# ANNUAL REVIEWS

## Quantitative Spatial Economics

### Stephen J. Redding and Esteban Rossi-Hansberg

Department of Economics and Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08544; email: reddings@princeton.edu, erossi@princeton.edu

ANNUAL
REVIEWS **Further**

Click here to view this article's online features:
• Download figures as PPT slides
• Navigate linked references
• Download citations
• Explore related articles
• Search keywords

## Keywords

agglomeration, cities, economic geography, quantitative models, spatial economics

## Abstract

The observed uneven distribution of economic activity across space is influenced by variation in exogenous geographical characteristics and endogenous interactions between agents in goods and factor markets. Until the past decade, the theoretical literature on economic geography had focused on stylized settings that could not easily be taken to the data. This article reviews more recent research that has developed quantitative models of economic geography. These models are rich enough to speak to first-order features of the data, such as many heterogeneous locations and gravity equation relationships for trade and commuting. At the same time, these models are sufficiently tractable to undertake realistic counterfactual exercises to study the effect of changes in amenities, productivity, and public policy interventions such as transport infrastructure investments. We provide an extensive taxonomy of the different building blocks of these quantitative spatial models and discuss their main properties and quantification.

## 1. INTRODUCTION

Economic activity is highly unevenly distributed across space, as reflected by the existence of cities and the concentration of economic functions in specific locations within cities, such as Manhattan in New York and the Square Mile in London. The relative strengths of the agglomeration and dispersion forces that underlie these concentrations of economic activity are central to a range of economic issues. The delicate balance between these two sets of forces helps to determine, for example, the incomes of mobile and immobile factors, the magnitude of investments, and both city and aggregate productivity. The impact of public policies differentiated by location (place-based policies) and of transport infrastructure investments, local taxation, and land regulation is crucially determined by how these policies affect the equilibrium balance between these centripetal and centrifugal forces.

The complexity of modeling spatial interactions between agents has meant that the theoretical literature on economic geography has traditionally focused on stylized settings—such as a small number of symmetric locations—that cannot easily be taken to the data. More recent research has developed quantitative models of the spatial distribution of economic activity. These models are rich enough to incorporate first-order features of the data, such as large numbers of locations with heterogeneous geography, productivity, amenities, and local factors, as well as trade and commuting costs. They are also able to incorporate key interactions between locations, such as trade in goods, migration, and commuting. At the same time, these models are sufficiently tractable to enable quantitative counterfactuals to evaluate empirically meaningful policies and counterfactual scenarios. In this article, we review this recent body of research on quantitative spatial economics, highlighting the key new theoretical and empirical insights and discussing remaining challenges and potential areas for further research. We provide an extensive taxonomy of the different building blocks of quantitative spatial models used in the literature and discuss their properties.

We interpret the field of economic geography as the study of the interactions between economic agents across geographic space. This field, in contrast to the study of international trade, typically assumes economic agents to be geographically mobile. Early theoretical research on new economic geography [as synthesized by Fujita et al. (1999), Fujita & Thisse (2002), and Baldwin et al. (2003)] concentrated on formalizing mechanisms for agglomeration and cumulative causation, including forward and backward linkages between economic activities. This literature stressed the combination of love of variety, increasing returns to scale, and transport costs as a mechanism for agglomeration. This mechanism provided a fundamental theoretical explanation for the emergence of an uneven distribution of economic activity even on a featureless plain of ex ante identical locations and highlighted the potential for multiple equilibria in location choices. However, the complexity of these theoretical models limited the analysis to stylized spatial settings such as a limited number of locations, a circle, or a line. Therefore, although this early theoretical literature stimulated a wave of empirical research, much of this empirical research was reduced form in nature. As a result, the mapping from the model to the empirical specification was often unclear, and it was difficult to give a structural interpretation to the estimated reduced-form coefficients. In the absence of such a structural interpretation, the coefficients of these reduced-form relationships need not be invariant to policy intervention (e.g., the Lucas critique). Furthermore, the extent to which theoretical results for stylized spatial settings would generalize qualitatively and quantitatively to more realistic environments is unclear (for reviews of the earlier theoretical and empirical literature on new economic geography, see, e.g., Overman et al. 2003; Redding 2010, 2011).

Following the introduction of quantitative models of international trade [in particular that of Eaton & Kortum (2002)], research in economic geography has developed a quantitative framework that connects closely to the observed data. In contrast to the previous theoretical work, this research does not aim to provide a fundamental explanation for the agglomeration of economic activity, but rather to provide an empirically relevant quantitative model to perform general equilibrium counterfactual policy exercises. Agglomeration in these models is simply the result of exogenous local characteristics augmented by endogenous economic mechanisms. These frameworks can accommodate many asymmetric locations that can differ from one another in terms of their productivity, amenities, and transport and mobility connections to one another. The analysis can admit many sectors with different factor intensities and observed input–output linkages between them. Furthermore, the same quantitative framework can be derived from an entire class of theoretical models of economic geography, highlighting the robustness of this framework to perturbations in theoretical assumptions. These theoretical models differ in assumptions (e.g., monopolistic competition versus perfect competition) and mechanisms (e.g., technological versus pecuniary externalities), in the structural interpretations of some reduced-form coefficients (e.g., whether the elasticity of trade with respect to trade costs corresponds to the elasticity of substitution or the dispersion of productivity), and in some of their predictions (e.g., when factors are mobile across locations, trade cost reductions have different effects on the spatial distribution of economic activity in models of constant versus increasing returns to scale). Nonetheless, these models are isomorphic to one another for a series of predictions (e.g., the gravity equation for bilateral trade and commuting, in which interactions between two locations increase with the product of their size and decrease with the distance between them).

The close connection between model and data in this quantitative research has a number of advantages. First, by accommodating many regions and a rich geography of trade costs, these models provide microfoundations for central features of the data. Second, by allowing for many regions that can differ in their productivity and amenities, as well as a number of other characteristics, these models are sufficiently rich to explain the observed data as an equilibrium of the model. These models are typically exactly identified, such that there exists a one-to-one mapping from the observed data on the endogenous variables of the model (e.g., employment and wages) to the exogenous primitives or structural fundamentals of the model (e.g., productivity and amenities). Therefore, this mapping can be inverted to identify the unique values of the estimated structural fundamentals that exactly rationalize the observed data as an equilibrium. Having recovered these estimated structural fundamentals, the observed variation in the data can be decomposed within the model into the contributions of each of the fundamentals. Inevitably, this analysis is conditional on the assumed model, and different models generally imply different estimated structural fundamentals and decompositions.

