Journal of Economic Perspectives—Volume 37, Number 2—Spring 2023—Pages 75–98

Quantitative Urban Models: From Theory to Data

Stephen J. Redding

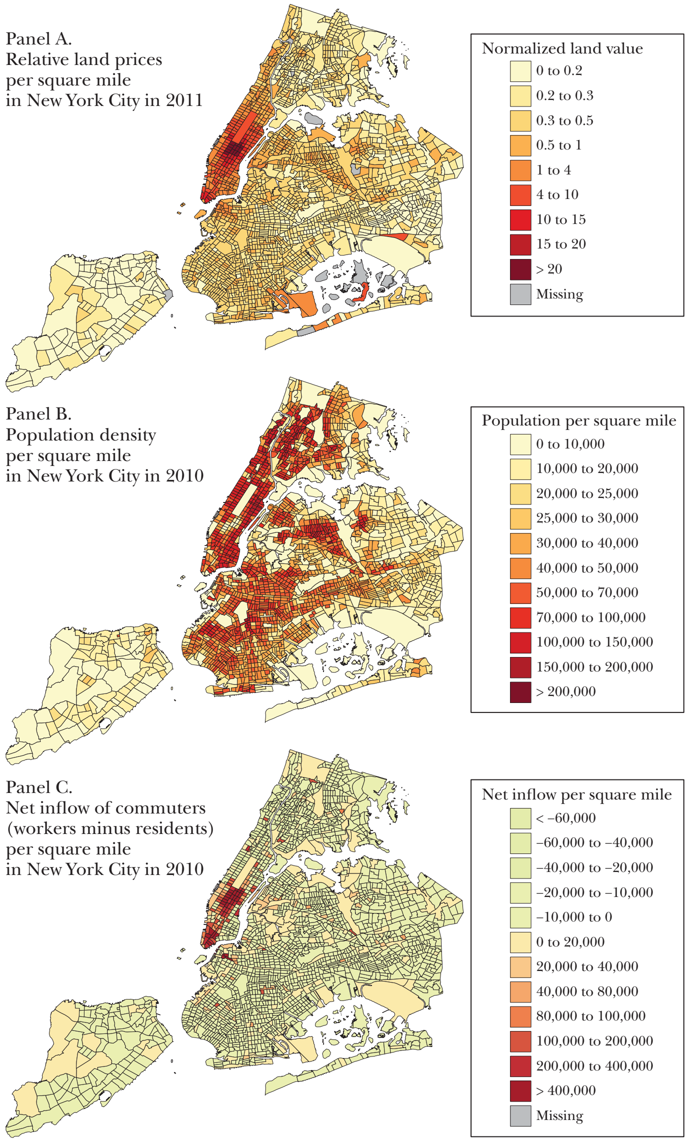
One of the most striking economic features of our world is the uneven distribution of economic activity across space. This concentration is evident in the existence of cities. By 2018, 55 percent of the world's population lived in urban areas, with one in eight urbanites residing in 33 megacities with more than ten million inhabitants (United Nations 2019). But similar concentration is observed within cities as well. 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. As a pedestrian walks from one city block to another, land use can change sharply from residential to commercial land use and back again.

The three panels of Figure 1 illustrate this within-city variation by using census tract data across the five boroughs of New York City. The borough of Staten Island is in the lower left of the figure. The land directly east (across the Hudson River) includes Brooklyn to the west and Queens to the east. The island further north (between the Hudson River and the East River) includes Manhattan to the south and the Bronx to the north. Census tracts are intended to include about 4,000 people, although for localized reasons they can be half or twice that size.

■ *Stephen J. Redding is Professor in Economics, Princeton University, Princeton, New Jersey. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Fellow, Centre for Economic Policy Research, London, United Kingdom. His email address is reddings@princeton.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.2.75.

Figure 1
Economic Geography of New York City



Source: Panel A: Primary Land Use Tax Lot Output, NYC Department of City Planning. Panel B: US population census. Panel C: Data on commuting flows from the LEHD Origin-Destination Employment Statistics (LODES).

Note: Panel A: Assessed land values per square mile in 2011, normalized by the mean across census tracts in New York City, including the five boroughs of Bronx, Brooklyn, Manhattan, Queens, and Staten Island. A value above one corresponds to an above average land value per square mile. Land values are from property taxation assessments. Panel B: 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). Panel C: 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.

Panel A shows the variation in land prices across New York City, measured using the assessed tax value of the land per square meter in 2011.[1] Midtown Manhattan has by far the highest land prices, with a smaller secondary peak in downtown Manhattan, and an area of lower land values in between. Towards the bottom of Manhattan, the Lower East Side has noticeably lower land prices than other nearby neighborhoods. Across the East River, central Brooklyn is the site of another smaller peak in land prices. Finally, the areas bordering Central Park in Manhattan are relatively more expensive than those further away from the park. The higher land prices in some neighborhoods relative to others can be explained in terms of the demand for either commercial or residential land use. In 2010, the top 10 percent of census tracts in New York City with the highest land value per square mile accounted for 60 percent of total land value and 64 percent of employment, but only 14 percent of population and 7 percent of land area.[2]

Panel B of Figure 1 displays the number of residents per square mile in 2010 for each census tract in New York City. Manhattan is the most densely-populated county in the United States: In 2010, 1,518,500 people lived in an area of 22.8 square miles, with a population density of 66,579 people per square mile. In contrast, Staten Island is relatively sparsely populated, with a population density an order of magnitude smaller at 7,923 people per square mile.[3] Even within Manhattan, population density displays dramatic variation, with relatively low densities in the commercial districts of midtown and downtown, and relatively high densities in the residential suburbs of the Upper West and Upper East Sides.

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

Location specialization is even more dramatic by this measure. The two land price peaks of midtown and downtown Manhattan are highly-specialized commercial

[1] Similar patterns are observed using land prices estimated from property transactions data. For evidence from land prices estimated using property transactions data, see for example Barr, Smith, and Kulkarni (2018) and Haughwout, Orr, and Bedoll (2008).

[2] These authors' estimates are from the Primary Land Use Tax Lot Output, NYC Department of City Planning, the US population census, and LEHD Origin-Destination Employment Statistics (LODES).

