Spatial Economics*

Stephen J. Redding†
Princeton University, NBER and CEPR

December 11, 2024

Abstract

This paper reviews recent research in spatial economics. The field of spatial economics is concerned with the determinants and effects of the location of economic activity in geographic space. It analyses how geographical location shapes the economic activities performed by agents, their interactions with one another, their welfare, and the effects of public policy interventions. Research in this area has benefited from the simultaneous development of new theoretical techniques, new sources of geographic information systems (GIS) data, rapid advances in computing power, machine learning and artificial intelligence, and renewed public policy interest in infrastructure and appropriate policies towards places "left-behind" by globalization and technology. Among the insights from this research are the role of goods and commuting market access in determining location choices; the conditions under which the location of economic activity is characterized by multiple equilibria; the circumstances under which temporary shocks can have permanent effects (hysteresis or path dependence); the heterogeneous and persistent impact of local shocks; the magnitude and spatial decay of agglomeration economics; and the role of both agglomeration forces and endogenous changes in land use in shaping the impact of transport infrastructure improvements.

Keywords: cities, economic geography, regions, spatial economics

JEL Classification: F15, R10, R12

*This paper was commissioned for the *Oxford Research Encyclopedia of Economics and Finance*. I am grateful to Princeton University for research support. I would like to thank Zhuokai Huang for excellent research assistance. I am grateful to Dave Donaldson, Ed Glaeser and Jacques-François Thisse for helpful comments. Any errors, limitations and omissions are the responsibility of the author alone.

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

1 Introduction

The observed distribution of economic activity across geographical space is extremely uneven. This unevenness is evident from the comparison between urban and rural areas. By the year 2050, approximately two thirds of the world's population is projected to live in cities and the urban population is predicted to increase by around 2.5 billion people. Africa and Asia are expected to account for 90 percent of this incease in the urban population.[1]

Spatial economics is concerned with the determinants and effects of the location of economic activity in geographic space. It analyses how geographical location shapes the economic activities performed by agents, their interactions with one another, their welfare, and the effects of public policy interventions.

The remainder of this article reviews insights from recent research in spatial economics. Section 2 discusses the scope of spatial economics and distinguishes between two main lines of research: (i) systems of cities and regions versus (ii) internal city structure. Section 3 highlights an important distinction between first and second-nature geography. Section 4 reviews research on systems of cities and regions. Section 5 examines work on internal city structure. Section 6 summarizes conclusions and considers areas for future research.

2 Scope of Spatial Economics

In some respects, spatial economics is a subset of network economics, because it focuses on a particular type of network, namely one arrayed in geographic space. In general, we can represent a network as a graph of nodes (e.g., people, firms, etc.) and edges (connections between these nodes). In spatial economics, the nodes are arrayed in geographic space (e.g., regions, cities, etc.) and the edges correspond to relationships that take place in geographic space (e.g., distance, travel time, trade flows, migration flows, commuting flows).

In other respects, spatial economics is a superset of traditional fields of economics, such as international trade. While international trade focuses on interactions between countries, spatial economics considers interactions for any level of spatial aggregation (countries, regions, counties, cities, city blocks, etc.). Besides this difference in spatial scale, research in international trade typically assumes that factor endowments are exogenous, whereas factor mobility is central to spatial economies, with typically only geographical land area considered as exogenous. While urban economics typically focuses on cities, spatial economics includes both rural and urban areas, and considers spatial units ranging from individual blocks within cities through countries to the global economy as a whole.

[1] See for example the World Urbanization Prospects, United Nations (2018). For broader historical evidence on patterns of urbanization, see Michaels et al. (2012) and Desmet and Henderson (2015).

Perhaps one of the exciting aspects of spatial economics is that it lies at the intersection of several traditional fields of economics, including international trade, urban economics, labor economics, public finance and development economics. Researchers in these fields are frequently concerned with geographical location. This concern can arise because geographical location is central to the question at hand (e.g., the local impact of the Tennessee Valley Authority (TVA)). Or it can emerge because geographical locations are a source of observational data or exogenous variation. Partly through including researchers from several traditional fields, spatial economics encompasses the full range of methods in economics, including theoretical modeling, reduced-form microeconometrics, structural microeconometrics, and quantitative macroeconomics.

A first strand of research in spatial economics analyzes system of cities or regions (the network of economic interactions between cities or regions). A second strand of work considers internal city structure (the network of economic interactions within a single city).[2] The main distinction between these two lines of work is that their different spatial scales change the relative importance of alternative economic mechanisms. Between cities, goods trade and migration are more salient. Within cities, commuting (the separation of residence and workplace) and consumption travel (the separation of residence and consumption) are much more pertinent.

3 First and Second-Nature Geography

Within each of these lines of research, there is an important distinction between two different ways in which geographical location can matter for the spatial distribution of economic activity. "First-nature geography" corresponds to exogenous differences in locational fundamentals or natural advantages, including access to rivers or natural harbors. "Second-nature geography" refers to endogenous decisions by economic agents of where to locate relative to one another. For example, people and firms can endogenously cluster together to reduce the transportation costs for goods, people and ideas (through what are known as *agglomeration forces*), even without differences in locational fundamentals or natural advantages.[3]

Whether the uneven spatial distribution of economic activity is driven by first-nature or second-nature geography is fundamental to many economic and public policy debates. Explanations based on the agglomeration forces of second-nature geography typically feature exter-

[2] For related discussions of systems of cities and regions, see Abdel-Rahman and Anas (2004), Rossi-Hansberg (2019), Venables (2019), Chen and Peng (2020) and Redding (2022b). For discussions of internal city structure, see Anas et al. (1998), Thisse (2019), Fujita (2020), and Redding (2023, 2024). For a review of quantitative spatial economics, see Redding and Rossi-Hansberg (2017). On spatial economics more widely, see Proost and Thisse (2019).

[3] Sometimes the term "second-nature" is reserved for *historical* man-made factors that are fixed in location, durable, and sunk, such as canals, railroads, or highways (e.g., Lin and Rauch 2022), excluding contemporaneous man-made forces. We follow Krugman (1993) in using "second-nature" to refer to all effects of the location of economic agents relative to one another (whether historical or contemporaneous).

nalities: when one agent chooses where to locate, she does not take into account the impact of her decision on other agents choices of where to locate. A distinction can be drawn between technological externalities (e.g., knowledge spillovers) and pecuniary externalities that are mediated through markets (e.g., demand for non-traded goods and services). If these externalities exist, the market equilibrium potentially can be inefficient, which opens up the possibility for public policy interventions to raise welfare. Therefore, understanding the magnitude of agglomeration forces is key to evaluating the impact of local taxation, and a range of place-based policies such as Empowerment Zones, zoning and building regulations, and transport improvements, among other policies.

Spatial economics has enumerated a number of different potential sources of agglomeration forces. Marshall (1920) made an influential distinction between three sets of forces. First, there is labor market pooling: workers and firms may have an incentive to colocate, in order to make is easier for firms to find suitable workers, and for workers to find suitable firms. Second, there are non-traded inputs: firms have an incentive to cluster together, in order for buyers to gain improved access to suppliers of non-traded inputs, and for these suppliers of these non-traded inputs to benefit from improved access to buyers. Third, there are knowledge spillovers, whereby concentrating people together may facilitate the invention and diffusion of knowledge.

