to a single metropolitan planner, municipalities underinvest in areas near their boundaries and overinvest in core areas away from the boundary. As a result, infrastructure investment choice by decentralized municipalities results in higher cross-jurisdiction commuting costs, more dispersed employment, and more polycentric patterns of economic activity than with a single metropolitan planner.

7.4 Remote Working and the Future of Cities

Back in 1965, the share of full days worked from home (WFH) was less than 0.5 percent of all paid workdays in the United States. With improvements in communications technology, such as email, the internet and videoconferencing, this share had risen to around 7 percent just before the Covid-19 pandemic. During this pandemic, social distancing mandates and fear of infection

[48] Relating 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.

[49] For further discussion of optimal transport policy, and optimal spatial policy more broadly, see Fajgelbaum and Gaubert (2024).

[50] Building on this framework, Fajgelbaum et al. (2023) show how to use the observed placement of transport infrastructure together with an economic model of its impact to back out implied political preferences.

led to a large-scale shift to WFH, which has persisted in the years following the pandemic, even after widespread vaccination and the removal of social distancing mandates. In June 2023, the share of full days worked from home was 28 percent of all paid workdays in the United States, approximately four times its value in 2019.

Although there remains uncertainty about the permanence of this shift in working practices, it seems clear that hybrid working (spending some days at home and some days in the office) is likely to remain important for the foreseeable future. In many ways, this shift towards remote or hybrid working can be interpreted as a reduction in commuting costs (from travelling into the office fewer days each week), analogous to the reduction in commuting costs from new transport technologies, such as the railway and the automobile. Therefore, one might expect remote working to lead to a similar decentralization of economic activity.

Already this shift towards WFH has reshaped the organization of economic activity within cities. Ramani and Bloom (2021) provide evidence of a "donut effect," in the form of a shift in real estate demand for households and businesses away from central business districts (CBDs) and towards lower density suburbs. Gupta et al. (2022a) provide evidence of an "office real estate apocalypse," in the form of a large decline in lease revenues, occupancy and market rents in the commercial office sector in large cities. Althoff et al. (2022) find that the centers of large cities were most exposed to the shift to remote working, because they had the largest concentrations of business services jobs that can be done remotely. As business services workers ceased commuting into the center of these large cities, this led to a decline in visits to establishments in non-tradable services sectors. Delventhal and Parkhomenko (2023) find rich income distributional consequences of the shift to WFH, depending on the extent to which workers are employed in jobs where telecommuting is possible. Monte et al. (2023) provide theory and evidence that the Covid-19 pandemic induced a shift in working patterns between multiple equilibria in large versus small cities. While foot traffic has returned to pre-pandemic levels in the centers of small cities, it has stabilized at levels that are only 60 percent of pre-pandemic levels in large cities.

Looking ahead, the shift to WFH has reduced the demand for office space in central cities, which could lead to a reallocation of floor space from commercial to residential use. Given the costs of converting existing buildings to residential use, this reallocation is likely to occur gradually over time, and is subject to the constraints of local zoning and building regulations. As central cities become relatively less important as places of work, they could become relatively more important as places of consumption, connecting with our discussion of consumption access in Section 6.6 above.

8 Areas for Further Research

While the latest decade has seen a wave of research on quantitative urban models, there remain many exciting opportunities and areas for further research.

First, a limitation of most existing specifications of travel time in quantitative urban models is that they abstract from the industrial organization of the transportation sector. In reality, the travel costs incurred by agents do not only involve travel time but also depend on the prices charged by suppliers of transportation services. These prices are themselves determined in equilibrium by technology and market structure within the transportation sector. Following the emergence of new technologies such as ride-hailing, which allow the measurement of spatial mobility with an unprecedented level of detail, there has been a proliferation of research on the industrial organization of the transportation sector, including for example Buchholz (2022), Fréchette et al. (2019), Buchholz et al. (2020), Brancaccio et al. (2020) and Castillo (2024). The coming years will surely see more research on the two-way interaction between market structure within the transportation sector and the spatial distribution of economic activity.