[3] Even so, Staten Island is densely-populated relative to many rural locations in the United States, with the state of Wyoming having a population density in 2010 of 5.6 people per square mile.

districts, with net imports of commuters greater than 400,000 per square mile.[4] Although Manhattan as a whole is a net importer of commuters, we find that parts of the island specialize as residences, with high exports of commuters per square mile in some areas on the Upper West and Upper East Side. Additionally, Brooklyn and the Bronx also have some census tracts with high imports of commuters per square mile. 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 New York City and the tri-state area (with an inflow of 1.6 million and an outflow of 0.19 million). Moreover, this specialization occurs at the macro scale of New York City and its economic hinterland across the states of Connecticut, New Jersey, and New York: In 2010, New York City was a net importer of 0.5 million workers from the rest of this tri-state area (with an inflow of 0.97 million and an outflow of 0.46 million).

These rich patterns of the concentration of economic activity can be explained by a three-way interaction between natural advantages, agglomeration forces, and dispersion forces. In the context of New York City, the waterfront areas around the edge of Manhattan historically had *natural advantages* for production, 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. Discussions of *agglomeration* typically feature externalities, such that an agent making a location decision does not take into account how that decision will affect the location decisions of other agents. These externalities can be either technological (say, knowledge spillovers between agents) or mediated through markets (say, demand for locally-traded goods and services). These agglomeration forces promote the concentration of economic activity, but are offset by *dispersion* forces. For example, when the price of local factors of production that are in inelastic supply are bid higher, such as land, incentives arise that shift production or residential activity to areas with lower land prices. More broadly, the concentration of economic activity can give rise to congestion or facilitate the spread of disease between people, both of which can act as dispersion forces.

The complexity of modeling these forces in spatial equilibrium has meant that the traditional theoretical literature on cities focused on stylized settings, such as a monocentric city with one central business district, a one-dimensional city on a line, or a perfectly symmetric circular city. I begin with a brief review of these earlier models, but as New York City and many other cities readily illustrate, such models cannot capture the rich internal variation in patterns of economic activity within real world cities, nor can they be easily used for detailed analysis of events or policies affecting a specific city.

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

The main focus of this paper is to describe the recent development of quantitative urban models that connect directly with observed data on real world cities. These models allow for many locations within a city that can differ in productivity, amenities, land area, the supply of floor space, and transport connections. This heterogeneity across locations reflects the impact of natural advantages, agglomeration forces, and dispersion forces. Given the richness and flexibility of these quantitative urban models, they have been used to analyze a host of issues in urban economics: the strength of agglomeration forces, zoning and building regulations, the impact of transport infrastructure improvements, the sorting of heterogeneous groups of workers across space, and congestion pricing, among many others. These frameworks are sufficiently tractable that they permit a mathematical analysis of their properties, such as the conditions under which there is a unique equilibrium versus multiple equilibria in the model. In the presence of multiple equilibria, even small public policy interventions can have substantial effects, by shifting the location of economic activity between multiple equilibria. These theories also can be used to examine the effects of exogenous shocks affecting a city, like the division of Berlin by the Berlin Wall or the invention of a new mass-transit technology.

In the aftermath of the Covid-19 pandemic, these models can also provide insights as urban areas react to a range of issues; for example, concerns about disease, the practice of social distancing, a rise in remote work, changes in public transit systems, and further innovations in transport technology, such as ride-hailing and -sharing and autonomous vehicles.

Traditional Theoretical Models of Urban Economics

Traditional theoretical models in urban economics are focused on explaining stylized features of the data, such as the existence of a land price gradient, in which land prices are typically higher in the city center and on average decline with distance from the city center. Often these traditional models assume that economic activity is *monocentric*, in the sense that there is a well-defined central business district with a single peak of land prices. They may also consider a restrictive geography, such as identical locations along the real line or a perfectly symmetric circular city.

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

Many cities with long histories of settlement (like London) are well approximated by this assumption of a monocentric pattern of economic activity. In contrast, other cities that developed more recently (like Los Angeles) are better described by a polycentric structure, in which there are multiple business districts spread throughout the metropolitan area. One polycentric structure is an "edge city," which consists of multiple concentrations of business, shopping, and entertainment outside a traditional downtown or central business district, often beside a major road in what had previously been a suburban residential or rural area.

To allow for the possibility of polycentricity, the assumption that all employment is concentrated in the city center can be relaxed to allow for an endogenous allocation of land between commercial and residential use throughout the city. Fujita and Ogawa (1982) consider the case of a one-dimensional city along the real line, while Lucas and Rossi-Hansberg (2002) analyze a perfectly symmetric circular city. In these frameworks, whether monocentric or polycentric patterns of economic activity emerge depends on the strength of agglomeration and dispersion forces. On the one hand, a nonmonocentric pattern of alternating areas of commercial and residential land use reduces commuting costs, because workers can 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 summary, key insights from this theoretical literature are the role of the trade-off between agglomeration forces and commuting costs in generating urban rent gradients, and in determining whether these rent gradients are monocentric or polycentric.

Quantitative Urban Models

Although traditional models in urban economics explain certain features of the data, their simplifying assumptions of monocentricity or symmetry limit their usefulness for empirical work. These simplifying assumptions abstract from empirically relevant differences in natural advantage across locations, such as access to natural harbors or green parks. No city in the real world is perfectly monocentric or symmetric.

To address these limitations, recent quantitative urban models allow for empirically relevant differences in natural advantage while also incorporating agglomeration forces. These models are designed to connect directly to observed data on cities, which feature rich asymmetric patterns of economic activity—say, higher land prices in western than in eastern suburbs—and scattered clusters of employment and residents throughout a given city. Because these models connect directly to the observed data, they can be used to estimate the strength of agglomeration forces, or to undertake counterfactuals to predict the impact of realistic public policy interventions, such as the construction of a new subway line along a specific route within a given city.

We begin by describing a baseline quantitative version of the canonical urban model following Ahlfeldt et al. (2015), before discussing a number of extensions and generalizations.[5] Redding and Rossi-Hansberg (2017) review quantitative spatial models more broadly and Redding (2022a) surveys the wider literature on trade and geography.