An alternative trichotomy is proposed by Duranton and Puga (2004). A first set of agglomeration forces are based on sharing, which includes sharing indivisible facilities, the gains from a wider variety of input suppliers, the benefits from a finer level of specialization, and risks. A second group of agglomeration forces are based on matching, which includes a higher expected match quality, higher matching probabilities, and a reduction in hold-up within matches. A third category of agglomeration forces are based on learning, which includes the creation, diffusion, and accumulation of knowledge.

Within each of these different classifications of agglomeration forces, a distinction can be drawn between agglomeration forces that operate within industries ("localization economies") versus those that extend across industries ("urbanization economies"). Localization economies encourage the formation of specialized cities based around a single industry (Henderson 1974). In contrast, urbanization economies foster the development of cities with a diversified industrial structure (Jacobs 1961). Traditionally, research in either of these areas has focused on "static agglomeration forces" that operate within each time period, such as contemporaneous knowledge spillovers. More recently, research has begun to explore "dynamic agglomeration forces" that operate across time periods, such as young workers moving to cities to learn when young, and then retiring elsewhere when old (e.g., De La Roca and Puga 2017).

These agglomeration forces (sometimes termed centripetal forces) pull economic activity together. Working against them are congestion or dispersion forces (sometimes termed centrifugal forces) that push economic activity apart. These congestion forces also take a number of forms, including commuting costs, immobile factors of production such as land, various forms of congestion (including traffic congestion), and the spread of disease.[4] The observed spatial distribution of economic activity reflects the combined influence of exogenous first-nature geography and endogenous second-nature geography (both agglomeration and dispersion forces).

4 Systems of Cities and Regions

Having introduced these general conceptual distinctions, we now turn to research on systems of cities or regions, in which goods trade, migration and knowledge spillovers are the key economic interactions between locations.

Rosen-Roback Model One of the most influential theoretical frameworks for thinking about the spatial distribution of economic activity across cities and regions is the Rosen-Roback model (Rosen 1979 and Roback 1982). This framework is also related to research on systems of cities following Henderson (1974). The Rosen-Roback model highlights the role of differences in productivity and amenities in shaping the location of economic activity. Markets are assumed to be competitive. All locations are assumed to produce a single final good that is costlessly traded. Preferences depend on consumption of this final good, residential land use and amenities. Output of this single final good depends on labor input, commercial land use and productivity. Workers are assumed to be identical and perfectly mobile across locations. Land is perfectly geographically immobile, but is endogenously allocated between residential and commercial use, to arbitrage away any differences in the rate of return from these alternative land uses.

A key concept that emerges from the Rosen-Roback model is the notion of spatial equilibrium, in which no worker or firm has an incentive to change their location choices. If workers are assumed to be identical, then to induce some of them to pay high land prices in densely-populated locations, these high land prices must be compensated by either high wages or high amenities, in order to ensure that workers are indifferent across all locations with positive population. If firms produce an identical product and markets are perfectly competitive, then to induce some firms to pay high land prices and high wages in densely-populated locations, these higher factor prices must be compensated by higher productivity, in order to ensure that firms make zero profits across all locations with positive production. Combining these two insights, a key implication of the concept spatial equilibrium is that the concentration of economic activity ultimately must be explained by either higher productivity or higher amenities, where both can be influenced by first-

[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).

nature geography (natural advantages) and second-nature geography (agglomeration forces).

Sorting Models An alternative approach to thinking about the spatial distribution of economic activity is provided by assignment or sorting models (building on Roy 1951 and Sattinger 1993). Whereas the Rosen-Roback model focuses on the case in which workers are *ex ante* homogeneous, sorting models allow both workers and locations to be *ex ante* heterogeneous. Consider a continuum of locations that are differentiated along a single vertical dimension (location quality) and a continuum of workers that are differentiated along another vertical dimension (worker skill). Markets are assumed to be competitive. All locations are again assumed to produce a single final good that is costlessly traded across locations. Productivity in the production of this single final good is assumed to depend on both worker skill and location quality.

In models of this form (see in particular Davis and Dingel 2020), the equilibrium pattern of worker sorting depends on the degree of substitutability or complementarity between location quality and worker skill. A common assumption is that the production technology is log supermodular in these two characteristics, which requires that the return to higher location quality is greater for higher worker skill, and the return to higher worker skill is greater for higher location quality. Under this assumption, the spatial equilibrium is generically characterized by positive assortative matching (PAM), such that higher-quality locations are populated by higher-skill workers. In this spatial equilibrium, no individual worker of a given skill has an incentive to change location, but workers with different skills obtain different levels of utility.

New Economic Geography Models In both of these traditional approaches to modelling spatial equilibrium, locations are connected through population mobility and goods trade. However, there is no notion of geographical space, because both population mobility and trade in goods are assumed to be frictionless. A major breakthrough in modelling spatial equilibrium came with the development of the "new economic geography," which explicitly models costly trade in geographical space (Krugman 1991, Krugman and Venables 1995, Fujita et al. 1999).

Given the complexity of modelling interactions in geographical space, early theoretical research in new economic geography assumed away any differences in first-nature geography to characterize the mechanisms of second-nature geography. Researchers typically considered a small number of symmetric regions, or a "featureless plain" or "seamless world," in which locations are *ex ante* homogeneous. Nevertheless, these locations become *ex post* heterogeneous in equilibrium, through the emergence of an uneven spatial distribution distribution of economic activity driven by agglomeration and dispersion forces.

The canonical model of Krugman (1991) considers two regions that are *ex ante* identical (North and South). There are two production sectors: agriculture and manufacturing. Agricultural goods are homogeneous and produced under conditions of constant returns to scale and perfect competition. Manufacturing consists of horizontally-differentiated varieties that are produced under conditions of increasing returns to scale and monopolistic competition. Agricultural farmers are geographically immobile and equally distributed between the two regions. Manufacturing workers are perfectly mobile between the two regions.

Agglomeration forces arise from the combination of love of variety preferences over manufacturing goods, increasing returns to scale and transport costs.[5] Together these three assumptions create a backward linkage: Increasing returns to scale provide firms with an incentive to concentrate manufacturing production in a single location, while transport costs provide the incentive for this concentration to occur close to large markets. Together these three assumptions also create a forward linkage: Love of variety preferences provide workers with an incentive to consume the manufacturing goods supplied by all firms, while transport costs provide the incentive to locate close to where those goods are produced.