Second, optimal transport infrastructure and the political economy of investments in transport infrastructure are areas that remain ripe for further work. In Fajgelbaum and Schaal (2020), the transport network is used for goods trade, whereas in Bordeu (2023) it is used for commuting. In reality, the transport network within large metropolitan areas is used for both of these purposes. We still know relatively little about the political economy forces shaping observed transport infrastructure investments and the explanations for the discrepancies between the observed networks and the optimal networks predicted by quantitative models.

Third, at small spatial scales in the presence of substantial agglomeration forces, multiple equilibria, hysteresis and path dependence are likely to be empirically relevant. An important area for further research is developing better approaches to characterizing the set of possible equilibria and selecting between equilibria when undertaking counterfactuals (for a recent approach see Ouazad 2024). In dynamic settings, in which there are multiple groups of agents, who are large relative to markets and behave strategically, these problems are especially challenging, as in the literature on dynamic games.

Fourth, although climate change is a global phenomenon, different locations around the world are unevenly exposed to this global shock, depending on their geographical position and initial climatic conditions. There is already a rapidly-growing literature on the uneven spatial incidence of global climate change through coastal flooding, amenities and sectoral productivity, including Costinot et al. (2016), Desmet et al. (2021), Cruz and Rossi-Hansberg (2024) and Bilal and Rossi-Hansberg (2024). Existing public policies, such as flood insurance and disaster relief, affect agents incentives to respond to this climate change. Further research is needed comparing the observed effects of climate change in the data to the predictions of our quantitative models, and evaluating alternative potential public policies.

Fifth, further research is needed on how the organization of production activity within cities has changed over time and how this affects our understanding of the microfoundations of agglomeration economies. Michaels et al. (2019) provide evidence on these changes in the tasks undertaken in urban areas over time using using verbs from occupational descriptions. 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. Consistent with substantial changes in the organization of production within urban areas over time, Autor (2019) finds large-scale 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. These changes in the organization of economic activity within cities could affect both the strength and spatial decay of agglomeration forces.

Sixth, recent years have seen a proliferation of new sources of data containing location information at a high-level of spatial resolution, including not only smartphone data (e.g., Büuchel et al. 2020, Athey et al. 2021, Miyauchi et al. 2022, Kreindler and Miyauchi 2023, Almagro et al. 2024, Couture et al. 2024a), but also credit card data (e.g., Klenow et al. 2022, Fuchs et al. 2024), ride-hailing data (Gorback 2022), Google-maps travel time data (Akbar et al. 2023, Akbar et al. 2024), barcode scanner data (Handbury and Weinstein 2015), Yelp restaurant reservations (Davis et al. 2019), short-term rental data (e.g., Almagro and Domínguez-lino 2024), and transit-fare-card data (e.g., Lee and Tan 2024).[51] We are now able to observe spatial mobility within urban areas with an unprecedented level of detail. The role of this spatial mobility to consume non-traded services in understanding the sources of agglomeration is an exciting area for further research. Models of spatial mobility to consume non-traded services have the potential to provide microfoundations for the residential externalities assumed in the specification of endogenous amenities in equation (2) in Section 4 above (see Miyauchi et al. 2022).

More generally, there remain many exciting areas for further research inspired by ongoing changes in the economy around us. 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, information and communication technologies,

[51] For further discussion of these and other new sources of data, see Abramitzky et al. (2024).

including ride hailing, online retail, autonomous vehicles and artificial intelligence?

9 Conclusions

Modern metropolitan areas have economies as large as entire countries. These cities feature complex internal structures, with individual locations specializing as either workplaces or residences, and segregation across residential neighborhoods by income, education, race and ethnicity.

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. While these traditional theories captured some general tendencies in the data, real-world cities are not well-approximated by a one-dimensional line or a perfectly symmetric circle, which limited the usefulness of these stylized settings for empirical work.