The cost of enriching theoretical models to connect more closely to the data is typically a loss of analytical tractability. However, a major contribution of this quantitative economic geography literature has been to preserve sufficient analytical tractability to provide conditions under which there exists a unique spatial equilibrium distribution of economic activity and to permit some analytical comparative statics (see, in particular, Allen & Arkolakis 2014, Allen et al. 2015). Another central advantage of this structural empirical approach relative to the earlier reduced-form empirical literature is the ability to undertake counterfactuals for policy interventions or other out-of-sample changes in model primitives. For these exercises to be valid, one must assume that the identified structural fundamentals are stable and invariant to the analyzed policy interventions (for a general review of structural estimation approaches in urban economies, see Holmes & Sieg 2015). Under this assumption, these counterfactuals yield general equilibrium predictions for the spatial

distribution of economic activity, which take full account of all the complex spatial interactions between locations.[1] These interactions and general equilibrium effects are typically not identified in reduced-form difference-in-differences approaches because differencing between the treatment and control group eliminates any effect that is common to both groups. Thus, a key implication of this analysis is that locations are not independent observations in a cross-sectional regression but rather are systematically linked to one another through trade, commuting, and migration flows. Not recognizing this interdependence in reduced-form empirical analysis can lead to significant biases and substantial heterogeneity in treatment effects that threaten the external validity of the results (see, e.g., Monte et al. 2015). Finally, the use of the model's structure makes it possible to compute the counterfactual change in welfare, which is usually unobservable in reduced-form approaches and yet is typically the object of ultimate interest for policy intervention.

Quantitative spatial models share many similarities with the earlier theoretical literature on economic geography. The mechanisms are typically the same, although there is greater scope to combine multiple mechanisms within a single framework. The broad questions are also largely the same. For example, how important is physical geography (e.g., mountains, coasts) versus economic geography (the location of agents relative to one another)? What is the impact of reductions in transport costs on the spatial distribution of economic activity? However, there are three key differences in focus and specificity relative to the earlier theoretical research. First, this new research connects in a meaningful way with the observed data and thus provides quantitative rather than qualitative answers to these questions. The emphasis is therefore on combining, measuring, and quantifying existing theoretical mechanisms. Second and relatedly, this work identifies the key structural parameters that need to be estimated to undertake such quantification. Third, the meaningful connection with the data permits specificity in addressing counterfactual questions of interest to policy makers: For example, if a railroad is built between these cities in this country at this time, what is the quantitative effect on these particular regions, sectors, and factors of production? Not only can this specificity address important policy questions, but the ability to contrast the model's predictions with real-life policy allows us to gauge the empirical importance of different theoretical mechanisms.

In addition to the quantitative evaluation of specific counterfactuals and policy exercises, the existing research on quantitative spatial models has yielded two main sets of general insights that are not present in the earlier literature on economic geography. The first set of general insights are methodological. These include an improved understanding of the conditions for the existence and uniqueness of equilibrium in economic geography models, the conditions under which these models can be inverted to separate out the contributions of physical and economic geography, and methods for undertaking counterfactuals to evaluate comparative statics with respect to changes in the model's parameters. Perhaps even more important is that the literature has provided a set of model components that allow us to introduce, in a unified theoretical framework, a large variety of agglomeration and congestion forces in a simple and practical way. Together, these insights facilitate the quantification and measurement that are at the heart of this body of research.

The second set of general insights is substantive in terms of the quantitative importance of theoretical mechanisms. First, market access is an empirically relevant causal determinant of the spatial distribution of activity. This mechanism can account for the observed decline of approximately one-third in the relative size of West German cities close to the new border with East Germany

---

[1]Quantitative spatial models share some features with the earlier computable general equilibrium literature in international trade, as reviewed by Shoven & Whalley (2005). The key difference is that quantitative spatial models can both obtain sharp analytical results and permit transparent counterfactuals, in addition to the focus on economic geography rather than international trade.

following the division of Germany after World War II (Redding & Sturm 2008). Similarly, assuming that the railroads constructed up to 1890 had not been built, the value of agricultural land in the United States would have been reduced by approximately 60%, with limited potential for mitigating these losses through feasible extensions to the canal network (Donaldson & Hornbeck 2016). Second, canonical models of urban economics (e.g., Fujita & Ogawa 1982, Lucas & Rossi-Hansberg 2002) can account quantitatively for the observed gradients of economic activity within cities (e.g., Ahlfeldt et al. 2015). The estimated parameter values imply substantial and highly localized agglomeration externalities for both production and residential choices.

Third, the local incidence of economic shocks is shaped in an important way by spatial linkages in goods and factor markets, which give rise to heterogeneous treatment effects of changes in the local economic environment (Monte et al. 2015) as well as heterogeneous aggregate implications of local shocks (Caliendo et al. 2014). Fourth, the distribution of economic activity across cities and regions is shaped in a quantitatively important way not only by productivity and amenity differences but also by a number of other spatial frictions, such as local infrastructure and governance (e.g., Desmet & Rossi-Hansberg 2013, Behrens et al. 2014). Fifth, the distribution of economic activity shapes the dynamics of local innovation and growth by determining the market size of firms. This link is quantitatively relevant for understanding the evolution of the spatial distribution of economic activity over time (e.g., Desmet & Rossi-Hansberg 2014) and the counterfactual dynamic response of the economy to global migration, trade policy changes, and global shocks such as climate change (e.g., Desmet & Rossi-Hansberg 2015, Desmet et al. 2016, Nagy 2016).

The remainder of this review is structured as follows. In Section 2, we outline a menu of building blocks or model components that can be combined in different ways in quantitative spatial models. We discuss the criteria for choosing between these building blocks and the trade-offs involved. In Section 3, we develop an example of such a quantitative spatial model based on a canonical new economic geography model. In this framework, a system of cities and regions is linked together through costly goods trade and labor mobility. We solve the model numerically and perform policy exercises that reduce external and internal trade costs. In Section 4, we provide another example based on the canonical urban model, which focuses instead on the internal structure of economic activity within a city. In both cases, we discuss the analytical characterization of the existence and uniqueness of the equilibrium, the inversion of the model to recover unobserved location characteristics from observed endogenous variables, and the use of the model to undertake counterfactuals for transport infrastructure improvements or other policy interventions. In Section 5, we review the empirical evidence on the predictions of these models. Section 6 concludes and discusses some potential areas for further research.

## 2. A MENU OF QUANTITATIVE SPATIAL MODELS

Each of the quantitative spatial models considered in this review makes implicit or explicit assumptions about a number of building blocks or model components. In this section, we review the key building blocks and menu of assumptions in existing studies. In addition to preferences, production technologies, endowments, and market structure, these building blocks include the three main reasons why agents' location relative to one another in geographic space matters: frictions for the movement of goods, ideas, and people. Combining different building blocks and assumptions allows researchers to capture different dimensions of the spatial economy. We discuss the criteria for selecting building blocks and choosing between assumptions for each block. We provide examples of existing studies that have selected particular items from the menu. In Sections 3 and 4, we pick particular combinations of building blocks and assumptions and show how the resulting framework can be used for the quantitative analysis of the spatial economy.

## 2.1. Preferences

Assumptions about preferences play a central role in shaping consumers' location decisions. Five main sets of assumptions about preferences can be distinguished.