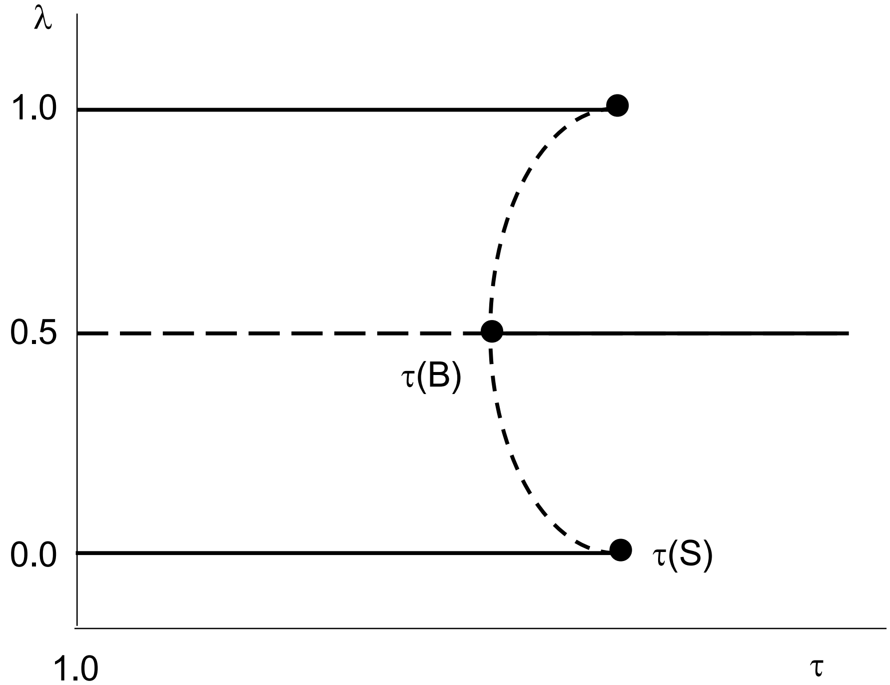
These forward and backward linkages mutually reinforce one another in a circular process of cumulative causation. This process of cumulative causation favors the concentration of manufacturing activity, encouraging the formation of a manufacturing core in one region and an agricultural periphery in the other region. Dispersion forces arise from the assumptions of transport costs and geographically-immobile farmers. Together these two assumptions encourage some manufacturing firms to locate in each region to serve agricultural farmers.[6]

Whether a core-periphery structure emerges in equilibrium depends on the relative strength of these agglomeration and dispersion forces. As transport costs fall, both agglomeration and dispersion forces become weaker, but the dispersion forces decline in absolute magnitude faster than the agglomeration forces. As shown in Figure 1, for sufficiently high transport costs (τ), there is a symmetric equilibrium, with manufacturing and agriculture located in both regions. As transport costs fall below a critical value (the "sustain point," $\tau(S)$), it becomes possible to sustain a core-periphery equilibrium, in which manufacturing is only located in one of the two regions. As transport costs fall further below another critical value ("the break point," $\tau(B)$), a core-periphery equilibrium becomes the only stable spatial equilibrium. Only when transport costs fall to zero ($\tau = 1$) does location in geographical space become irrelevant.

[5]In terms of Marshall (1920)'s classification, these agglomeration forces correspond to a form of locally-traded inputs, while in terms of Duranton and Puga (2004)'s trichotomy, they correspond to a form of sharing.

[6]While agricultural farmers provide the dispersion force in Krugman (1991), an alternative source of a dispersion force is an inelastic supply of land, as in Helpman (1998).

Figure 1: Core-Periphery Bifurcation



Note: The figure shows how the configuration of equilibria in Krugman (1991) varies with transportation costs (τ), where $\tau > 1$ is the fraction of a good that must be shipped in order for one unit to arrive, such that $\tau = 1$ corresponds to zero transport costs. Solid lines denote stable equilibria. Dashed lines denote unstable equilibria. λ is the share of manufacturing workers in a region. $\tau(B)$ is the break point and $\tau(S)$ is the sustain point.

Since the two regions are *ex ante* identical, it follows that for parameter values for which a core-periphery equilibrium is sustainable, it is indeterminant which of the two regions becomes the manufacturing core. Therefore, a key prediction of this theoretical literature on new economic geography is that the spatial distribution of economic activity can be characterized by multiple equilibria. In the presence of these multiple equilibria, small policy interventions can potentially have discontinuous effects by shifting the economy between multiple equilibria.

This possibility of multiple equilibria stimulated a long line of empirical research examining whether temporary shocks can have permanent effects ("hysteresis" or "path dependence") by shifting the location of economic activity between multiple steady states. Davis and Weinstein (2002) uses the bombing of Japanese cities during the Second World War as a temporary large-scale shock and finds no evidence of path dependence. Bleakley and Lin (2012) examines portage sites in the United States, which had historical natural advantages for transshipment for water-borne trade. Although these natural advantages have long since ceased to be relevant, they are found to have permanent effects on the spatial distribution of economic activity. Determining the circumstances under which the location of economic activity either is or is not characterized by path dependence remains a lively area of ongoing research.[7]

[7] See also Redding et al. (2011) and Michaels and Rauch (2018).

Quantitative Spatial Models Abstracting from first-nature geography allowed early theoretical research on new economic geography to characterize the mechanisms of second-nature geography. However, real world economies are not well approximated by a small number of symmetric regions or a "featureless plain" or "seamless world," which limited the usefulness of these models for empirical research.

A major breakthrough has been the development of quantitative spatial models (Allen and Arkolakis 2014, Desmet and Rossi-Hansberg 2014, Redding 2016, Caliendo et al. 2018). These quantitative spatial models are rich enough to connect to key features of the observed data, such as many heterogeneous locations that differ in productivity, amenities and trade costs. To do so, they incorporate both second-nature geography (agglomeration and dispersion forces) and first-nature geography (exogenous differences in productivity, amenities, land supply and trade costs). Nevertheless, these models remain tractable and amenable to a theoretical analysis of their properties, including the existence and uniqueness of the equilibrium. In contrast to earlier computable general equilibrium (CGE) models, these quantitative spatial models typically have only a small number of structural parameter to estimate. Therefore, they lend themselves to credible identification of these parameters, using quasi-experimental sources of exogenous variation. Since these quantitative spatial models are able to rationalize the observed spatial distribution of economic activity as an equilibrium, they can be used to undertake counterfactuals for the impact of empirically-relevant public-policy interventions (e.g., the construction of a particular highway link) on this observed spatial distribution.

In general, there exists an entire class of quantitative spatial models that are isomorphic with respect to their theoretical properties of existence and uniqueness and their counterfactual predictions (Allen and Arkolakis 2014, Allen et al. 2019, Allen et al. 2024). This class is defined by a constant elasticity structure, in which economic outcomes in one location are a constant elasticity function of economic outcomes in all locations weighted by a network of bilateral frictions between locations (e.g., trade costs). This class includes models in which goods are differentiated by origin (Armington 1969), models in which specialization arises from Ricardian technology differences (Eaton and Kortum 2002), and new economic geography models in which specialization arises from love of variety and increasing returns to scale (Helpman 1998).[8] Given this constant elasticity structure, sufficient conditions for the existence and uniqueness of the equilibrium can be derived, which depend only on the model's structural parameters (elasticities), and hence hold for any network of bilateral frictions between locations.

Given the model's structural parameters (elasticities) and the observed endogenous variables

[8] For tractability, quantitative urban models have focused on immobile land as the congestion force, as in Helpman (1998), instead of immobile agricultural farmers, as in Krugman (1991), even though the comparative statics of these two models with respect to transport costs are quite different.