Building Blocks

Consider a city, embedded in a wider economy. The city consists of a set of discrete "blocks" or census tracts. Each block has a supply of floor space that depends on its geographical land area and the density of development (the ratio of floor space to land area). Floor space can be used either commercially or residentially, or with some mixture of the two, a choice which will be influenced by zoning regulations.

The city is populated by workers, who are mobile between the city and the larger economy. Workers first decide whether to move to the city, and if so, they then consider each possible pair of residence and workplace blocks within the city. Workers have idiosyncratic preferences for living and working in different locations within the city. They consider all the personal, work-related, or amenity-related reasons for living in one place and working in another, and pick the residence-workplace pair that yields the highest utility. Commuting costs increase with the travel time between the worker's residence and workplace. Residential amenities depend on both natural advantages, such as leafy streets and scenic views, and agglomeration forces in the form of residential externalities, including positive externalities from nontraded goods and negative externalities from crime.

Because the model is focused on location choices within the city, it assumes away different kinds of final goods. Instead it assumes a single final good that is costlessly traded both within the city and with the wider economy, within perfectly competitive markets. This final good is produced using inputs of labor and commercial floor space according to a constant returns to scale technology. However, productivity of the final good can differ across locations within the city and depends both on natural advantages, such as access to natural water, and on agglomeration forces in the form of production externalities, which depend on employment density in surrounding locations (and are influenced by knowledge spillovers).

Economic Forces

We now discuss the three sets of economic forces that shape the equilibrium organization of economic activity within the city: productivity differences across locations, amenity differences across locations, and the transportation network.

High productivity in a location raises the marginal productivities of labor and land, which increases wages and the price of commercial floor space. In contrast, high amenities in a location raise the utility of living there, which attracts residents,

[5] An accompanying online Appendix provides a more detailed development of the model and a formal characterization of its theoretical properties.

and bids up the price of residential floor space. Transportation networks allow workers to separate where they live from where they work to take advantage of these differences in productivity and amenities. Through this separation of home and work, some locations specialize as workplaces (often but not always in central cities), while other locations specialize as residences (often but not always in outlying suburbs).

In locations with high productivity relative to amenities, the return to commercial land use exceeds the return to residential land use. Therefore, these locations specialize as workplaces, with higher employment than residents and net imports of commuters. In contrast, in locations with high amenities relative to productivity, the converse is true. The return to residential land use exceeds the return to commercial land use, such that these locations specialize as residences, with lower employment than residents and net exports of commuters. If a location has both positive employment and positive residents, either the return to commercial land use equals the return to residential land use, or zoning regulations sustain a wedge between the returns to commercial and residential land uses.

The resulting commuting flows between locations are typically assumed to satisfy a *gravity equation*, so-called because of the parallel with the Newtonian theory of gravity. According to this specification, the bilateral flow of commuters between a residence and workplace is decreasing in bilateral commuting costs, increasing in the attractiveness of the workplace (for example, as captured by its wage), and increasing in the attractiveness of the residence (for example, as shaped by its amenities). The attractiveness of each residence also depends on its overall commuting costs to all workplaces, often referred to as "multilateral resistance." Even if a specific workplace-residence pair has high bilateral commuting costs (high *bilateral resistance*), we may still observe substantial bilateral commuting flows if that residence has even higher commuting costs for all other workplaces (high *multilateral resistance*). This gravity equation specification is both theoretically tractable and provides a good approximation to observed commuting patterns (for example, as in Fotheringham and O'Kelly 1989; McDonald and McMillen 2010). Bilateral commuting costs depend on bilateral travel times, which can be computed using the observed transport network (say, underground and overground railway lines) and assumptions about the average speed of each mode of transport.[6]

The differences in productivity and amenities that induce location specialization and commuting reflect the combined impact of natural advantages and agglomeration forces. As employment concentrates in a location because of natural advantages for production, this raises employment density, which further increases productivity through production externalities, thereby magnifying the impact of differences in natural advantage. Similarly, most empirical studies find net residential externalities to be positive. Therefore, as residents concentrate in a location because of natural advantages for amenities, this further increases amenities

[6] A long line of research in transportation economics following McFadden (1974) models individuals' choice of transport mode.

through residential externalities, again magnifying the impact of differences in natural advantage.

These production and residential agglomeration forces are offset by dispersion forces from commercial and residential floor space use. As employment concentrates in a location, this bids up the price of commercial floor space, which raises firms' cost, and encourages the dispersion of employment to lower density locations. Similarly, as residents concentrate in a location, this bids up the price of residential floor space, which reduces worker utility and encourages the dispersion of residents to lower density locations. Although the overall demand for floor space in a location can be reduced by separating workplace and residence, this gives rise to commuting costs, which themselves provide another force for the dispersion of economic activity.

The strength of this dispersion force from a limited supply of floor space depends on assumptions about the supply elasticity for floor space. Floor space is typically assumed to be produced by a competitive construction sector that uses land and capital as inputs. Land itself is in perfectly inelastic supply. In contrast, for a city that is small relative to the wider economy, capital is in perfectly elastic supply at an exogenous cost of capital. As a result, the smaller is the share of land in construction costs, the larger is the supply elasticity for floor space, and the weaker are dispersion forces.

If production and residential agglomeration forces are sufficiently strong relative to these dispersion forces, the spatial organization of economic activity within the city can be subject to multiple equilibria. If employment is expected to concentrate in some locations and residents are expected to concentrate in other locations, this itself generates differences in productivity and amenities that can support such specialization as an equilibrium outcome. At small spatial scales within cities where natural advantages are similar, it is particularly plausible that the location of economic activity could be subject to such multiple equilibria, especially for individual economic functions (for example, whether shoe stores are clustered on one street rather than a neighboring street).

The general equilibrium of the model satisfies the following equilibrium conditions: (1) cost minimization and zero profits in production; (2) utility maximization and population mobility; (3) cost minimization and zero profits in construction; (4) demand for commercial floor space equals the supply of commercial floor space; (5) demand for residential floor space equals the supply of residential floor space; (6) no-arbitrage between alternative uses of floor space, such that locations are either specialized as workplaces, specialized as residences, or incompletely specialized, depending on the relative values of the prices of commercial and residential floor space; (7) employment in each workplace is equal to the number of residents choosing to commute to that workplace.