Real-world cities can exhibit either monocentric or polycentric structures, 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 sufficiently tractable as to permit an analytical characterization of their theoretical properties. With only a small number of structural parameters (elasticities) to be estimated, they lend themselves to transparent identification. As they rationalize the observed spatial distribution of economic activity within cities, they can be used to undertake counterfactuals for the impact of empirically-realistic public-policy interventions on this observed distribution.

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.

Another 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.

Cities have changed dramatically over time, from primarily market places, to the site of manufacturing industry, to clusters of office and retail development, and to centers of consumption. With the development of new transportation, information and communication technologies, such as remote working, online retail, autonomous vehicles and artificial intelligence, they are likely to continue to evolve. With the availability of new theoretical techniques and new spatially-disaggregated data, there remain many exciting opportunities for further research to understand the implications of these innovations for the organization of economic activity within cities.

References

Abramitzky, R., L. Boustan, and A. Storeygard (2024): "New Sources of Data on Cities," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 2.

Adão, R., J. S. Becko, A. Costinot, and D. Donaldson (2024): "Putting Quantitative Models to the Test: An Application to the US-China Trade War," MIT, mimeograph.

Adão, R., C. Arkolakis, and F. Esposito (2019): "Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence," *NBER Working Paper*, 25544.

Agarwal, S., B. Jensen, and F. Monte (2023): "Consumer Mobility and the Local Structure of Consumption Industries," *NBER Working Paper*, 23616.

Ahlfeldt, G., N. Baum-Snow, and R. Jedwab (2023): "The Skyscraper Revolution: Global Economic Development and Land Savings," *CEPR Discussion Paper*, 18604.

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 (2022): "The Economics of Skyscrapers: A Synthesis," *Journal of Urban Economics*, 129.

Ahlfeldt, G. M. and D. P. McMillen (2018): "Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago 1870-2010," *Review of Economics and Statistics*, 100, 861–875.

Akbar, P., V. Couture, G. Duranton, and A. Storeygard (2023): "Mobility and Congestion in Urban India," *American Economic Review*, 113, 1083–1111.

——— (2024): "The Fast, the Slow, and the Congested: Urban Transportation in Rich and Poor Countries," University of Pennsylvania, mimeograph.

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," *Journal of Economic Perspectives*, 37, 3–28.

——— (2022b): "The Welfare Effects of Transportation Infrastructure Improvements," *Review of Economic Studies*, 89, 2911–2957.

——— (2024): "Quantitative Regional Economics," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 6.

Allen, T., C. Arkolakis, and X. Li (2017): "Optimal City Structure," Yale University, mimeograph.

——— (2024a): "On the Equilibrium Properties of Spatial Models," *American Economic Review: Insights*, forthcoming.

Allen, T., D. Atkin, S. Cantillo, and C. Hernandez (2024b): "Trucks," Dartmouth College, mimeograph.

Allen, T. and D. Donaldson (2020): "Persistence and Path Dependence in the Spatial Economy," *NBER Working Paper*, 28059.

Almagro, M., F. Barbieri, J. C. Castillo, N. Hickok, and T. Salz (2024): "Optimal Urban Transportation Policy: Evidence from Chicago," *NBER Working Paper*, 32185.

Almagro, M. and T. Domínguez-lino (2024): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *NBER Working Paper*, 32304.

Alonso, W. (1964): *Location and Land Use*, Cambridge MA: Harvard.

Althoff, L., F. Eckert, S. Ganapati, and C. Walsh (2022): "The Geography of Remote Work," *Regional Science and Urban Economics*, 93, 103770.

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

Artuç, E., S. Chaudhuri, and J. McLaren (2010): "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 100, 1008–45.

Arzaghi, M. and J. V. Henderson (2008): "Networking Off Madison Avenue," *Review of Economic Studies*, 75, 1011–1038.