(e.g., population, wages), quantitative spatial models in this class have the property that they can be inverted to recover unique values of the unobserved structural residuals (exogenous components of productivity, amenities and trade costs) that exactly rationalize the observed data as an equilibrium. Therefore, these quantitative spatial models provide a framework for assessing the contributions of first and second-nature geography to the observed spatial distribution of economic activity. In order to estimate the model's structural parameters (elasticities), researchers require additional information, typically in the form of orthogonality conditions on either the levels or changes in these structural residuals.

Another property of this class of constant elasticity quantitative spatial models is that they lend themselves to solving for counterfactuals using "exact-hat algebra." According to this approach, researchers solve for a counterfactual equilibrium, using only the observed values of the model's endogenous variables in the initial equilibrium in the data and assumed changes in exogenous location characteristics (e.g., reductions in trade costs from a transport improvement), without needing to know the levels of the unobserved location characteristics in the initial equilibrium. Implicitly, the observed endogenous variables together with the equilibrium conditions of the model contain enough information to control for the levels of the unobserved location characteristics. For parameter values for which there is a unique equilibrium in the model, these counterfactuals yield determinate predictions for the impact of public policy interventions on the spatial distribution of economic activity. When implementing quantitative spatial models using spatially-disaggregated data, one empirical challenge is that the observed endogenous variables can include small-sample variation (granularity). One approach to this empirical challenge is to use the estimated model's predictions for the endogenous variables instead of their observed values when undertaking counterfactuals (Dingel and Tintelnot 2020).

By explicitly modelling location in geographical space, quantitative spatial models provide micro foundations for the role of market access in shaping the spatial distribution of economic activity. An earlier reduced-form literature proposed measures of market potential, such as the distance-weighted average of populations, but these measures lacked theoretical foundations. Quantitative spatial models not only provide these theoretical foundations, but highlight the role of income and relative prices, as well as population, in determining market access. Using the division of Germany in the aftermath of the Second World War and its reunification following the fall of the Iron Curtain as an exogenous source of variation, Redding and Sturm (2008) provides evidence of a causal impact of market access on the spatial distribution of city populations. Using the construction of the 19th-century railway network in the United States, Donaldson and Hornbeck (2016) provide evidence on the role of market access in determining land prices. Removing all railroads in 1890 is estimated to decrease the total value of U.S. agricultural land by 60 percent, with limited potential for mitigating these losses through feasible extensions to the canal network or improvements to country roads.

The class of quantitative spatial models discussed so far falls within the Rosen-Roback tradition, in which workers are *ex ante* identical, while deviating from from this tradition by explicitly modelling location in geographical space. To explore issues of income distribution, researchers have also developed quantitative spatial models in the Roy-Sattinger tradition, which feature multiple groups of workers that are *ex ante* heterogeneous. Diamond (2016) finds that endogenous amenities play a key role in explaining the increased geographic sorting of workers by skill in the United States from 1980-2000. Fajgelbaum and Gaubert (2020) develop a quantitative framework for evaluating optimal spatial policies, in a setting with spillovers and spatial sorting of heterogeneous workers. Designing these optimal spatial policies relates to important popular debates about the role of place-based policies and appropriate policy responses towards places that have been "left-behind" by globalization and technological change.[9]

Empirical Applications The increasing availability of geographic information systems (GIS) data has revolutionized our ability to take into account observed transport networks in computing bilateral trade costs between locations. In an empirical application to the construction of the railway network in British India, Donaldson (2018) uses a measure of lowest-cost route effective distance, in which bilateral trade costs are modeled using graph theory as depending on a set of nodes, the arcs between those nodes, and the cost of traveling along each arc. In quantitative spatial models, nodes are typically the centroids of spatial units (e.g., county centroids) and arcs are the available transportation modes between these nodes (e.g., rail, road). The cost of traveling along each arc is a vector that summarizes the per unit distance cost for each available transport mode. Given this vector and the transport network, the lowest-cost route effective distance between any pair of locations equals the cost of traveling along the least-cost path using the transport network. For any discrete set of nodes and arcs, this lowest-cost route effective distance can be computed efficiently using Dijkstra's shortest-path algorithm (Ahuja et al. 1993). In an empirical application to U.S. states, Allen and Arkolakis (2014) use related methods for continuous space based on the Fast Marching Method (FMM) of Sethian (1996). A complementary approach computes least cost path travel times for a given mode of transport and then estimates a discrete choice model across modes of transport (e.g., McFadden 1974, Ahlfeldt et al. 2015). More generally, these two approaches can be combined to estimate the demand for travel as a function of observed characteristics, including travel time, price and a range of other observed characteristics.

One of the most exciting areas of empirical applications of quantitative spatial models is to transport improvements. Allen and Arkolakis (2022) develops methods to evaluate these trans-

[9]For a broader discussion of place-based policies, see Kline and Moretti (2014) and Fajgelbaum and Gaubert (2024).

port improvements in the presence of congestion and implements these methods for the U.S. interstate highway system and the Seattle road network. Evaluating transportation improvements in the presence of congestion is challenging, because the spatial distribution of economic activity affects congestion through the induced demand for travel, but congestion in turn feeds back to affect the spatial distribution of economic activity through bilateral travel costs. A first key empirical finding for the interstate highway network is that there is a high annual rate of return on investment (measured as the ratio of annualized benefits to costs) for many links in the network. A second key empirical finding is that there is substantial heterogeneity in these annual rates of return. Although the average annual rate of return is 108 percent, it varies from over 400 percent for some important connector links in the North-East of the United States to negative values for some remote mountain links. This heterogeneity highlights the potential welfare gains from targeting transport infrastructure improvements using a network approach. More broadly, Fajgelbaum and Schaal (2020) develop a quantitative framework for evaluating optimal transport networks, which allows researchers to compute the welfare losses from deviations between the observed and optimal transport networks.

Dynamic Quantitative Spatial Models Most existing research on quantitative spatial models has considered static frameworks, because of the challenges of modelling forward-looking optimization decisions in environments with a high-dimensional state space that includes many heterogeneous locations. Nevertheless, the development of dynamic spatial models remains an active and rapidly-evolving area of research.

A first source of dynamics is costly migration decisions. Caliendo et al. (2019) develops a dynamic discrete choice model of migration, in which agents face bilateral mobility frictions, and take into account continuation values when deciding whether to move between locations. In such a dynamic setting, international trade or technology shocks that are uneven across locations have distributional consequences across workers, depending on the location in which these workers are initially located.