**2.1.1. Homogeneous versus differentiated goods.** Following Krugman (1991a,b), new economic geography models such as that of Helpman (1998) emphasize firm product differentiation and consumers' love of variety. More recent research has shown that similar properties hold in models in which goods are homogeneous (e.g., Eaton & Kortum 2002) and labor is mobile (e.g., Rossi-Hansberg 2005, Redding 2016) or models in which goods are differentiated only by country of origin (e.g., Armington 1969) and labor is mobile (e.g., Allen & Arkolakis 2014).

**2.1.2. Single versus multiple sectors.** To preserve analytical tractability, theoretical models of economic geography have often restricted attention to a single production sector (e.g., Helpman 1998) or distinguished between aggregate sectors such as agriculture and manufacturing (e.g., Krugman 1991b, Puga 1999). With the development of tractable quantitative models and efficient computational methods, researchers have become able to handle multiple disaggregated sectors (e.g., Caliendo et al. 2014). This introduction of multiple sectors permits the analysis of issues such as structural transformation and development, as in the work of Desmet & Rossi-Hansberg (2014), Fajgelbaum & Redding (2014), Coşar & Fajgelbaum (2016), and Nagy (2016).

**2.1.3. Exogenous and endogenous amenities.** Early new economic geography models, such as that of Krugman (1991b), assumed a featureless plain in which locations were ex ante identical, and ex post differences in the spatial distribution of economic activity emerged endogenously. To incorporate real-world differences across locations (e.g., climate, access to water, and other characteristics of physical geography), quantitative models typically allow for exogenous differences in amenities across locations. In the spirit of the seminal work of Rosen (1979) and Roback (1982), amenities are understood as any characteristic that makes a location a more or less desirable place of residence, as examined empirically for US metropolitan areas by Albouy (2016). Several studies have also argued that, to match the response of the local economy to external shocks, it is important to allow for endogenous amenities (e.g., low crime rates) as well as exogenous amenities (e.g., scenic views), as in the work of Ahlfeldt et al. (2015) and Diamond (2016), among others.

**2.1.4. Fixed local factors in utility.** The presence of fixed factors in utility, such as residential land, acts as a congestion or dispersion force (see, e.g., Helpman 1998, Monte et al. 2015).

**2.1.5. Common versus idiosyncratic preferences.** A standard benchmark in the quantitative spatial literature is the assumption that agents have common preferences and are perfectly mobile across locations. In this case, a no-arbitrage condition ensures that real wages are equalized across locations, and each location faces a perfectly elastic supply of labor at the common real wage. A tractable approach to departing from this benchmark is to allow agents to have idiosyncratic preferences for each location that are drawn from an extreme value distribution. In this case, individual agents pick their preferred location, and each of these locations faces a supply curve for labor that is upward sloping in real wages, as higher real incomes have to be paid to attract workers with lower idiosyncratic preferences. The elasticity of labor supply with respect to the real wage is determined by the degree of heterogeneity in agents' preferences (see, e.g., Artuç et al. 2010, Grogger & Hanson 2011, Kennan & Walker 2011, Busso et al. 2013). Although much of the literature has focused on idiosyncratic differences in preferences across locations, models with

idiosyncratic differences in worker productivity across locations have many similar properties, although with different predictions for wages (see, e.g., Galle et al. 2015).

## 2.2. Production Technology

Assumptions about production technology critically influence firms' location decisions. Four main sets of assumptions concerning production technology can be distinguished.

**2.2.1. Constant versus increasing returns.** Following Krugman (1991a,b), the new economic geography literature assumes increasing returns to scale, which generates the potential for a self-reinforcing process of agglomeration (often termed cumulative causation) and the emergence of multiple equilibrium spatial allocations even on a featureless plain of ex ante identical locations. However, even under the assumption of constant returns to scale, agents' locations relative to one another in geographic space have implications for prices and allocations. Indeed, there are conditions under which models of constant returns to scale and transport costs are isomorphic for endogenous outcomes of interest to those of models with local increasing returns to scale (see, in particular, Allen & Arkolakis 2014). Both Armington differentiation by location of origin (e.g., Armington 1969) and Ricardian technology differences (e.g., Eaton & Kortum 2002) can provide alternative mechanisms for specialization from the love of variety and increasing returns to scale in new economic geography models.

**2.2.2. Exogenous and endogenous productivity differences.** Although early theoretical models of economic geography focused almost exclusively on endogenous production externalities (e.g., knowledge spillovers), a long intellectual tradition in international trade emphasizes exogenous productivity differences (e.g., mineral resources), and quantitative spatial models have typically found it necessary to allow for such exogenous differences across locations to rationalize the observed employment and income data (e.g., Allen & Arkolakis 2014, Caliendo et al. 2014, Ahlfeldt et al. 2015, Desmet et al. 2016).

**2.2.3. Input–output linkages.** Input–output linkages play a key role in determining how productivity shocks in a particular sector or region spread through the wider economy and shape local multipliers (i.e., the extent to which an increase in expenditure in one sector leads to more than proportionate increases in overall expenditure through increased demand in other sectors). Such input–output linkages provide an additional mechanism for agglomeration (e.g., Krugman & Venables 1995), and the observed linkages between sectors in real-world input–output matrices can now be incorporated in a relatively tractable way into quantitative spatial models (following Caliendo et al. 2014).

**2.2.4. Fixed local factors in production.** The presence of fixed local factors in production, such as commercial land, acts as a congestion force (e.g., Rossi-Hansberg 2005, Ahlfeldt et al. 2015).

## 2.3. Costs of Trading Goods

Several mechanisms can explain the importance of the location of agents relative to one another in quantitative spatial models. The first of these mechanisms is the cost of trading goods. Four main sets of assumptions concerning the costs of trading goods can be delineated.

**2.3.1. Variable versus fixed trade costs.** A widespread assumption used for analytical tractability is that of iceberg variable transport costs, whereby $d_{ni} > 1$ units of a good must be shipped from location $i$ to location $n \neq i$ for one unit to arrive (i.e., some of each unit melts in transit).[2] Combining assumptions about the functional form of trade costs with those about preferences and production technology generates predictions for bilateral trade. Arguably, any plausible quantitative spatial model should explain the gravity equation, a strong empirical feature in which bilateral trade increases with exporter and importer size and declines with geographical distance (e.g., surveyed in Head & Mayer 2014).

**2.3.2. Asymmetric versus symmetric transport costs.** Whether transport costs are symmetric or asymmetric (i.e., whether or not $d_{ni} = d_{in}$) has implications both for the characterization of equilibrium and for patterns of trade and income (see Waugh 2010, Allen et al. 2015). Although transport costs are necessarily symmetric if they depend solely on geographic distance, departures from symmetry can arise from a variety of geographic and economic factors (e.g., land gradient and trade volumes).

**2.3.3. Geographic versus economic frictions.** Both geographic frictions (e.g., mountains) and economic frictions (e.g., borders, road and rail networks) can influence bilateral transport costs. With the diffusion of geographic information system data and software, advances have been made in the detailed modeling of observed determinants of transport costs (e.g., mountains, rivers, and coastlines) using algorithms that determine the lowest-transport-cost path, such as the Djikstra or Fast Marching algorithm used by Allen & Arkolakis (2014), Ahlfeldt et al. (2015), Donaldson & Hornbeck (2016), Desmet et al. (2016), Donaldson (2016), and Nagy (2016).