From the zero-profit condition in production, firms must make zero profits in all locations with positive employment. Therefore, high productivity in a location must be offset in equilibrium by some combination of higher wages and/or a higher price of commercial floor space. Given data on wages and the price of commercial floor space, and an assumption about the production technology, it follows that one can use this zero-profit condition to back out the productivity required for the observed data to be an equilibrium of the model.

From the population-mobility condition, residents must be indifferent across all locations with positive residents. Hence, high amenities in a location must be offset in equilibrium by some combination of lower expected income net of commuting costs and/or a higher price of residential floor space. Given data on wages, travel times, and the price of residential floor space and an assumption about the utility function, it also follows that one can use this population mobility condition to back out the amenities required for the observed data to be an equilibrium of the model.

Through these differences in productivity and amenities across locations, quantitative urban models are able to rationalize the rich polycentric and asymmetric patterns of economic activity observed in real-world cities. Given the values of productivity and amenities backed out from the observed data, quantitative urban models can be used to structurally estimate the role of agglomeration forces in determining these variables. Before illustrating this, we next provide further intuition for the determination of equilibrium in these models.

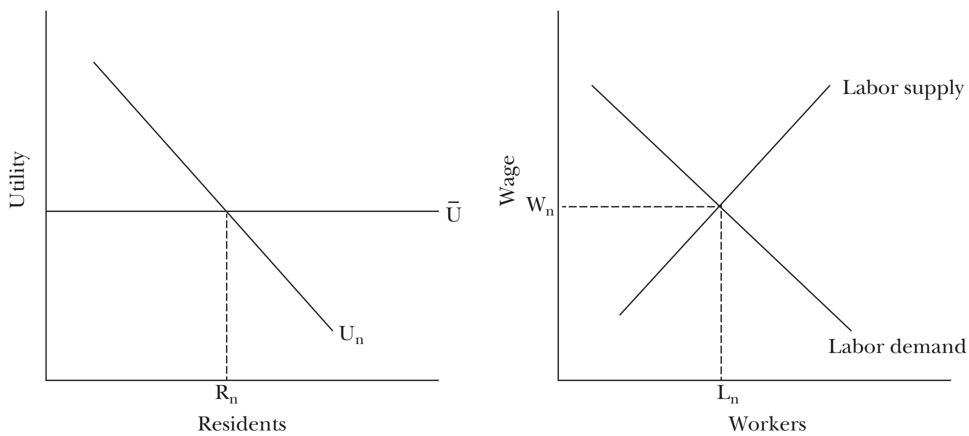
Equilibrium City Structure

We can gain a better sense of how internal city structure is determined at an intuitive level by looking at residential and workplace choices in partial equilibrium, where these two sets of decisions are then linked in general equilibrium. This partial equilibrium analysis connects closely with conventional models of labor demand and supply.

In the left-hand panel of Figure 2, we illustrate the determination of the number of residents (R_n) who choose to live in a certain location in the city. The horizontal line shows the reservation level of utility in the wider economy ($\bar{U}$). The downward-sloping line shows expected utility from residence n (U_n). Expected utility is decreasing in the number of residents for two reasons. First, as we increase the number of residents for a given supply of residential floor space, this bids up the price for residential floor space, which reduces expected utility. Second, as we increase the number of residents in a given location, we attract workers with lower idiosyncratic preferences for that location, which reduces expected utility through a composition or batting-average effect.

The equilibrium number of residents is determined by the intersection of the reservation level of utility in the wider economy ($\bar{U}$) and the expected utility from living in location n. The position of the expected utility curve (U_n) depends on amenities in location n and expected income net of commuting costs from access to workplaces from that location. An increase in amenities in location n (as an example, perhaps from improved access to green spaces) shifts outwards the expected utility curve (U_n), which increases the number of residents in location n (R_n). Similarly, an improvement in location n's connections to the transport network increases expected income net of commuting costs from that location, which shifts outwards

Figure 2
Residence and Workplace Choices



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

the expected utility curve (U_n), and increases the number of residents in location n (R_n).

In the right-hand panel of Figure 4, we illustrate the determination of the number of workers in each workplace (L_n). The downward-sloping line shows labor demand in workplace n, as determined by the equality between the wage and the value marginal product of labor. An increase in the number of workers employed in a location leads to a decrease in the wage because of diminishing marginal physical productivity of labor in the production technology. The upward-sloping line shows labor supply for workplace n, as determined by worker choices of residence and workplace. In order to increase labor supply, firms must offer a higher wage in order to attract workers with lower idiosyncratic preferences for that workplace.

The position of the labor demand curve depends on productivity in location n, while the position of the labor supply curve depends on the transport network and access to commuters from surrounding locations. An increase in productivity in location n (as an example, perhaps from improved access to natural water) raises the marginal product of labor, which shifts outwards the labor demand curve, and increases employment in location n (L_n). An improvement in location n's connections to the transport network increases the supply of commuters at a given wage, which shifts outwards the labor supply curve and increases employment in location n (L_n).

From the left-hand panel, a location has zero residents, and hence completely specializes as a workplace, if expected utility (U_n) always lies below the reservation level of utility in the wider economy ($\bar{U}$) for all positive values of residents (R_n). From the right panel, a location has zero employment, and hence completely specializes as a residence, if the labor demand curves lies below the labor supply curve for all positive values of workers (L_n). More generally, a location can either be a net importer of commuters if employment exceeds residents ($L_n > R_n$), or a net exporter of commuters if employment falls short of residents ($L_n < R_n$).

Although Figure 4 provides useful intuition, it is important to keep in mind that it only provides a partial equilibrium analysis and does not capture all general equilibrium relationships in the model. First, this figure focuses on the commuter market for residents and workers, but the position of the expected utility and labor demand curves is also influenced by the land market. Second, the expected utility curve in the left panel (U_n) is jointly determined with the labor supply curve in the right panel within a given location, because residents can work locally. Third, the expected utility curve in the left panel (U_n) for one location is jointly determined with the labor supply curve for other locations, because residents in one location commute to work in other locations. Therefore, an increase in amenities in surrounding locations, which raises the number of surrounding residents, increases a location's own supply of labor. Fourth, amenities and productivity depend on surrounding concentrations of residents and workers, respectively, through agglomeration forces. In Figure 4, we have assumed that these agglomeration forces are not too strong, such that expected utility (U_n) is downward-sloping in a location's own residents (R_n), and labor demand is downward-sloping in a location's own workers (L_n).