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt (2021): "Estimating Experienced Racial Segregation in US Cities Using Large-Scale GPS Data," *Proceedings of the National Academy of Science*, 118, e2026160118.

Autor, D. (2019): "Work of the Past, Work of the Future," *American Economic Review*, 109, 1–32, richard Ely Lecture.

Autor, D., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–68.

Autor, D. H., D. Dorn, and G. H. Hanson (2016): "The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade," *Annual Review of Economics*, 8, 205–240.

Balboni, C., G. Bryan, M. Morten, and B. Siddiqi (2020): "Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies," London School of Economics, mimeograph.

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.

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.

Baum-Snow, N. (2007): "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122, 775–780.

Baum-Snow, N. (2020): "Urban Transport Expansions and Changes in the Spatial Structure of US Cities: Implications for Productivity and Welfare," *Review of Economics and Statistics*, 102, 929–945.

Baum Snow, N. and G. Duranton (2024): "Housing Supply and Affordability," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 6.

Baum-Snow, N. and L. Han (2024): "The Microgeography of Housing Supply," *Journal of Political Economy*, 132, 1897–1946.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2016): "A Dynamic Model of Demand for Houses and Neighborhoods," *Econometrica*, 84, 893–942.

Bilal, A. and A. Rossi-Hansberg (2024): "Anticipating Climate Change Across the United States," .

Bleakley, H. and J. Lin (2012): "Portage: Path Dependence and Increasing Returns in U.S. History," *Quarterly Journal of Economics*, 127, 587–644.

Bordeu, O. (2023): "Commuting Infrastructure in Fragmented Cities," Chicago Booth School of Business.

Borusyak, K., R. Dix-Carneiro, and B. Kovak (2024): "Understanding Migration Reponses to Local Shocks," University of California Berkeley, mimeograph.

Brancaccio, G., M. Kaloupsidi, and T. Papageorgiou (2020): "Geography, Transportation and Endogenous Trade Costs," *Econometrica*, 88, 657–691.

Brueckner, J. K. (1987): "The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model," in *Handbook of Regional and Urban Economics*, ed. by E. S. Mills, Amsterdam: Elsevier North Holland, vol. 2, chap. 20, 821–845.

Buchholz, N. (2022): "Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry," *Review of Economic Studies*, 89, 556–591.

Buchholz, N., L. Doval, J. Kastl, F. Matejka, and T. Salz (2020): "Personalized Pricing and The Value of Time: Evidence from Auctioned Cab Rides," Princeton University, mimeo.

Büuchel, K., M. V. Ehrlich, D. Puga, and E. Viladecans (2020): "Calling from the Outside: The Role of Networks in Residential Mobility," *Journal of Urban Economics*, 119, 103277.

Caliendo, L., M. Dvorkin, and F. Parro (2019): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87, 741–835.

Castillo, J. C. (2024): "Who Benefits from Surge Pricing?" *Econometrica*, forthcoming.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–87.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106, 855–902.

Combes, P.-P., G. Duranton, and L. Gobillon (2021): "The Production Function for Housing: Evidence from France," *Journal of Political Economy*, 129, 2766–2816.

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

COSTINOT, A., D. DONALDSON, AND C. SMITH (2016): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, 124, 205–248.

COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2014): "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by E. Helpman, G. Gopinath, and K. Rogoff, Amsterdam: North-Holland, vol. 4, chap. 4, 197–261.

COUTURE, V. (2016): "Valuing the Consumption Benefits of Urban Density," Vancouver School of Economics, mimeograph.

COUTURE, V., J. DINGEL, A. GREEN, AND J. HANDBURY (2024a): "Demographic Preferences and Income Segregation," Wharton, mimeo.

COUTURE, V., G. DURANTON, AND M. A. TURNER (2018): "Speed," *Review of Economics and Statistics*, 100, 725–739.

COUTURE, V., C. GAUBERT, J. HANDBURY, AND E. HURST (2024b): "Income Growth and the Distributional Effects of Urban Spatial Sorting," *Review of Economic Studies*, 91, 858–898.