A second source of dynamics is capital accumulation. Kleinman et al. (2023) combines a dynamic discrete choice model of migration with forward-looking capital investments. In the presence of these two sources of dynamics, local trade or technology shocks have heteroegeneous and persistent effects across locations, because of the complementary between labor and capital in the production technology. As labor migrates away from a region experiencing a negative shock, this reduces the marginal product of capital and leads to a decline in the capital stock, which in turn reduces the marginal product of labor, and leads to further outmigration from the region experiencing the negative shock.[10]

[10]For empirical evidence on local labor market shocks from globalization and technological change, see in partic-

A third source of dynamics is endogenous innovation. Desmet et al. (2018) develops a model of innovation and growth, in which migration frictions play a key role in determining market size, innovation incentives, and the evolution of technology. Within this framework, relaxing migration frictions increases welfare around threefold, and leads to large changes in the evolution of the relative economic size of different regions of the world.[11]

Dynamic spatial models can be used to analyze the conditions under which the location of economic activity is characterized by path dependence. Allen and Donaldson (2020) develop a model, 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 location of economic activity, with the result that which equilibrium is selected depends on agents' expectations. For the estimated parameter values, small and temporary shocks have permanent effects on the location of economic activity and a substantial impact on welfare.

5 Internal City Structure

We now turn to research on internal city structure, in which the separation of residence and workplace (commuting) and the separation of residence and consumption (shopping travel) become relatively more important mechanisms for economic interactions between locations.

Monocentric Cities The traditional theoretical framework for modelling the internal structure of cities is the Alonso-Muth-Mills model (Alonso 1964, Muth 1969 and Mills 1967). In this traditional framework, cities are monocentric by assumption.[12] The model considers a city on the real line. A single final good can be traded across locations at zero cost. All employment in the city is concentrated in a single central business district (CBD). Workers choose where to reside along the real line. Workers face commuting costs that are increasing in the distance between the CBD and their residence. In a spatial equilibrium in which workers are indifferent across all residential locations, the higher commuting costs for residences further from the CBD must be compensated by lower land rents. The geographical boundary of the city is endogenous determined by the requirement that the return to using land residentially must equal to the return to land in its alternative use for agricultural production. Therefore, a central prediction of this tradi-

ular Moretti (2011), Autor et al. (2013), Autor et al. (2016) and Dix-Carneiro and Kovak (2017).

[11]See also Peters (2022) for an analysis of the impact of market size on spatial innovation and growth.

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

tional theoretical literature on internal city structure is that land rents decline monotonically with distance from the city center. This qualitative prediction is consistent with the observed feature of many cities that central locations typically have higher land rents that outlying locations.

Non-monocentric Cities Some historical cities with relatively long periods of settlement are relatively well approximated by a monocentric structure with a single CBD (e.g., London). Other cities with shorter periods of settlement display a polycentric structure, in which there are multiple business districts located in different parts of the metropolitan area (e.g., Los Angeles).

To allow for possibility of non-monocentric city structures, the the assumption that all employment is concentrated in a single central business district 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. Again a single final good can be traded across locations at zero cost. But the key difference is that the locations of both employment and residents within the city are now endogenously determined. 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.

Whether city internal structure is monocentric or polycentric in these frameworks depends on a trade-off between agglomeration and dispersion forces. In a non-monocentric city structure, there are alternating areas of commercial and residential land use. This alternating pattern reduces commuting costs, because workers on average live closer to their workplaces than in monocentric city structure. But this alternating pattern of commercial and residential land use in non-monocentric cities also reduces the spatial concentration of employment, which in turn decreases agglomeration forces relative to a monocentric city, in which all employment is concentrated in a single central business district.

Overall, this traditional theoretical literature highlights the role of agglomeration and dispersion forces in explaining observed urban land rent gradients, and in determining whether these land rent gradients are characterized by a monocentric or polycentric structure.

Quantitative Urban Models The stylized settings considered by these traditional theoretical frameworks reveal important mechanisms. But no real world city is well described by a one-dimensional line or a perfectly symmetric circle. In many urban areas, land prices can fluctuate dramatically between neighborhood that lie relative to close to one another. Often, one side of a city has higher average land prices than the other side of the city, as in the contrast between the West and East Ends of London (see Heblich et al. 2021). Some parts of a city have dense concentrations of industrial or commercial activity. Other nearby parts of the city are home to large-scale concentrations of residents. As one walks a relatively short distance with a city, land use can change notably from one block to another.

A key research breakthrough has been the development of quantitative urban models, which are able to rationalize these observed patterns in the data as an equilibrium of the model. We begin by developing a baseline quantitative urban model following Ahlfeldt et al. (2015). We consider a city that is located in a wider economy. The city is composed of a set of discrete blocks or census tracts. Each block has a supply of floor space that is determined by its geographical land area and its density of development (as measured by the ratio of floor space to land area). Floor space is owned by absentee landlords and can be used either commercially or residentially. Blocks can be either completely specialized in commercial use, completely specialized in residential use, or incompletely specialized between these two uses. We allow for a potential tax wedge between residential and commercial use. This tax wedge can differ across blocks and captures the tax equivalent of zoning regulations.

The city is populated by an endogenous measure of workers. These workers are assumed to be perfectly mobile within the city. We consider both a closed-city specification (an exogenous supply of workers) and an open-city specification (the supply of workers is endogenously determined by population mobility with the wider economy that provides a reservation level of utility). After observing idiosyncratic preference shocks for each possible pair of residence and workplace within the city, each worker chooses her preferred residence and workplace. These idiosyncratic preference shocks capture all the idiosyncratic factors that can determine why an individual worker chooses to live in one place and work in another place.

Worker utility is assumed to depend on consumption of a single homogeneous final good, consumption of residential floor space, commuting costs and residential amenities. Commuting costs increase with the travel time between the worker's residence and workplace, which is determined by the observed transport network (e.g., underground and suburban rail lines and driving times). Residential amenities capture features of a location that make it a more or less attractive place to live. These residential amenities include both natural advantages (residential fundamentals) and agglomeration forces (residential externalities). Residential fundamentals summarize the exogenous characteristics of locations that make them more or less attractive places to live independent of surrounding patterns of economic activity (e.g. leafy streets and scenic views). Residential externalities encapsulate agglomeration forces that are assumed to depend on the travel-time weighted sum of the density of residents in surrounding locations. These residential externalities can include both positive externalities (e.g., non-traded goods) and negative externalities (e.g., crime).

The homogeneous final good is assumed to be traded across location at zero cost and is chosen as the numeraire. Markets are assumed to be perfectly competitive. This final good is produced with a constant returns to scale technology using inputs of labor and commercial floor space.

Productivity is assumed to differ across locations and can depend on both natural advantages (production fundamentals) and agglomeration forces (production externalities). Production fundamentals summarize the exogenous characteristics of locations that make them more or less attractive places to produce independent of surrounding patterns of economic activity (e.g., access to natural water). Production externalities encapsulate agglomeration forces that are assumed to depend on the travel time weighted sum of surrounding employment density (capturing, for example, knowledge spillovers).

The resulting quantitative urban framework allows for a rich range of differences in characteristics across locations in order to connect directly with the observed data. The internal structure of economic activity within the city is shaped by productivities, amenities and the transport network. High productivity increases the marginal productivities of labor and land in a location, which raises wages and the price of commercial floor space, thereby reallocating floor space towards commercial use. In contrast, high amenities raise the utility of living in a location, which attracts residents, thereby bidding up the price of residential floor space, and reallocating floor space towards residential use. Transportation networks facilitate the separation of workplace and residence to allow workers to take advantage of these differences in productivity and amenities. As a result, some locations can specialize as workplaces, while other locations specialize as residences. These differences in productivity and amenities are influenced by both first-nature geography (production and residential fundamentals) and second-nature geography (production and residential externalities).