Extensions and Generalizations

Although we have considered a relatively parsimonious quantitative urban model, the tractability of these frameworks lends itself to a large number of extensions and generalizations, which can be used to address a range of public policy issues.

First, this class of models can accommodate nontraded goods, as in Heblich, Redding, and Sturm (2020). Second, the models can accommodate other reasons for travel apart from commuting, such as consumption trips, as in Miyauchi, Nakajima, and Redding (2022). Third, they can allow for multiple final goods with costly trade and technology differences, as in Eaton and Kortum (2002) and Redding (2016). Fourth, they can encompass final goods that are differentiated by origin and costly trade, as in Armington (1969), Allen and Arkolakis (2014), and Allen, Arkolakis, and Li (2017). Fifth, these quantitative urban models can encapsulate horizontally-differentiated firm varieties with costly trade, as in Helpman (1998), Redding and Sturm (2008), and Monte, Redding, and Rossi-Hansberg (2018).

Sixth, they can be used to quantify the impact of zoning regulations on internal city structure, as in Allen, Arkolakis, and Li (2017). Seventh, they can be used as a platform for evaluating neighborhood development programs, as in the analysis of the redevelopment of Detroit in Owens, Rossi-Hansberg, and Sarte (2020). Eighth, this kind of model can incorporate forward-looking investments in capital accumulation, as in Kleinman, Liu, and Redding (2023). Ninth, it can allow for multiple groups of workers that are heterogeneous, as in Redding and Sturm (2016) and Tsivanidis (2018). Tenth, whereas travel time was treated as exogenous and independent of commuting flows above, congestion can be introduced, as in Allen and Arkolakis (2022). Which of these specifications is most useful for empirical work depends on the data available and the public policy issue of interest.

Recent events have drawn attention to a range of public policy issues that can be addressed using quantitative urban models. The outbreak of the Covid-19 pandemic reminded us that disease contagion has been a powerful dispersion force throughout human history (for example, as discussed in Glaeser and Cutler 2021). New technologies and forms of managerial organization that allow remote or hybrid working can be interpreted as reductions in commuting costs, as workers no longer need to travel from their home to their workplace or are only required to do so for a smaller number of days each week (see the discussion in Barrero, Bloom, and Davis 2021). Similarly, autonomous vehicles can be interpreted as another technological innovation that reduces commuting costs. To the extent that workers no longer need to pay attention while driving, this will free up additional time for work or leisure. If an active ride-hailing market develops for autonomous vehicles, this may also free up substantial areas of land in urban areas that are currently used for parking private vehicles. In the empirical applications below, we provide another example from history of how a technological innovation (the invention of the steam passenger railway) reduced commuting costs and reshaped patterns of specialization by residence and workplace within urban areas.

Application 1: The Division of Berlin

Throughout the long literature on economic geography and urban economics, it has been empirically challenging to distinguish agglomeration and dispersion forces from variation in natural advantages. After all, high land prices and levels of economic activity in a group of neighboring locations are consistent with strong agglomeration forces, but equally consistent with shared amenities that make these locations attractive places to live (like leafy streets and scenic views) or common natural advantages that make these locations attractive for production (like access to natural water).[7] To disentangle these two alternative explanations for location choices, one requires a source of exogenous variation in the surrounding concentration of economic activity. 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.

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

The Qualitative Story

A protocol signed in London in September 1944 near the end of World War II designated separate occupation sectors in Berlin, Germany, 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. 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.

Ahlfeldt et al. (2015) provides evidence that Berlin's land price gradient in 1936 was approximately monocentric, with the highest values concentrated in the prewar central business district in the neighborhood of Mitte, with concentric rings of progressively lower land prices in the surrounding areas. However, Mitte was east of the future line of the Berlin Wall and thus was cut off when the wall was built. If one looks only at the areas of Berlin in 1936 that were going to become part of the future West Berlin, the two parts of the future West Berlin with the highest land prices in 1936 were an area just west of the prewar central business district and the future line of the Berlin Wall, and the Kudamm ("Kurfürstendamm") further west, which had developed into a fashionable shopping area in the decades leading up to World War II.

By 1986, looking at West Berlin following division, we find that the first prewar land price peak just west of the prewar central business district is entirely eliminated. This area ceased to be an important center of commercial and retail activity. Instead, the second prewar price peak in what had been the secondary area of the Kudamm develops into West Berlin's central business district during the period of division.

By 2006, after the reunification of Berlin, the prewar central business district that had been in the former East Berlin reemerges as a land price peak, as does the area just west of this prewar central business district and the former line of the Berlin Wall, which is now again a concentration of office and retail development.

These patterns are consistent with the qualitative predictions of the model developed above. Following division, the biggest declines in land prices are observed in the parts of West Berlin closest to the pre–World War II city's central business district. These parts of West Berlin experience the greatest reductions in access to production agglomeration forces, residential agglomeration forces, supplies of commuters, and employment opportunities from the areas of the prewar city that became East Berlin. There is also 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.

These observed changes in the land price gradient are accompanied by a similar reorientation of employment and residents within West Berlin.

Quantitative Evidence

To examine whether the quantitative urban model developed above can account for the observed changes in the spatial distribution of land prices, employment, and residents, Ahlfeldt et al. (2015) estimate the structural model's parameters. Using a given set of parameters, the structure of the model can be used to solve for the unobserved values of natural advantages for production, natural advantages for amenities, and the density of development (as measured by the ratio of floor space to land area). With this estimation procedure, the model exactly rationalizes the observed data on land prices, employment, and residents in each year before and after division and reunification as an equilibrium outcome.