CRUZ, J.-L. AND E. ROSSI-HANSBERG (2024): "The Economic Geography of Global Warming," *Review of Economic Studies*, 91, 899–939.

DAVIS, D. R., J. DINGEL, J. MONRAS, AND E. MORALES (2019): "How Segregated is Urban Consumption?" *Journal of Political Economy*, 127, 1684–1738.

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

——— (2008): "A Search for Multiple Equilibria in Urban Industrial Structure," *Journal of Regional Science*, 48, 29–65.

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

DELVENTHAL, M. J. AND A. PARKHOMENKO (2023): "Spatial Implications of Telecommuting," University of Southern California, mimeo.

DESMET, K., R. E. KOPP, S. A. KULP, D. K. NAGY, M. OPPENHEIMER, E. ROSSI-HANSBERG, AND B. H. STRAUSS (2021): "Evaluating the Economic Cost of Coastal Flooding," *American Economic Journal: Macroeconomics*, 13, 444–86.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018): "The Geography of Development," *Journal of Political Economy*, 126, 903–983.

Desmet, K. and F. Parro (2024): "Spatial Dynamics," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 6.

Desmet, K. and E. Rossi-Hansberg (2014): "Spatial Development," *American Economic Review*, 104, 1211–1243.

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

Diamond, R. and J. C. Suarez Serrato (2024): "Spatial Sorting and Inequality," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 2.

Dingel, J. I. and F. Tintelnot (2020): "Spatial Economics for Granular Settings," *NBER Working Paper*, 27287.

Dix-Carneiro, R. and B. K. Kovak (2017): "Trade Liberalization and Regional Dynamics," *American Economic Review*, 107, 1908–2946.

Donaldson, D. (2015): "The Gains from Market Integration," *Annual Review of Economics*, 7, 619–647.

——— (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

——— (2022): "Blending Theory and Data: A Space Odyssey," *Journal of Economic Perspectives*, 36, 185–210.

——— (2024): "Transportation," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier, vol. 6.

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

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.

——— (2015): "Urban Land Use," in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. C. Strange, Amsterdam: Elsevier North Holland, vol. 5, chap. 8, 467–560.

——— (2020): "The Economics of Urban Density," *Journal of Economic Perspectives*, 34, 3–26.

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

Eaton, J. and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

Eaton, J., S. Kortum, and S. Sotelo (2013): "International Trade: Linking Micro and Macro," in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, Cambridge: Cambridge University Press.

Epple, D., B. Gordon, and H. Sieg (2010): "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 100, 905–924.

Fajgelbaum, P. and C. Gaubert (2020): "Optimal Spatial Policies, Geography, and Sorting," *Quarterly Journal of Economics*, 135, 959–1036.

——— (2024): "Optimal Spatial Policies," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Amsterdam: Elsevier North Holland, vol. 2.

Fajgelbaum, P., C. Gaubert, N. Gorton, E. Morales, and E. Schaal (2023): "Political Preferences and Transport Infrastructure: Evidence from California's High-Speed Rail," *NBER Working Paper*, 31438.

Fajgelbaum, P. D. and E. Schaal (2020): "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 88, 1411–1452.

Fogli, A. and V. Guerrieri (2019): "The End of the American Dream? Inequality and Segregation in US Cities," Chicago Booth School of Business, mimeograph.

Fortheringham, S. and M. O'Kelly (1989): *Spatial Interaction Models: Formulations and Applications*, Dordrecht: Kluwer.

Fréchette, G. R., A. Lizzeri, and T. Salz (2019): "Frictions in a Competitive, Regulated Market: Evidence From Taxis," *American Economic Review*, 109, 2954–2992.

Fuchs, S., S. Ganapati, A. Graziano, R. Madera, and J. Montoriol-Garriga (2024): "Urban Welfare: Tourism in Barcelona," Dartmouth College, mimeograph.