Properties of Quantitative Urban Models If workers idiosyncratic preferences for locations are drawn from an extreme value distribution, this quantitative urban model implies a constant elasticity commuting gravity equation, in which bilateral commuting flows depend on bilateral travel costs, origin characteristics and destination characteristics. A large empirical literature finds that this gravity equation provides a good approximation to observed bilateral commuting flows, as summarized in Fortheringham and O'Kelly (1989) and McDonald and McMillen (2010). This gravity equation provides microfoundations for measures of residents' and workers' commuting market access, which play an analogous role to goods market access in quantitative models of systems of cities or regions (Ahlfeldt et al. 2015, Redding 2022a, Tsivanidis 2024). The number of residents in each location can be expressed as a function of residential amenities, the cost of living and residents' commuting market access, which depends on the travel-time weighted sum of wages in each workplace. Similarly, the number of workers in each location can be expressed as a function of the wage and workers' commuting market access, which depends on the travel-time weighted sum of the amenity-adjusted cost of living in each residence.

The extreme value specification for idiosyncratic preferences has a further implication that expected utility is equalized across all pairs of residences and workplaces. The intuition for this result is as follows. Residence-workplace pairs with desirable economic characteristics (high wages and low costs of living) attract additional commuters with lower realizations for idiosyncratic preferences, until expected utility (incorporating these idiosyncratic preferences) is equalized across all residence-workplace pairs. In an open-city specification, this common level of expected utility is determined by the reservation level of utility in the wider economy. Therefore, changes in transport infrastructure or other public policies affect total city population and the welfare of landlords, but leave expected worker utility unchanged. In contrast, in a closed-city specification, these changes in public policies affect both the welfare of landlords and expected worker utility, with total city population unchanged.

Consistent with our discussion of quantitative models of systems of cities or regions above, there exists an entire class of quantitative urban models that are isomorphic with respect to their gravity equation predictions (Heblich et al. 2020). This class is again defined by a constant elasticity structure, in which economic outcomes in one location are a constant elasticity function of economic outcomes in all locations weighted by a network of bilateral frictions between locations. Within this class of models, sufficient conditions for the existence and uniqueness of the equilibrium can be derived, which depend only on the model's structural parameters (elasticities), and hence hold for any network of bilateral interactions (Allen et al. 2024). In the special case of this class of models, in which residential and production externalities depend only on own location characteristics (and do not spillover across locations), changes in workers, residents and land rents can be expressed up to a first-order approximation in terms of changes in workers' and residents' market access, as shown in Tsivanidis (2024).

Quantitative urban models are again typically invertible. Given the structural parameters (elasticities) and the observed endogenous variables in the data, one can recover unique values of the unobserved structural residuals (production fundamentals, residential fundamentals and the ratio of floor space to land area) that exactly rationalize the observed data as an equilibrium. This invertibility property can hold even if there are multiple equilibria, because this procedure conditions on the observed equilibrium in the data. Intuitively, the observed endogenous variables and the equilibrium conditions of the model can together contain enough information to uniquely determine these structural residuals, even though there could have been another (unobserved) equilibrium given the same value of the model's parameters. Therefore, the parameters of quantitative urban models can be estimated even in the presence of multiple equilibrium, using orthogonality conditions on these structural residuals (e.g., Ahlfeldt et al. 2015).

Quantitative urban models also typically lend themselves to solving for exact-hat algebra counterfactuals. Given the observed endogenous variables in the initial equilibrium in the data and assumed changes in exogenous location characteristics (e.g., reductions in travel costs from a transport improvement), one can solve for a counterfactual equilibrium without requiring information on the unobserved location characteristics in the initial equilibrium. Since quantitative urban models are able to rationalize the rich asymmetric patterns of economic activity observed in the data, they can be used to undertake counterfactuals for empirically-relevant public policy interventions, such as the opening of a new subway line between between a specific set of locations within a city. Therefore, quantitative urban models provide a useful supplement to conventional cost-benefit approaches for evaluating transport improvements that takes into account the resulting general equilibrium reorganization of economic activity. Multiple equilibria pose more of a problem for undertaking counterfactuals than for parameter estimation, because a researcher must specify an equilibrium selection rule when solving for a counterfactual equilibrium. Developing robust methods for undertaking counterfactuals in the presence of multiple equilibria remains an exciting area for further research.

We have so far discussed quantitative models of systems of cities and internal city structure separately. For some research questions, it may make sense to focus on the distribution of economic activity across cities, abstracting from internal city structure. For other research questions, it may be more reasonable to concentrate on the internal structure of economic activity within a single city, abstracting from spillover effects on other cities. However, for yet other research questions, it may be important to incorporate economic interactions both across and within cities. 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 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. One of the key implications of this framework is that elasticity of local economic activity with respect to local shocks (e..g., productivity shocks) can be heterogeneous across locations, depending on the network of connections to other locations in goods and commuting markets.

Although we have concentrated on quantitative models of internal city structure in which workers are *ex ante* homogenous, variants of these models can be developed in the Roy-Sattinger tradition, in which workers are *ex ante* heterogenous and there is endogenous spatial sorting (Almagro and Domínguez-lino 2019, Couture et al. 2024, Tsivanidis 2024, Redding and Sturm 2024). Introducing *ex ante* worker heterogeneity is central to thinking about issues of segregation (by income, race and ethnicity) and gentrification. In the presence of this heterogeneity, public policy interventions such as transport infrastructure improvements have distributional consequences across these different groups of workers. Furthermore, these distributional consequences depend not only on who is initially living in each location, but on how the public policy interventions change endogenous patterns of spatial sorting. For example, policies to revitalize low-income neighborhoods need not benefit the initial residents of those neighborhoods, because these policies can lead to gentrification, as higher-income residents move into the neighborhood. This gentrification bids up rents and house prices, which can either help or hurt initial low-income residents, depending on whether they are owner-occupiers or renters.