The model's parameters are estimated using the identifying assumption that changes in natural advantages for production and amenities in each city block are uncorrelated with the change in the surrounding concentration of economic activity induced by Berlin's division and reunification. Because the city's division stemmed from military considerations during World War II and its reunification originated in the wider collapse of Soviet communism, the resulting changes in the surrounding concentration of economic activity are plausibly exogenous to changes in natural advantages in individual city blocks. In particular, these changes in natural advantages in West Berlin are assumed to be orthogonal to indicator variables for distance of grid cells to the prewar central business district. This identifying assumption requires that the systematic change in the gradient of economic activity in West Berlin relative to the prewar central business district following the city's division is explained by the mechanisms of the model—that is, by the changes in commuting access and production and residential agglomeration forces—rather than by systematic changes in natural advantages for production and amenities. 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.

The parameters are estimated for both division and reunification separately, and then by pooling all of the data together. All three specifications yield a similar pattern of estimated coefficients, with evidence of substantial agglomeration forces from production and residential externalities. 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. These agglomeration forces are highly localized. The estimates imply that both production and residential externalities fall to close to zero after around ten minutes of travel time, which corresponds to around 0.83 kilometers by foot (at an average speed of five kilometers per hour) and about four kilometers by underground and suburban railway (at an average speed of 25 kilometers per hour).

Other Evidence

These parameter estimates from Berlin's division and reunification 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 from the survey by Rosenthal and Strange (2004), but less than the elasticities from some quasi-experimental studies (for example, Greenstone, Hornbeck, and Moretti 2010; Kline and Moretti 2014).

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 nine 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 (Glaeser, Kolko, and Saiz 2001; Diamond 2016; Almagro and Domínguez-Lino 2022). Similarly, using data on an urban revitalization program in Richmond, Virginia, Rossi-Hansberg, Sarte, and Owens (2010) also find residential externalities are highly localized, with housing externalities falling by approximately one-half every 1,000 feet.

Taking the empirical findings of this section together, the quantitative urban model developed above is able to rationalize the rich patterns of spatial variation in land prices, employment, and residents observed in the data. Furthermore, for the estimated parameter values, the model is quantitatively successful in predicting the change in the internal city structure in response to the large-scale shock of Berlin's division and reunification.

Application 2: The Nineteenth-Century Steam Railway Revolution in London

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 and its trains travel around 76 million kilometers (about 47 million miles) each year. What is the role of London's transport network in sustaining its dense concentrations of economic activity? Heblich, Redding, and Sturm (2020) use the mid-nineteenth-century invention of steam railways as a natural experiment to explore this question. The key idea is that steam railways dramatically reduced travel time for a given distance, thereby lowering commuting costs, and 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 to 1921, before and after this transport innovation. Data are available for a number of different geographical definitions of London. Greater London, as defined by the boundaries of the modern Greater London Authority (GLA), includes a 1921 population of 7.39 million and an area of 1,595 square kilometers. The historical County of London has a 1921 population of 4.48 million and an area of 314 square kilometers. The City of London has a 1921 population of 13,709 and an area of about three square kilometers, and its boundaries correspond approximately to the Roman city wall.

The Qualitative Story

At the beginning of the nineteenth century, the most common mode of transport in London was walking, with average travel speeds in good road conditions of around three miles per hour. When the horse omnibus started in London in the 1820s, average travel reached perhaps six miles per hour. However, the opening of the London and Greenwich railway in 1836 as the first steam railway to be built specifically for passengers transformed the relationship between travel time and distance, with average travel speeds of around 21 miles per hour.

The availability of the steam passenger railway was followed by a large-scale change in the organization of economic activity within Greater London. In the first half of the nineteenth century, population in the City of London was relatively constant (at around 130,000), while population in Greater London grew substantially (from 1.14 million to 2.69 million). From 1851 onwards, shortly after the first steam passenger railways, population in the City of London falls sharply 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.

In the City of London, we observe the emergence of the first large-scale separation between the night population (where people sleep) and the day population (where they work). In the opening decades of the nineteenth century, the night and day populations for the City of London are relatively similar at about 150,000. But in the decades following the first steam passenger railways, in the City of London day censuses for 1866, 1881, 1891 and 1911, the sharp decline in night population is combined with a steep rise in day population. By 1911, the day population of the City of London was approaching 400,000, while the night population had fallen to only 10,000. This pattern of empirical results is consistent with the idea that the reduction in commuting costs from this new transport technology allowed the City of London to specialize as a workplace (importing commuters), while the surrounding suburbs specialized as residences (exporting commuters).

Quantitative Evidence

To rationalize these empirical findings, Heblich, Redding, and Sturm (2020) develop an estimation procedure that illustrates how quantitative urban models can be used to undertake counterfactuals for transport infrastructure improvements or other public policy interventions. Given data on economic activity in an initial observed equilibrium and estimates of the changes in travel times from a transport improvement, these models can be solved for the predicted change in the spatial organization of economic activity. Using these predictions, the economic benefit from the transport improvement can then be compared to estimates of its construction costs.

As a first step in implementing this procedure, the relationship between commuting costs and travel times is estimated using data on bilateral commuting flows and the observed transport network in London in the year 1921. This transport network includes overground and underground railways, buses and trams, and walking, since commuting by private car was negligible in London in 1921. Because the placement of transport infrastructure is potentially endogenous, this estimation uses an instrumental variable for travel time using the transport network in the form of bilateral geographical distance between locations. Given these estimates and the observed evolution of the transport network over time, predicted changes in commuting costs from the expansion of the railway network can be calculated.

Armed with these estimates of changes in commuting costs, observed data on bilateral commuting flows for 1921, and data on property values and employment by residence in earlier decades, the model can be solved for predicted employment by workplace and commuting flows back to the beginning of the nineteenth century. An advantage of using these historical data on property values and employment by residence is that the values of these variables in earlier decades can be used to control for other factors that changed over time in addition to the transport network, such as productivity or amenities.

The model successfully captures the observed sharp divergence between the night and day populations 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). 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 employment density with respect to distance from the center of the City of London declines between 1831 and 1921, and the share of the 13 boroughs within five 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 long line of empirical research that finds evidence of employment (and population) decentralization in response to transport improvements, as reviewed in Redding and Turner (2015) and Redding (2022b). These findings suggest that present technological changes, such as innovations in remote working and autonomous vehicles, have the potential to further decentralize economic activity.