Fujita, M. and P. Krugman (1995): "When is the Economy Monocentric? Von Thünen and Chamberlain Unified," *Regional Science and Urban Economics*, 25, 505–528.

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge MA: MIT Press.

Fujita, M. and H. Ogawa (1982): "Multiple Equilibria and Structural Transformation of Non-monocentric Urban Configurations," *Regional Science and Urban Economics*, 12, 161–196.

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

Gechter, M. and N. Tsivanidis (2022): "Spatial Spillovers from Urban Redevelopments: Evidence from Mumbai's Textile Mills," University of California, Berkeley.

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

Glaeser, E. (2008): *Cities, Agglomeration and Spatial Equilibrium*, Oxford: Oxford University Press.

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. and J. Gyourko (2005): "Urban Decline and Durable Housing," *Journal of Political Economy*, 113, 345–375.

Glaeser, E. L., J. Gyourko, and R. Saks (2005): "Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices," *Journal of Law and Economics*, 48, 331–369.

Glaeser, E. L., J. Kolko, and A. Saiz (2001): "Consumer City," *Journal of Economic Geography*, 1, 27–50.

Gorback, C. (2022): "Ridesharing and the Redistribution of Economic Activity," University of Texas at Austin, mimeo.

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.

Guimarães, P., O. Figueirdo, and D. Woodward (2003): "A Tractable Approach to the Firm Location Decision Problem," *Review of Economics and Statistics*, 85, 201–204.

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.

Handbury, J. and D. E. Weinstein (2015): "Goods Prices and Availability in Cities," *Review of Economic Studies*, 82, 258–296.

Hanson, G. H. (2005): "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economics*, 67, 1–24.

Harari, M. (2020): "Cities in Bad Shape: Urban Geometry in India," *American Economic Review*, 10, 2377–2421.

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.

Heblich, S., A. Trew, and Y. Zylberberg (2021): "East-Side Story: Historical Pollution and Persistent Neighborhood Sorting," *Journal of Political Economy*, 129, 1319–1666.

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.

Herzog, I. (2023): "The City-Wide Effects of Tolling Downtown Drivers: Evidence from London's Congestion Charge," University of Toronto, mimeograph.

Hornbeck, R. and E. Moretti (2024): "Estimating Who Benefits From Productivity Growth: Local and Distant Effects of City Productivity Growth on Wages, Rents, and Inequality," *Review of Economics and Statistics*, 106, 587–607.

Hulten, C. R. (1978): "Growth Accounting with Intermediate Inputs," *Review of Economic Studies*, 511–518.

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

Kleinman, B., E. Liu, and S. Redding (2023): "Dynamic Spatial General Equilibrium," *Econometrica*, 91, 385–424.

Klenow, P., L. Einav, J. Levin, and R. Murciano-Goroff (2022): "Customers and Retail Growth," Stanford University, mimeograph.

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.

Kreindler, G. E. (2024): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, forthcoming.

Kreindler, G. E. and Y. Miyauchi (2023): "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records," *Review of Economics and Statistics*, 105, 899–909.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

——— (1993): "First Nature, Second Nature, and Metropolitan Location," *Journal of Regional Science*, 33, 129–144.

Lee, K. H. and B. J. Tan (2024): "Urban Transit Infrastructure and Inequality," *Review of Economics and Statistics*, forthcoming.

Lin, J. and F. Rauch (2022): "What Future for History Dependence in Spatial Economics?" *Regional Science and Urban Economics*, 94, 103628.

Loumeau, G. (2024): "Accommodating the Rise in Urbanization: Are New Towns a Good Solution?" *Economic Journal*, forthcoming.

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.

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.

McMillen, D. P. (1996): "One Hundred and Fifty Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics*, 40, 100–124.

Michaels, G. and F. Rauch (2018): "Resetting the Urban Network: 117-2012," *Economic Journal*, 128, 378–412.