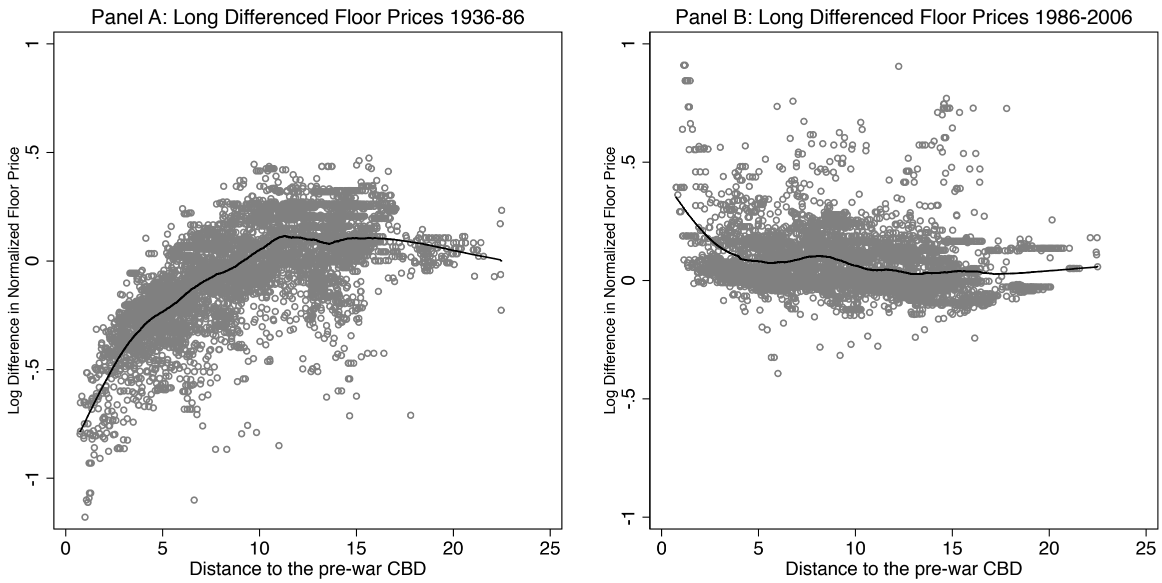
In modelling the internal structure of cities, quantitative urban models have largely focused on the separation of residence and workplace (and commuting decisions). However, another striking feature of cities is the separation of residence and locations of consumption (Miyauchi et al. 2022). Millions of people move each day through the complex transportation networks of large cities. Access to both employment opportunities and consumption possibilities are some of the key attractions of living in these large metropolitan areas. Additionally, most existing research on quantitative models of internal city structure has focused on static specifications, but developing dynamic specifications that allow for gradual adjustment in response to shocks remains an exciting area for further research.

Empirical Applications We now illustrate two empirical applications of quantitative urban models, one to estimate the strength of agglomeration and dispersion forces, and the other to evaluate the impact of transport infrastructure improvements.

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 an exogenous source of variation in surrounding economic activity to estimate the strength of agglomeration forces. Following the city's division, the areas of West Berlin close to the pre-war CBD in East Berlin experience larger reductions in access to nearby economic activity than other locations in West Berlin. In the quantitative urban model discussed above, this leads to larger reductions in commuting market access and production and residential externalities close to the pre-war CBD. To restore equilibrium, both employment and residents reallocate away from the areas of West Berlin close to the pre-war CBD, until wages and the price of floor space in these locations fall, such that there zero profits in each location with positive production of the final good, workers are indifferent across all locations with positive quantities of residents, and there is no-arbitrage between the use of floor space for commercial and residential activity. Consistent with these predictions, Figure 2 shows a strong negative relationship between the change in the price of floor space and distance from the pre-war CBD after the division of (1936-88). This pattern is reversed after the reunification of the city (1988-2006), with a marked positive relationship between the change in the price of floor space and distance from the pre-war CBD.

To examine whether the quantitative urban model developed above can successfully account for the observed changes in the spatial distribution of land prices, employment and residents in the data, Ahlfeldt et al. (2015) structurally estimate the model's parameters using the generalized

Figure 2: Changes in the Price of Floor Space in West Berlin Following Division and Reunification



Note: Changes in the Price of Floor Space in each city block in West Berlin following division (1936-88) and reunification (1988-2006); distance from the pre-war central business district (CBD) is measured as straightline distance from the intersection of Friedrich Strasse and Leipziger Strasse, close to the underground station City Center ("Stadtmitte"). Source: Ahlfeldt et al. (2015).

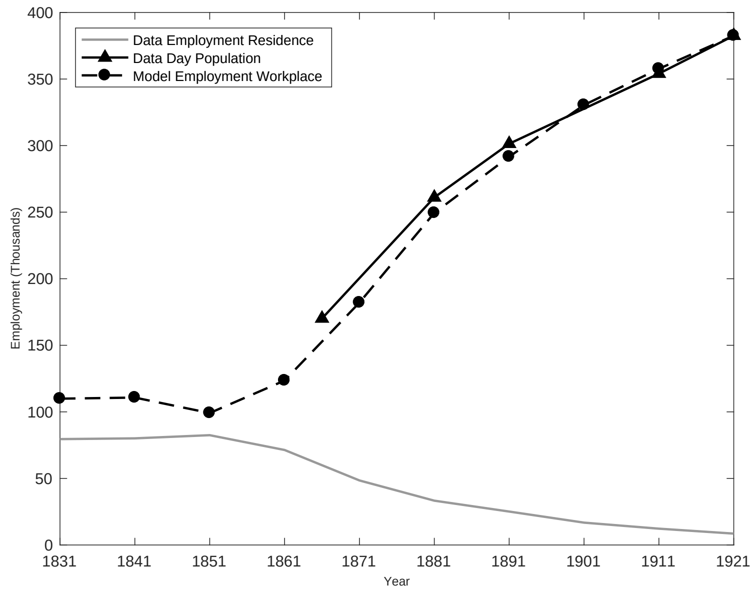
method of moments (GMM). The identifying assumption in this estimation is that the log changes in production and residential fundamentals in each block in West Berlin are uncorrelated with the change in the surrounding concentration of economic activity caused by the city's division and reunification. The estimated parameters imply substantial and highly localized agglomeration forces. The estimated elasticity of productivity to employment density is 0.07, while the estimated elasticity of amenities to residential density is 0.15. The estimated exponential rates of decay of production and residential externalities with travel time are 0.36 and 0.76, respectively. Based on these estimated parameter values, production and residential externalities fall to close to zero by around 10 minutes of travel time, or approximately 0.83 kilometers by foot (average speed equal to 5 kilometers per hour) and approximately 4 kilometers by railway (average speed equal to 25 kilometers per hour).

Heblich et al. (2020) examines the impact of transport infrastructure on the internal structure of cities using the mid-19th century invention of steam railways as a natural experiment. The key idea is that the slow travel times by human or horse power ensured that most people lived close to their workplace when these were the main travel modes. In contrast, steam railways substantially increased the distance that could be travelled in a given amount of time, and hence enabled the first large-scale separation of workplace and residence.

Following the invention of the steam passenger railway, there is evidence of a large-scale change in the organization of economic activity within the metropolitan area of Greater London.

Figure 3 shows residential (night-time) population and employment by workplace (day-time population) in the historical center of the metropolitan area, which is termed the City of London and roughly corresponds to the Roman city (the Square Mile). Shortly after the invention of the first steam passenger railway (the London and Greenwich Railway in 1836), there is a sharp decline in residential population and a steep increase in employment by workplace, as residents dispersed from the center to the suburbs, and the center specialized as a workplace.

Figure 3: Night and Day Population in the Historical City of London



Note: The City of London is the historical center of the metropolitan area, corresponding approximately to the boundaries of the old Roman city (the Square Mile); "Data Employment Residence" 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 employment by workplace from the population census for 1921. Source: Heblich et al. (2020).

To assess the ability of a quantitative urban model to account for this change in patterns of specialization, the paper undertakes counterfactuals for the removal of London's railway network, starting at the end of the sample in 1921, which is the first year for which data on bilateral commuting flows are available in the population census. This empirical approach conditions on observed changes in residential population and property values (to control for changes in amenities and productivities) and generates predictions for employment by workplace going backwards in time, using estimates of changes in commuting costs from the invention of the steam passenger railway. As shown in Figure 3, the estimated model is quantitatively successful in capturing this large-scale change in internal city structure.