**2.3.4. Role of nontraded goods.** Nontraded goods can typically be thought of as the limiting case in which iceberg trade costs for a particular good are infinite ($d_{ni} \to \infty$). These non-traded goods play an important role in shaping input–output linkages and local multipliers (see, e.g., Moretti 2011b, Caliendo et al. 2014). A given productivity difference in the traded sector has a larger proportionate impact on overall employment with nontraded goods because the relocation of workers in the traded sectors shifts around demand and thus employment for nontraded goods.

## 2.4. Technology for Idea Flows

The second mechanism explaining the importance of the location of agents relative to one another in quantitative spatial models is friction in idea flows. Three main sets of assumptions concerning the technology of idea flows are considered.

**2.4.1. Knowledge externalities and diffusion.** An externality arises whenever an economic agent takes an action that affects another economic agent without internalizing this effect when evaluating the cost and benefits of the action. In the flow of ideas, such externalities can be the result of the lack of a market or can be mediated by prices, as in the case of pecuniary externalities. An obvious example would be a case in which ideas discovered by a researcher or firm in one location diffuse to other researchers and firms in the same location or in different locations. The standard approach to modeling such knowledge externalities is to assume that they are a function of the distance-weighted sum of employment in surrounding locations (see, e.g., Fujita

---

[2]The spatial economics literature has devoted less attention to the distinction between fixed and variable trade costs than, for example, the recent literature on heterogeneous firms in international trade (an exception is Baldwin & Okubo 2005).

& Ogawa 1982, Lucas & Rossi-Hansberg 2002). This reduced-form specification can be derived from alternative microfoundations. The standard classification of these microfoundations is that of Marshall (1920), which distinguishes between knowledge spillovers, externalities due to thick labor markets, and backward and forward linkages. More recently, Duranton & Puga (2004) propose sharing, matching, and learning as three different classes of mechanisms that can result in similar reduced-form specifications. Other research has sought to measure and distinguish between these and other microeconomic mechanisms (see Jaffe et al. 1993, Ellison et al. 2010, Comin et al. 2013). Most empirical studies find that these externalities are highly localized and decay rapidly with geographical, technological, or economic proximity (e.g., Arzaghi & Henderson 2008, Rossi-Hansberg et al. 2010, Ahlfeldt et al. 2015). A key question in setting up spatial models is whether these externalities are present only within the spatial unit of analysis (e.g., Allen & Arkolakis 2014) or across spatial units, as well (e.g., Rossi-Hansberg 2005).

**2.4.2. Innovation.** Spatial models must also determine whether the level of local productivity is constant and exogenous or the result of intentional investments in innovation. The incentives to undertake these investments depend critically on the ability to appropriate the returns from the development of ideas and, thus, the speed with which these ideas diffuse to other agents. Most research in economic geography is concerned with the static spatial distribution of economic activity at a point in time; however, innovation is an inherently dynamic activity. Modeling these dynamics is challenging because of the high dimensionality of the state space across locations and over time. However, tractable quantitative models of spatial innovation and the corresponding evolution of economic activity have recently been developed by Desmet & Rossi-Hansberg (2015), Desmet et al. (2016), and Nagy (2016). In these frameworks, the spatial economy influences the profitability of local innovations by determining the market size of firms and, therefore, the extent to which the cost of innovation can be shared among consumers. The key to the tractability of these frameworks is the fact that a competitive market for land, together with local diffusion of technology, implies that future returns from an innovation are fully capitalized in land rents.

**2.4.3. Transferability of ideas.** Another important factor is the extent to which ideas developed in one location can be costlessly transferred to other locations. In the international trade literature, several studies have explored the implications of frictions that reduce the productivity of ideas when transferred to other countries through foreign direct investment (see, in particular, Arkolakis et al. 2014). Within countries, a firm that enters and develops a blueprint for production in one location may face costs when transferring that blueprint to other locations, as shown by Fajgelbaum et al. (2015).

## 2.5. Costs of Moving People

The third mechanism that explains the importance of the spatial location of agents relative to one another in quantitative spatial models is friction in the movement of people. In this section, four main sets of assumptions are described.

**2.5.1. Migration costs.** The first assumption relates to frictions in the migration of people. Such frictions provide an alternative explanation for real wage differences across locations, in contrast to the idiosyncratic differences in preferences discussed in Section 2.1. This raises the question of the extent to which observed urban–rural wage differentials within countries reflect migration frictions, nonrandom selection of worker productivity, the cost of land and other nontraded goods, and amenity differences (see, e.g., Young 2013, Bryan & Morten 2015). Although these migration

frictions can exist within countries, they are typically thought to be much larger between countries. To the extent that these migration frictions involve sunk costs, agents' location decisions again become inherently dynamic. In this case, these location decisions depend not only on current real wages, but also on expected continuation values, as analyzed by Artuç et al. (2010), Caliendo et al. (2015), and Morten & Oliviera (2016). Desmet et al. (2016) measure the moving costs at the origin and the destination that rationalize the observed net population flows across regions of the world using a dynamic spatial model.

**2.5.2. Commuting.** A second assumption concerns whether agents can separate their workplace and residence by commuting between them. In the canonical monocentric city of urban economics, all production activity is assumed to occur at the center of the city, and commuting costs play the key role in determining the land price gradient with respect to distance from the center of the city (see Alonso 1964, Mills 1967, Muth 1969, Lucas 2000). An important contribution of more recent research has been to allow for nonmonocentric patterns of economic activity within cities; in the nonmonocentric case, the interaction of agglomeration forces and commuting costs remains central to determining internal city structure (see Fujita & Ogawa 1982, Lucas & Rossi-Hansberg 2002, Ahlfeldt et al. 2015, Brinkman 2016). In models of systems of cities, the efficiency of the commuting technologies within each city is an important determinant, alongside the amenities and productivities of each city, in shaping the distribution of city sizes (see Desmet & Rossi-Hansberg 2013, Behrens et al. 2014). In the literature on local labor markets, many studies have examined the impact of local shocks and policy interventions on local employment (for reviews, see Moretti 2011a, Kline & Moretti 2014b, Neumark & Simpson 2015). However, relatively little attention has been devoted to commuting and the resulting distinction between employment by workplace and employment by residence. When locations are connected by bilateral commuting flows, a shock to one location can spill over to other locations (see Monte 2015), and a given local shock can have heterogeneous employment effects across locations depending on commuting networks (see Monte et al. 2015).