As a specification check, the model's predictions for commuting flows are compared to historical data from the personnel ledgers of Henry Poole Tailors, a high-end bespoke tailoring firm, which was founded in 1802. The firm collected data on workers' residential addresses at the time they were first hired, thus allowing an estimate of commuting distances to the firm. There are of course a number of possible reasons why the pattern of employee commutes to a particular firm at a specific site might differ from the model's predictions. Nevertheless, the model is remarkably successful in capturing the change in the distribution of commuting distances between these time periods. In the opening decades of the railway age in the 1850s and 1860s, most workers in Westminster in both the model and data lived within five kilometers of their workplace. By the turn of the twentieth century, commuting distances up to 20 kilometers are observed in both the model and data.

Evaluation of Transport Infrastructure Investments

The estimated model also can be used to evaluate the economic benefits of the construction of London's railway network, holding constant all other factors, such as productivity and amenities. In this analysis, the impact of the railway network on worker utility depends on assumptions about labor mobility and land ownership. In particular, suppose that the economy consists of workers who own only labor and landlords who own only land, and assume that workers are perfectly mobile between London and the wider economy at an unchanged reservation level of utility. In this case, as the construction of London's railway network reduces commuting costs and raises expected worker utility, it attracts a population inflow, which bids up the price of land, until expected worker utility in London in the new equilibrium is equal to the unchanged reservation level of utility in the wider economy. Under these assumptions, all economic benefits from the construction of London's railway network accrue to landlords through a higher price of land. More generally, if labor is imperfectly mobile between London and the wider economy, the economic benefits from the railway network are enjoyed by both workers and landlords.

Under a range of different assumptions about labor mobility, the economic benefits from the construction of London's railway network are found to exceed historical estimates of its construction costs based on the capital issued by railway companies. The ratio of benefits to costs is substantially larger once production and residential agglomeration forces are taken into account. In the presence of these forces, the population inflow induced by the reduction in commuting costs induces endogenous increases in productivity and amenities. Similarly, the ratio of benefits to costs is enhanced by taking into account complementary investments in buildings and structures. The reason is that the resulting population inflow raises the demand for commercial and residential floor space, which leads to an endogenous increase in the supply of floor space from the construction sector. An important takeaway for these findings is the need to take into account agglomeration forces and complementary investments in buildings and structures in conventional cost-benefit analyses of transport infrastructure investments.

Looking beyond this empirical application, policymakers are often interested in comparing alternative possible transport investments, such as which links in a railway or highway network to improve. To develop a framework to address this question, Allen and Arkolakis (2022) embed a specification of endogenous route choice in a quantitative spatial model. In their approach, individuals consider travel costs and choose the least-cost route. A key implication of this framework is that the welfare effects of a small improvement in a transport link are equal to the percentage cost saving multiplied by the initial value of travel along that link.[8] Barwick et al. (2020) use an approach along these lines for an analysis of China's High Speed Rail Network, while Gupta, Van Nieuwerburgh, and Kontokosta (2022) provide evidence on the quantitative impact of the latest expansion to New York's subway network, the Second Avenue Subway.

More generally, Fajgelbaum and Schaal (2020) develop a framework for characterizing optimal transport networks in spatial equilibrium. This characterization is challenging, because the problem is high dimensional. However, they show that the problem of finding the optimal transport network can be transformed into the problem of finding the optimal flow in a network, which has been studied in the operations research literature. While this approach has so far been applied to trade in goods between cities, incorporating commuting within cities is an exciting avenue for further research.

Conclusion

Real-world cities feature complex internal structures, with a rich specialization by residential and commercial land use and an intricate division of labor. The real-world cities in which people live often 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 and to explore these observed features of the data. These frameworks can accommodate many locations that differ in productivity, amenities, land area, the supply of floor space, and transport connections. Nevertheless, these models remain tractable and amenable to theoretical analysis with a manageable number of parameters to be estimated.

[8] Although this result is derived for particular functional 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.

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. Another insight is the role of advanced transport networks in sustaining dense concentrations of economic activity in modern metropolitan areas.

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

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

Although these mechanisms are well understood conceptually, there is relatively little evidence on their empirical importance, with a few exceptions such as Ellison, Glaeser, and Kerr (2010). Over time, the nature of economic activity undertaken within cities has changed dramatically, from the marketplaces and ports of pre-industrial Europe, through the centers of manufacturing of the industrial revolution, through the concentrations of office space of the mid-twentieth century, to an increasing focus on the consumption of nontraded goods and services in the twenty-first century. Using the verbs from occupational descriptions, Michaels, Rauch, and Redding (2018) quantify the change in the tasks undertaken by workers in cities over time. Whereas the tasks most concentrated in cities in 1880 involved the manipulation of the physical world, such as "Thread and Sew," those most concentrated in cities in 2000 involve human interaction, such as "Advise and Confer."

Given these large-scale changes in the types of economic activities performed in urban areas over time, it is plausible to think that the nature and scope of agglomeration economies could have evolved as well. Consistent with this idea, Autor (2019) finds substantial changes in the urban wage premium for workers with different levels of skills over time. At the beginning of his sample period in the 1970s, average wages were sharply increasing in population density for both low-skill workers (high-school or less) and high-skill workers (some college or greater). By the end of the sample period in 2015, this wage premium to population density had increased for high-skill workers but almost disappeared for low-skill workers.

Looking ahead, the wealth of newly-available sources of Geographical Information Systems (GIS) data promises to offer new opportunities to distinguish between different mechanisms for agglomeration, including ride-hailing data (from firms like Uber and Lyft), smartphone data with Global Positioning System (GPS) information, firm-to-firm data from sales (or value-added tax) tax records, credit card data with consumer and firm location, barcode scanner data with consumer and firm location, public transportation commuting data, work-from-home data, and satellite imaging data.

Over the centuries, cities have changed drastically—from marketplaces, to the locus of manufacturing industry, to clusters of office and retail development, and to centers of consumption. But as long as there are benefits to reduced costs of moving people, goods, and ideas, cities in some form are likely to thrive and prosper.

■ *I am grateful to Princeton University for research support. Thanks to Nan Xiang for excellent research assistance. I am very grateful to the editors for their extremely helpful comments and suggestions. I also would like to thank Gabriel Ahlfeldt, Treb Allen, Jason Barr, Dave Donaldson and Clifford Winston for helpful comments and suggestions. Any opinions, errors and omissions are the author's responsibility alone.*

References

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica* 83 (6): 2127–89.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–1140.