Undertaking counterfactuals for the impact of the new transport technology, holding the exogenous components of productivity and amenities constant, the change in the net present value of land and buildings exceeds historical estimates of railway construction costs. Therefore, the large-scale investments in London's 19th-century railway network can be justified based on their economic impact. Introducing agglomeration forces and/or allowing the supply of land and buildings to endogenously respond to changes in the price of floor space substantially magnifies the new transportation technology's economic impact. Therefore, these findings highlight the relevance of complementary investments in building structures and agglomeration forces for cost-benefit analyses of transport infrastructure improvements.

6 Conclusions

Spatial economics is concerned with the determinants and effects of the location of economic activity in space. Two main lines of research can be distinguished, one concerned with systems of cities or regions, and the other concerned with internal city structure.

Within each of these lines of research, the traditional theoretical literature in spatial economics considered stylized settings, such as two symmetric regions or a one-dimensional line. A major breakthrough in recent research is the development of quantitative spatial 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 they remain sufficiently tractable as to permit an analytical characterization of their theoretical properties, such as the existence and uniqueness of the equilibrium. With only a small number of parameters to be estimated, these models lend themselves to transparent identication. Since they rationalize the observed distribution of economic activity in the data, they can be used to undertake counterfactuals for the impact of empirically-realistic public-policy interventions on this observed distribution.

Among the insights that have emerged from these quantitative spatial models are the role of goods and commuting market access in determining location choices; the conditions under which the location of economic activity is characterized by multiple equilibria; the circumstances under which temporary shocks can have permanent effects (hysteresis or path dependence); the heterogeneous and persistent impact of local shocks; the magnitude and spatial decay of agglomeration economics; and the role of both agglomeration forces and endogenous changes in land use in shaping the impact of transport infrastructure improvements.

Spatial economics in recent years has benefited from the simultaneous development of new theoretical techniques, new sources of geographic information systems (GIS) data, rapid advances in computing power, machine learning and artificial intelligence, and renewed public policy interest in infrastructure and appropriate policies towards places "left-behind" by globalization and technology. Looking ahead, there remain many exciting opportunities to combine quantitative spatial models with new sources of big data containing geographic information, including ride-hailing (e.g., Uber and Lyft), smartphone data, firm-to-firm VAT sales, credit card transactions, public transportation fare cards, and satellite imaging data, among others.

References

Abdel-Rahman, H. M. and A. Anas (2004): "Theories of Systems of Cities," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J.-F. Thisse, Amsterdam: Elsevier North Holland, vol. 4, 2293–2339.

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015): "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993): *Network Flows: Theory, Algorithms, and Applications*, Upper Saddle River, NJ: Prentice-Hall.

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

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

Allen, T., C. Arkolakis, and X. Li (2024): "On the Equilibrium Properties of Spatial Models," *American Economic Review: Insights*, forthcoming.

Allen, T., C. Arkolakis, and Y. Takahashi (2019): "Universal Gravity," *Journal of Political Economy*, 128, 393–433.

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

Almagro, M. and T. Domínguez-lino (2019): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," New York University, mimeograph.

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

Anas, A., R. Arnott, and K. A. Small (1998): "Urban Spatial Structure," *Journal of Economic Literature*, 36, 1426–1464.

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

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.

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

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.

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.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2018): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *Review of Economic Studies*, 85, 2042–2096.

Chen, C.-m. and K. Peng (2020): "New Economic Geography," *Oxford Research Encyclopedia of Economics and Finance*, https://doi.org/10.1093/acrefore/9780190625979.013.338.

Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2024): "Income Growth and the Distributional Effects of Urban Spatial Sorting," *Review of Economic Studies*, 91, 858–898.

Davis, D. R. and J. I. Dingel (2020): "The Comparative Advantage of Cities," *Journal of International Economics*, 123, 103291.

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

De La Roca, J. and D. Puga (2017): "Learning by Working in Big Cities," *Review of Economic Studies*, 84, 106–142.

Desmet, K. and J. V. Henderson (2015): "The Geography of Development Within Countries," in *Handbook of Regional and Urban Economics*, Amsterdam: North-Holland, vol. 5, chap. 22, 1457–1517.

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

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.

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

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.

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

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

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

Fujita, M. (2020): "General Equilibrium Theory of Land," *Oxford Research Encyclopedia of Economics and Finance*, https://doi.org/10.1093/acrefore/9780190625979.013.549.

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.

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.

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.

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

Jacobs, J. (1961): *The Death and Life of Great American Cities*, New York: Random House.

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

Kline, P. and E. Moretti (2014): "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Policies," *Annual Review of Economics*, 6, 629–662.

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.

Krugman, P. and A. Venables (1995): "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, 110, 857–880.

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

Lucas, R. E. and E. Rossi-Hansberg (2002): "On the Internal Structure of Cities," *Econometrica*, 70, 1445–1476.

Marshall, A. (1920): *Principles of Economics*, London: Macmillan.

McDonald, J. and D. McMillen (2010): *Urban Economics and Real Estate: Theory and Policy*, Hoboken, NJ: John Wiley & Sons.

McFadden, D. (1974): "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3, 303–328.

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

Michaels, G., F. Rauch, and S. J. Redding (2012): "Urbanization and Structural Transformation," *Quarterly Journal of Economics*, 127, 535–586.

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

Miyauchi, Y., K. Nakajima, and S. Redding (2022): "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data," *NBER Working Paper*, 28497.

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

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.

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

Proost, S. and J.-F. Thisse (2019): "What Can be Learned from Spatial Economics?" *Journal of Economic Literature*, 57, 575–643.

Redding, S. J. (2016): "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148–167.

——— (2022a): "Suburbanization in the United States 1970-2000," *Economica*, 89, S110–S136, 100th Anniversary Special Issue.

——— (2022b): "Trade and Geography," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Amsterdam: Elsevier, chap. 3, 147–217.

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

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

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 Wartime Destruction in 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.

Roback, J. (1982): "Wages, Rents and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

Rosen, S. (1979): "Wages-based Indexes of Urban Quality of Life," in *Current Issues in Urban Economics*, ed. by P. Mieszkowski and M. Straszheim, Baltimore: John Hopkins University Press.

Rossi-Hansberg, E. (2019): "Geography of Growth and Development," *Oxford Research Encyclopedia of Economics and Finance*, https://doi.org/10.1093/acrefore/9780190625979.013.273.

Roy, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–46.

Sattinger, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31, 831–80.

Sethian, J. (1996): "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences*, 93, 1591–1595.

Thisse, J.-F. (2019): "Economics of Agglomeration," *Oxford Research Encyclopedia of Economics and Finance*, https://doi.org/10.1093/acrefore/9780190625979.013.152.

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

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

Venables, A. J. (2019): "Economic Geography and Trade," *Oxford Research Encyclopedia of Economics and Finance*, https://doi.org/10.1093/acrefore/9780190625979.013.332.