**2.5.3. Skills and heterogeneity.** For both migration and commuting decisions, a third modeling factor is whether agents have common or idiosyncratic preferences/productivities across locations. Under the assumption of extreme value–distributed idiosyncratic preferences/productivities, quantitative spatial models imply gravity equation relationships for migration or commuting (e.g., McFadden 1974, Kennan & Walker 2011, Ahlfeldt et al. 2015, Monte et al. 2015, Allen et al. 2016). Empirically, there is strong evidence that both migration and commuting flows are characterized by such gravity equation relationships, in which bilateral flows increase with origin and destination size and decline with geographical distance (see, e.g., Fortheringham & O'Kelly 1989). Whether agents are assumed to have common or idiosyncratic preferences/productivities, an additional modeling decision is whether to allow for multiple type of agents (e.g., workers with different observed levels of skills). In the presence of multiple types of agents who value location characteristics differentially, the equilibrium distribution of economic activity is typically characterized by spatial sorting, in which agents of a given type endogenously self-select into locations with a particular set of characteristics (see, e.g., Davis & Dingel 2015, Gaubert 2015, Redding & Sturm 2016).

**2.5.4. Congestion in transportation.** A fourth specification choice is the extent to which increased flows of people lead to greater congestion and higher travel costs and whether these can be relieved by transport infrastructure provision. Duranton & Turner (2011) provide evidence in support of a fundamental law of highway congestion [originally suggested by Downs (1962)],

according to which increased provision of highways leads to a proportionate increase in vehicle kilometers traveled with no reduction in congestion. Anderson (2014) finds that public transit plays an important role in alleviating congestion during peak travel times. Using hourly data on traffic speeds for all major Los Angeles freeways, Anderson finds that a 2003 strike by Los Angeles County Metropolitan Transportation Authority workers led to an abrupt increase in average delays of 47% (0.19 minutes per mile) during peak travel periods.

## 2.6. Endowments

The above choices about preferences; production technology; and the costs of the movement of goods, ideas, and people need to be combined with choices on the endowments of the economy.

**2.6.1. Population and skills.** A minimal endowment would be homogeneous labor alone, as in an Armington model (e.g., Allen & Arkolakis 2014). More generally, different types of labor can be distinguished, some of which may be more mobile across locations than others and some of which may have different skills or levels of wealth.

**2.6.2. Spatial scope and units.** In most cases, geographically mobile labor is combined with geographically immobile land, such that the model yields predictions for the prices of immobile factors of production (see, e.g., Rossi-Hansberg 2005, Redding 2016, among many others). Two further decisions are the spatial scope of the model and the spatial units for which it is quantified. Is the model concerned with a single city, a system of cities, a set of rural and urban regions within a country or group of countries, or the global economy as a whole? Is space ordered along one dimension (e.g., latitude) or two dimensions? Data are typically available for discrete spatial units. How disaggregated are these units? Are they points on a latitude and longitude grid, city blocks, municipalities, counties, commuting zones, metropolitan areas, states or provinces, regions, or countries as a whole? Clearly these two decisions are interrelated because the choice of spatial scope may limit the level of spatial disaggregation of the units for which data are available.

**2.6.3. Capital and infrastructure.** Other mobile factors of production, such as physical capital that is used in a construction sector (see, e.g., Epple et al. 2010, Combes et al. 2016, Ahlfeldt et al. 2015), can be introduced. Incorporating physical production capital that fully depreciates every period, as do Desmet & Rossi-Hansberg (2013), is also simple. However, incorporating, over time, local capital investments that do not depreciate fully introduces a dynamic forward-looking problem, with the whole distribution of capital across space as a state variable, that has not been tackled in the literature.[3] More generally, depending on the assumptions made about the costs of the movement of goods, ideas, and people, the economy's endowments can also include transport infrastructure networks, which up to now have largely been treated as exogenous in quantitative spatial models.

## 2.7. Equilibrium

Given the above assumptions about preferences; production technology; endowments; and the technologies for the movement of goods, ideas, and people, a final set of assumptions concerning the equilibrium conditions of the model must be considered.

---

[3]The closest frameworks are those following Desmet & Rossi-Hansberg (2014) that model the local accumulation of technology over time. However, those frameworks rely on technology diffusion to argue that all future rents of a local technology investment accrue to land owners (an assumption that is less attractive for capital).

**2.7.1. Market structure.** Two main market structures have been considered in the literature on quantitative spatial models. Models of constant returns to scale, such as those based on the work of Armington (1969) and Eaton & Kortum (2002), typically assume perfect competition (see, e.g., Allen & Arkolakis 2014, Caliendo et al. 2014). In contrast, models of increasing returns to scale, such as those in the new economic geography literature, typically assume monopolistic competition (see, e.g., Helpman 1998, Monte et al. 2015, Redding 2016). One reason for this assumption is the fact that internal increasing returns to scale require the assumption of imperfect competition because otherwise, according to Euler's theorem, factor payments would more than exhaust the value of output. Of the possible forms of imperfect competition, monopolistic competition is particularly tractable, and its assumption of free entry ensures zero equilibrium profits, which implies that all revenue is ultimately paid to factors of production.

**2.7.2. General versus partial equilibrium.** A central feature of quantitative spatial models is the discipline and internal consistency imposed by the equilibrium conditions of the model. However, researchers must choose the level at which these equilibrium conditions are imposed. If the model is of the internal structure of economic activity within a single city, the equilibrium conditions may hold within the city, which can be embedded within a larger economy that provides a reservation level of utility taken as given by the city. If the model is of a single country, these equilibrium conditions may hold within the country, which is assumed to face exogenous prices or levels of expenditure on a world market. At the most general level, if the model is of the global economy, the equilibrium conditions must be specified for the world as a whole. But if some factors are immobile across countries or some goods are nontraded, some of the equilibrium conditions will hold within each country separately.

**2.7.3. Land ownership and the distribution of rents.** If land is used for either residential or production purposes, then it will generate rents to its owners. Thus, specifying who are the owners of land in the different locations modeled is essential for the welfare properties of the model and can be important for the determination of the equilibrium allocation. The urban economics literature has a long tradition of abstracting from land rents by postulating the existence of absentee landlords who receive all the rents but are not explicitly modeled. This assumption, although sometimes convenient, eliminates the ability of the model to incorporate full general equilibrium effects. This limitation is particularly important because many changes in policy or productivity effects of innovation will ultimately be capitalized in land rents and therefore will be accrued to landowners. Of course, simply allowing for a land market where agents can buy and sell land would be ideal. However, it necessarily requires incorporating location-specific wealth effects. For example, if a region receives a positive productivity shock, its land will appreciate, which will make current owners richer and owners everywhere else relatively poorer. Although perhaps realistic, keeping track of these changes in individual wealth in a model with migration is extremely challenging. The key complication is that incorporating location-specific wealth effects makes agents heterogeneous as a function of their location history. The literature has devised three main methods of incorporating land rents into the analysis, and thus accounting for general equilibrium effects, without generating these types of heterogeneous wealth effects. The first is to introduce a global portfolio that aggregates the land rents of the whole economy and to give agents shares in this portfolio (see, e.g., Desmet & Rossi-Hansberg 2014). The second is to distribute land rents locally to current residents (see, e.g., Redding 2016). This second option generates inefficiencies because moving across locations imposes an externality on the rents received by other agents. One can also combine both to account for trade deficits, as do Caliendo et al. (2014). Finally, the third method assumes the presence of local immobile landlords who consume all their

land income locally. Then, local consumption will simply add up to total labor income (see, e.g., Monte et al. 2015).