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., C. Porcher, and E. Rossi-Hansberg (2023): "Remote Work and City Structure," *NBER Working Paper*, 31494.

Monte, F., S. Redding, and E. Rossi-Hansberg (2018): "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 108, 3855–3890.

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

Muth, R. (1969): *Cities and Housing*, Chicago: University of Chicago Press.

Oh, R. and J. Seo (2023): "What Causes Agglomeration of Services? Theory and Evidence from Seoul," Yale University, mimeograph.

Ouazad, A. C.-L. (2024): "Equilibrium Multiplicity: A Systematic Approach using Homotopies, with an Application to Chicago," *SSRN Working Paper*, http://dx.doi.org/10.2139/ssrn.4699361.

Owens III, R., E. Rossi-Hansberg, and P.-D. Sarte (2020): "Rethinking Detroit," *American Economic Journal*, 12, 258–305.

Peters, M. (2022): "Market Size and Spatial Growth: Evidence from Germany's Post-war Population Expulsions," *Econometrica*, 90, 2357–2396.

Ramani, A. and N. Bloom (2021): “The Donut Effect of Covid-19 on Cities,” *NBER Working Paper*, 28876.

Redding, S. J. (2010): “The Empirics of New Economic Geography,” *Journal of Regional Science*, 50, 297–311.

——— (2011): “Economic Geography: a Review of the Theoretical and Empirical Literature,” in *Palgrave Handbook of International Trade*, ed. by D. Bernhofen, R. Falvey, D. Greenaway, and U. Kreickemeier, London: Palgrave Macmillan.

——— (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.

——— (2023): “Quantitative Urban Models: From Theory to Data,” *Journal of Economic Perspectives*, 37, 75–98.

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.

——— (2024): “Neighborhood Effects: Evidence from the Streets of London,” *NBER Working Paper*, 32333.

Redding, S. J., D. M. Sturm, and N. Wolf (2011): “History and Industry Location: Evidence from German Airports,” *Review of Economics and Statistics*, 93, 814–831.

Redding, S. J. and M. A. Turner (2015): “Transportation Costs and the Spatial Organization of Economic Activity,” in *Handbook of Regional and Urban Economics*, ed. by G. Duration, J. V. Henderson, and W. Strange, Amsterdam: Elsevier, chap. 20, 1339–1398.

Redding, S. J. and A. J. Venables (2004): “Economic Geography and International Inequality,” *Journal of International Economics*, 62, 53–82.

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. (2004): "Optimal Urban Land Use and Zoning," *Review of Economic Dynamics*, 7, 69–106.

Rossi-Hansberg, E., P.-D. Sarte, and R. I. Owens (2010): "Housing Externalities," *Journal of Political Economy*, 118, 485–535.

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

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

Santos Silva, J. M. C. and S. Tenreyro (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88, 641–658.

Severen, C. (2023): "Commuting, Labor and Housing Effects of Mass Transportation: Welfare and Identification," *Review of Economics and Statistics*, 105, 1073–1091.

Shoven, J. B. and J. Whalley (1992): *Applying General Equilibrium*, Cambridge: Cambridge University Press.

Takeda, K. and A. Yamagishi (2023): "The Economic Dynamics of City Structure: Evidence from Hiroshima's Recovery," National University of Singapore.

Topalova, P. (2010): "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2, 1–41.

Tsivanidis, N. (2024): "The Aggregate And Distributional Effects Of Urban Transit Infrastructure: Evidence From Bogotá's TransMilenio," *American Economic Review*, forthcoming.

Vitali, A. (2024): "Consumer Search and Firm Location: Theory and Evidence from the Garment Sector in Uganda," Dartmouth College, mimeo.

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

Warnes, P. E. (2024): "Transport Infrastructure Improvements and Spatial Sorting: Evidence from Buenos Aires," Aalto University, mimeograph.

Weiwu, L. (2024): "Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities," MIT, mimeograph.