Allen, Treb, and Costas Arkolakis. 2022. "The Welfare Effects of Transportation Infrastructure Improvements." *Review of Economic Studies* 89 (6): 2911–57.

Allen, Treb, Costas Arkolakis, and Xiangliang Li. 2017. "Optimal City Structure." Unpublished.

Almagro, Milena, and Tomás Domínguez-Iino. 2022. "Location Sorting and Endogenous Amenities: Evidence from Amsterdam." University of Chicago Becker Friedman Institute for Economics Working Paper 2022-162.

Alonso, William. 1964. *Location and Land Use.* Cambridge MA: Harvard University Press.

Armington, Paul S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *International Monetary Fund Staff Papers* 16 (1): 159–78.

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. "Networking Off Madison Avenue." *Review of Economic Studies* 75 (4): 1011–38.

Autor, David H. 2019. "Work of the Past, Work of the Future." *American Economic Review* 109: 1–32.

Barr, Jason, Fred H. Smith, and Sayali J. Kulkarni. 2018. "What's Manhattan Worth? A Land Values Index from 1950 to 2014." *Regional Science and Urban Economics* 70: 1–19.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick." NBER Working Paper 28731.

Barwick, Panle Jia, Dave Donaldson, Shanjun Li, and Yatang Lin. 2020. "The Welfare Effects of Passenger

Transportation Infrastructure: Evidence from China." Unpublished.

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

Duranton, Gilles, and Hubert Jayet. 2011. "Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities." *Journal of Urban Economics* 69 (1): 56–71.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2063–2117. Amsterdam: Elsevier.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–1213.

Fajgelbaum, Pablo D., and Edouard Schaal. 2020. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica* 88 (4): 1411–52.

Fortheringham, A. S., and M. E. O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications.* Dordrecht: Kluwer Academic Publishers.

Fujita, Masahisa, and Hideaki Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations." *Regional Science and Urban Economics* 12 (2): 161–96.

Glaeser, Edward, and David Cutler. 2021. *Survival of the City: Living and Thriving in an Age of Isolation.* New York: Penguin Press.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. 2001. "Consumer City." *Journal of Economic Geography* 1 (1): 27–50.

Green, David R. 1988. "Distance to Work in Victorian London: A Case Study of Henry Poole, Bespoke Tailors." *Business History* 30: 179–94.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta. 2022. "Take the Q Train: Value Capture of Public Infrastructure Projects." *Journal of Urban Economics* 129: 103422.

Haughwout, Andrew, James Orr, and David Bedoll. 2008. "The Price of Land in the New York Metropolitan Area." *Current Issues in Economics and Finance* 14 (3): 1–7.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm. 2020. "The Making of the Modern Metropolis: Evidence from London." *Quarterly Journal of Economics* 135 (4): 2059–2133.

Helpman, Elhanan. 1998. "The Size of Regions." In *Topics in Public Economics: Theoretical and Applied Analysis*, edited by David Pines, Efraim Sadka, and Itzhak Zilcha, 33–54. Cambridge, UK: Cambridge University Press.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45 (3): 511–18.

Kleinman, Benny, Ernest Liu, and Stephen J. Redding. 2023. "Dynamic Spatial General Equilibrium." *Econometrica*, forthcoming.

Kline, Patrick, and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *Quarterly Journal of Economics* 129 (1): 275–331.

Lucas, Robert E., Jr., and E. Rossi-Hansberg. 2002. "On the Internal Structure of Cities," *Econometrica* 70 (4): 1445–76.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences.* Cambridge, MA: Harvard University Press.

Marshall, Alfred. 1920. *Principles of Economics.* London: Macmillan.

McDonald, John F., and Daniel P. McMillen. 2010. *Urban Economics and Real Estate: Theory and Policy.* 2nd ed. Hoboken, NJ: John Wiley and Sons.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3 (4): 303–28.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2018. "Task Specialization in U.S. Cities from 1880 to 2000." *Journal of the European Economic Association* 17 (3): 754–98.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review* 57 (2): 197–210.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J. Redding. 2022. "The Economics of Spatial Mobility:

Theory and Evidence Using Smartphone Data." NBER Working Paper 28497.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration and Local Employment Elasticities." *American Economic Review* 108 (12): 3855–90.

Muth, Richard. 1969. *Cities and Housing: The Spatial Patter of Urban Residential Land Use.* Chicago: University of Chicago Press.

Owens, Raymond, III, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte. 2020. "Rethinking Detroit." *American Economic Journal: Economic Policy* 12 (2): 258–305.

Redding, Stephen J. 2016. "Goods Trade, Factor Mobility and Welfare." *Journal of International Economics* 101: 148–67.

Redding, Stephen J. 2022a. "Trade and Geography." In *Handbook of International Economics*, Vol. 5, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 147–217. Amsterdam: North-Holland.

Redding, Stephen J. 2022b. "Suburbanization in the USA, 1970–2010." *Economica*, 89 (S1): S110–36.

Redding, Stephen J. 2023. "Replication data for: Quantitative Urban Models: From Theory to Data." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E185425V1.

Redding, Stephan J., and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics* 9: 21–58.

Redding, Stephen J., and Daniel M. Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review* 98 (5): 1766–97.

Redding, Stephen J., and Daniel M. Sturm. 2016. "Neighborhood Effects: Evidence from the Streets of London." Unpublished.

Redding, Stephen J. and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 1339–98. Amsterdam: North-Holland.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economics." In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–71. Amsterdam: Elsevier.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens III. 2010. "Housing Externalities." *Journal of Political Economy* 118 (3): 485–535.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: W. Strahan and T. Cadell.

Tsivanidis, John N. 2018. "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." PhD diss. University of Chicago.

United Nations. 2019. *World Urbanization Prospects: The 2018 Revision.* New York: United Nations Department of Economic and Social Affairs.

This article has been cited by:

1. Benny Kleinman, Ernest Liu, Stephen J. Redding. 2024. The Linear Algebra of Economic Geography Models. *AEA Papers and Proceedings* **114**, 328-333. [Abstract] [View PDF article] [PDF with links]