**2.7.4. Trade balance.** In any spatial model, one must choose the spatial unit for which trade is balanced. In quantitative trade models, this is often a country, although country trade accounts clearly exhibit long-lasting and persistent trade imbalances (see, e.g., Reyes-Heroles 2016). The decision is even more relevant when one focuses on smaller spatial units such as states, commuting zones, counties, or even zip codes or census tracks. The narrower the spatial unit, the less likely it is that trade is balanced for each location because agents can commute or migrate, taking with them their wealth balances, and regions can have accumulated assets and debts in other regions that result in future permanent trade flows. Even so, the assumption of trade balance at the local level is common in the quantitative spatial literature. Another popular option is to acknowledge the possibility of deficits, calibrate them using data, but keep them invariant in the counterfactual exercises (see, e.g., Allen & Arkolakis 2014). Yet another possibility is to calculate a baseline counterfactual economy without deficits from which all other counterfactual exercises are computed (thereby ameliorating the fact that deficits are fixed). In static models, a final possibility is to model changes in deficits as resulting from variations in the rents accrued to land owners (see, e.g., Caliendo et al. 2014). A full quantitative dynamic spatial model that endogenizes the consumption–savings decisions that determine whether a location saves and borrows over time has, to our knowledge, not been developed.

## 2.8. Criteria for Menu Choice

Having outlined the modules or building blocks that can be used to create quantitative spatial models, we now discuss some of the possible criteria for choosing among these modules.

**2.8.1. Tractability.** The first criterion includes both analytical and computational tractability. Traditionally, theoretical models of economic geography have focused on a small number of symmetric regions to preserve analytical tractability. Technical advances have now made it possible to obtain analytical results for the existence and uniqueness of equilibrium and for comparative statics even for large numbers of asymmetric locations connected by real-world transport networks (see, in particular, Allen & Arkolakis 2014, Allen et al. 2015). Other technical innovations have permitted analytical characterizations of the dynamics of the distribution of economic activity across space (see Desmet & Rossi-Hansberg 2013, Desmet et al. 2016). Related methodological improvements have developed a set of standard techniques for tractably undertaking counterfactuals in a class of theoretical models using the observed values of variables in an initial equilibrium (see Dekle et al. 2007). At the same time, advances in computing power and computational methods have made it possible to solve systems of nonlinear equations for large numbers of locations over realistic computational time periods.[4]

**2.8.2. Structural assumptions.** A second criterion involves the choice of structural parameters and exogenous location characteristics that are invariant to policy interventions. When a quantitative spatial model is used to undertake a counterfactual for the impact of a place-based policy or transport infrastructure improvement, the researcher must decide which components of the

---

[4]For example, Ahlfeldt et al. (2015) compute equilibrium for 15,937 city blocks in Berlin. Desmet et al. (2016) calculate equilibria for a grid of 68,400 locations across the world.

model are invariant to this intervention. Are productivity and amenities exogenous? Or are there agglomeration externalities for productivity and amenities such that only certain components of these location characteristics are fundamentals that are invariant to the intervention? Is there an outside level of utility in a wider economy that is constant? Or are there exogenous prices or expenditures in world markets? Answers to questions such as these will influence the selection of building blocks from the list outlined above. Only when the assumed structural parameters and locational fundamentals are indeed constant will the analysis not be subject to the Lucas critique.

**2.8.3. Connection between model and data.** A third criterion relates to the question being asked, the level of detail at which an answer is required, and what is observed in the data. Is the question about the aggregate effects of a policy or the impact on the distribution of economic activity across disaggregated spatial units? To what extent is understanding the distributional impact at the disaggregated level relevant for evaluating the aggregate effect? What are the spatial units for which the data are recorded? What types of data are available? These include the levels of the endogenous variables of the model for each location (e.g., population, wages); endogenous bilateral flows (e.g., trade and commuting flows); frictions in the movements of goods, ideas, and people (e.g., mountains and borders); and changes in the endogenous variables of the model for each location (e.g., changes in population or wages). Different types of data can sometimes be substitutes for one another. For example, quantitative models can typically be solved using either data on endogenous bilateral flows (e.g., bilateral trade) or data on exogenous frictions (e.g., the costs of traversing mountains and water). Do the data available permit a structural estimation of the model's parameters? Or will the model be calibrated using values of the model's parameters from elsewhere? Or can a subset of the parameters be estimated and the remaining parameters borrowed from other studies? When the model is taken to the data, is it exactly identified such that it has enough degrees of freedom to exactly explain the observed data as an equilibrium outcome? If so, is the model invertible such that there exists a one-to-one mapping from the parameters and observed data to the unobserved location characteristics or structural residuals? What overidentification checks can be undertaken using moments not used in the calibration or estimation to provide a check on the validity of the model's predictions?

# 3. A QUANTITATIVE SPATIAL MODEL

In this section, we outline a canonical quantitative spatial model that corresponds to a multiregion version of the new economic geography model of Helpman (1998). To study the determinants of the spatial distribution of economic activity across a set of regions connected by goods trade and factor mobility, we select the following items from the menu of building blocks above:

1. From Section 2.1, we select love of variety, single traded sector, no amenities, residential land use, and common preferences.
2. From Section 2.2, we select increasing returns to scale, exogenous productivity, no input–output linkages, and no commercial land use.
3. From Section 2.3, we select iceberg variable trade costs, symmetric trade costs, economic and geographic frictions, and no nontraded goods besides residential land use.
4. From Section 2.4, we select no knowledge externalities or diffusion, no innovation, and no transferability of ideas.
5. From Section 2.5, we select perfectly costless migration, no commuting, single worker type with no heterogeneity, and no congestion in transportation.
6. From Section 2.6, we select homogeneous labor, exogenous land endowments in regions within a single country, and no capital.

7. From Section 2.7, we select monopolistic competition, general equilibrium with a single country, land rents redistributed to residents, and balanced trade in each location.

This model has been widely used in empirical work, including work by Hanson (2005) and Redding & Sturm (2008).

We consider an economy consisting of a set $N$ of regions indexed by $n$. Each region is endowed with an exogenous quality-adjusted supply of land ($H_i$). The economy as a whole is endowed with a measure $\bar{L}$ of workers, where each worker has one unit of labor that is supplied inelastically with zero disutility. Workers are perfectly geographically mobile, and thus, in equilibrium, real wages are equalized across all populated regions. Regions are connected by a bilateral transport network that can be used to ship goods subject to symmetric iceberg trade costs, such that $d_{ni} = d_{in} > 1$ units must be shipped from region $i$ in order for one unit to arrive in region $n \neq i$, where $d_{nn} = 1$.[5]

## 3.1. Consumer Preferences

Preferences are defined over goods consumption ($C_n$) and residential land use ($h_n$) and are assumed to take the Cobb-Douglas form:[6]

$$U_n = \left( \frac{C_n}{\alpha} \right)^{\alpha} \left( \frac{h_n}{1-\alpha} \right)^{1-\alpha}, \qquad 0 < \alpha < 1. \qquad 1.$$

The goods consumption index $C_n$ is defined over consumption $c_{ni}(j)$ of each horizontally differentiated variety $j$ from the endogenous measures $M_i$ supplied by each region. This consumption index and its dual price index, $P_n$, are given by

$$C_n = \left[ \sum_{i \in N} \int_0^{M_i} c_{ni}(j)^{\rho} \, dj \right]^{\frac{1}{\rho}}, \qquad P_n = \left[ \sum_{i \in N} \int_0^{M_i} p_{ni}(j)^{1-\sigma} \, dj \right]^{\frac{1}{1-\sigma}}. \qquad 2.$$

## 3.2. Production

Varieties are produced under conditions of monopolistic competition and increasing returns to scale. To produce a variety, a firm must incur a fixed cost of $F$ units of labor and a constant variable cost in terms of labor that depends on a location's productivity $A_i$. Therefore, the total amount of labor, $l_i(j)$, required to produce $x_i(j)$ units of a variety $j$ in location $i$ is

$$l_i(j) = F + \frac{x_i(j)}{A_i}. \qquad 3.$$

Profit maximization and zero profits imply that equilibrium prices are a constant markup over the marginal cost of supplying a variety to a market,

$$p_{ni}(j) = \left( \frac{\sigma}{\sigma - 1} \right) d_{ni} \frac{w_i}{A_i}, \qquad 4.$$

and equilibrium output of each variety is equal to a constant that depends on location productivity, namely,

$$x_i(j) = \bar{x}_i = A_i(\sigma - 1)F, \qquad 5.$$

---

[5]Although we interpret the locations here as regions, they could also be interpreted as cities, as in the literature on systems of cities following Henderson (1974).

[6]For empirical evidence using US data in support of the constant housing expenditure share implied by the Cobb-Douglas functional form, readers are referred to Davis & Ortalo-Magne (2011).

which implies that equilibrium employment for each variety is the same for all locations, so

$$l_i(j) = \bar{l} = \sigma F. \qquad 6.$$

Given this constant equilibrium employment for each variety, labor market clearing implies that the total measure of varieties supplied by each location is proportional to the endogenous supply of workers choosing to locate there:

$$M_i = \frac{L_i}{\sigma F}. \qquad 7.$$

### 3.3. Price Indices and Expenditure Shares

Using equilibrium prices (Equation 4) and labor market clearing (Equation 7), one can express the price index dual to the consumption index (Equation 2) as

$$P_n = \frac{\sigma}{\sigma - 1}\left(\frac{1}{\sigma F}\right)^{\frac{1}{1-\sigma}}\left[\sum_{i \in N} L_i \left(d_{ni}\frac{w_i}{A_i}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \qquad 8.$$

Using the constant elasticity of substitution (CES) expenditure function, equilibrium prices (Equation 4), and labor market clearing (Equation 7), the share of location $n$'s expenditure on goods produced in location $i$ is

$$\pi_{ni} = \frac{M_i p_{ni}^{1-\sigma}}{\sum_{k \in N} M_k p_{nk}^{1-\sigma}} = \frac{L_i \left(d_{ni}\frac{w_i}{A_i}\right)^{1-\sigma}}{\sum_{k \in N} L_k \left(d_{nk}\frac{w_k}{A_k}\right)^{1-\sigma}}. \qquad 9.$$

The model therefore implies a gravity equation for goods trade, where the bilateral trade between locations $n$ and $i$ depends on both bilateral resistance (bilateral trade costs $d_{ni}$) and multilateral resistance (trade costs to all other locations $k$ $d_{nk}$), as in the work of Anderson & van Wincoop (2003). Together, Equations 8 and 9 imply that each location's price index can be written in terms of its trade share with itself, so

$$P_n = \frac{\sigma}{\sigma - 1}\left(\frac{L_n}{\sigma F \pi_{nn}}\right)^{\frac{1}{1-\sigma}}\frac{w_n}{A_n}. \qquad 10.$$

### 3.4. Income and Population Mobility

Expenditure on land in each location is redistributed in a lump sum to the workers residing in that location. Therefore, trade balance at each location implies that per capita income $v_n$ in each location equals labor income $w_n$ plus per capita expenditure on residential land, $(1-\alpha)v_n$, namely,

$$v_n L_n = w_n L_n + (1-\alpha)v_n L_n = \frac{w_n L_n}{\alpha}. \qquad 11.$$

Land market clearing implies that the supply of quality-adjusted land, $H_n$, equals the demand for land, $L_n b_n$. By combining this market clearing condition with the first-order condition of the consumer problem, we obtain the result that land rents, $r_n$, are given by

$$r_n = \frac{(1-\alpha)v_n L_n}{H_n} = \frac{1-\alpha}{\alpha}\frac{w_n L_n}{H_n}. \qquad 12.$$

Population mobility implies that workers receive the same real income in all populated locations, thus

$$V_n = \frac{v_n}{P_n^\alpha r_n^{1-\alpha}} = \bar{V}. \qquad 13.$$

By using the price index (Equation 10), the assumption that trade is balanced at each location such that income equals expenditure (Equation 11), and land market clearing (Equation 12) in the population mobility condition (Equation 13), one determines that real wage equalization implies that the population $L_n$ and domestic trade share $\pi_{nn}$ of each location must satisfy

$$\bar{V} = \frac{A_n^\alpha H_n^{1-\alpha} \pi_{nn}^{-\alpha/(\sigma-1)} L_n^{-\frac{\sigma(1-\alpha)-1}{\sigma-1}}}{\alpha \left(\frac{\sigma}{\sigma-1}\right)^\alpha \left(\frac{1}{\sigma F}\right)^{\frac{\alpha}{1-\sigma}} \left(\frac{1-\alpha}{\alpha}\right)^{1-\alpha}}. \qquad 14.$$

Therefore, the population share of each location ($\lambda_n \equiv L_n/\bar{L}$) depends on its productivity $A_n$, supply of land $H_n$, and domestic trade share $\pi_{nn}$ relative to those of all other locations,

$$\lambda_n = \frac{L_n}{\bar{L}} = \frac{\left[A_n^\alpha H_n^{1-\alpha} \pi_{nn}^{-\alpha/(\sigma-1)}\right]^{\frac{\sigma-1}{\sigma(1-\alpha)-1}}}{\sum_{k\in N} \left[A_k^\alpha H_k^{1-\alpha} \pi_{kk}^{-\alpha/(\sigma-1)}\right]^{\frac{\sigma-1}{\sigma(1-\alpha)-1}}}, \qquad 15.$$

where each location's domestic trade share $\pi_{nn}$ summarizes its market access to other locations.

## 3.5. General Equilibrium

The properties of the general equilibrium of the model can be characterized analytically by combining the trade share (Equation 9), price index (Equation 8), and population mobility condition (Equation 13). Under the assumption that trade costs are symmetric ($d_{ni} = d_{in}$), one can follow the arguments of Allen & Arkolakis (2014) to show that these three sets of relationships reduce to the following system of $N$ equations in the $N$ populations of each location:

$$L_n^{\tilde{\sigma}\gamma_1} A_n^{-\frac{(\sigma-1)(\sigma-1)}{2\sigma-1}} H_n^{-\frac{\sigma(\sigma-1)(1-\alpha)}{\alpha(2\sigma-1)}} = \bar{W}^{1-\sigma} \sum_{i\in N} \frac{1}{\sigma F} \left(\frac{\sigma}{\sigma-1} d_{ni}\right)^{1-\sigma} \left(L_i^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} A_i^{\frac{\sigma(\sigma-1)}{2\sigma-1}} H_i^{\frac{(\sigma-1)(\sigma-1)(1-\alpha)}{\alpha(2\sigma-1)}}, \qquad 16.$$

where the scalar $\bar{W}$ is determined by the requirement that the labor market clears ($\sum_{n\in N} L_n = \bar{L}$) and

$$\tilde{\sigma} \equiv \frac{\sigma-1}{2\sigma-1}, \qquad \gamma_1 \equiv \frac{\sigma(1-\alpha)}{\alpha}, \qquad \gamma_2 \equiv 1 + \frac{\sigma}{\sigma-1} - \frac{(\sigma-1)(1-\alpha)}{\alpha}.$$

Wages in turn are implicitly determined by

$$w_n^{1-2\sigma} A_n^{\sigma-1} L_n^{(\sigma-1)\frac{1-\alpha}{\alpha}} H_n^{-(\sigma-1)\frac{1-\alpha}{\alpha}} = \xi, \qquad 17.$$

where $\xi$ is a scalar that normalizes wages. Allen & Arkolakis (2014) use this argument, together with the mathematical results for fixed points of systems of equations of the form given by Equation 16 (Fujimoto & Krause 1985), to show that there exists a unique $L_n$ for each $n$ that satisfies Equation 16 as long as $\gamma_2/\gamma_1 \in (0, 1)$. Thus, given the land area and productivity parameters $\{H_n, A_n\}$ and symmetric bilateral trade frictions $\{d_{ni}\}$ for all locations $n, i \in N$, there exists a unique equilibrium as long as this parametric restriction is satisfied. Furthermore, if $\gamma_2/\gamma_1 \in (0, 1)$, then one can also guarantee that a solution to Equation 16 can be found by iteration from any initial distribution of populations.[7]

---

[7]If space is assumed to be continuous, one can set up an analogous model in which Equation 16 will have an integral rather than a sum on the right-hand side. In that case, Allen & Arkolakis (2014) show that similar results apply using theorem 2.19 of Zabreyko et al. (1975). Thus, if $\gamma_2/\gamma_1 \in (0, 1)$, then a solution exists, is unique, and can be found by iteration.

The parameter restrictions that guarantee that an equilibrium exists and is unique amount to imposing conditions that guarantee that congestion forces always dominate agglomeration forces. In our simple model, a sufficient condition for $\gamma_2/\gamma_1 \in (0, 1)$ is $\sigma (1 - \alpha) > 1$. Intuitively, as population concentrates in a location, the measure of varieties produced there expands, which, in the presence of trade costs, makes that location a more attractive residence (an agglomeration force). However, as population concentrates in a location, this also bids up land prices (a dispersion force). The higher the elasticity of substitution ($\sigma$), the weaker is the agglomeration force. The higher the share of land ($1 - \alpha$), the stronger is the dispersion force. For parameter values for which $\sigma (1 - \alpha) > 1$, the dispersion force dominates the agglomeration force, and a unique equilibrium distribution of economic activity exists.

The existence of such a unique equilibrium is important because it ensures that counterfactuals for transport infrastructure improvements or other public policy interventions have determinate implications for the spatial distribution of economic activity. Although this is a convenient property of the model for quantitative empirical work, a central feature of the theoretical literature on new economic geography was the presence of multiple equilibria [as in the original core–periphery model of Krugman (1991b)], and assuming $\sigma (1 - \alpha) > 1$ excludes this possibility. Thus, for the range of parameters where $\sigma (1 - \alpha) > 1$, the model cannot generate agglomerations when space is perfectly homogeneous; it can generate agglomerations only as a result of initial differences across locations. Of course, in this general class of models, the vector of initial differences can be multidimensional and quite rich, as discussed in the context of a richer model with many sources of heterogeneity across locations by Desmet et al. (2016).

At a more philosophical level, whether a model is characterized by multiple equilibria may depend on its level of abstraction. On the one hand, a model may exhibit multiple equilibria because it omits the relevant idiosyncratic factors that determine one allocation rather than another in the data. On the other hand, if a model includes all such idiosyncratic factors, it ceases to be a model and becomes a descriptive representation of reality. One possible interpretation of such idiosyncratic factors is as random realizations of a stochastic error that is drawn from a parameterized distribution. At smaller spatial scales (e.g., blocks within cities), one might expect such random idiosyncratic factors to be more important relative to the systematic deterministic components of a model (e.g., natural resource abundance) than at larger spatial scales (e.g., across regions or countries).

Throughout this section, we focus on a canonical new economic geography model with increasing returns to scale and monopolistic competition. However, similar properties hold in a wider class of models. Allen & Arkolakis (2014) demonstrate an isomorphism to a perfectly competitive Armington trade model (in which goods are differentiated by location of origin), extended to incorporate labor mobility and external economies. Allen & Arkolakis (2014) and Redding (2016) show that similar properties also hold in a perfectly competitive Ricardian trade model following Eaton & Kortum (2002) with labor mobility and external economies of scale.

## 3.6. Model Inversion

In this section, we describe how the quantitative spatial model can be used to rationalize observed data. We suppose that a researcher has estimates of the model's two key parameters: the share of residential land in consumer expenditure ($\alpha$) and the elasticity of substitution between varieties ($\sigma$). The researcher is also assumed to have parameterized symmetric bilateral trade costs ($d_{ni}$), for example, by assuming that they are a constant elasticity function of distance, and to observe an endogenous population, $\{L_n\}$, and nominal wages, $\{w_n\}$. One can show that there is a

one-to-one mapping from the model's parameters and the observed data to the unobserved values of quality-adjusted land $\{H_n\}$ and productivities $\{A_n\}$ (up to a normalization constant). That is, the model can be inverted to recover the unique values of unobserved quality-adjusted land and productivities that rationalize the observed data as an equilibrium outcome of the model.

Inverting the model amounts to using Equations 16 and 17 to solve for $\{A_n, H_n\}$ given $\{L_n, w_n\}$. This is exactly the opposite of what we do when we solve for an equilibrium of the model, where we solve for $\{L_n, w_n\}$ given $\{A_n, H_n\}$. To guarantee that there exists a unique set of values $\{A_n, H_n\}$ that rationalize the observed data (up to a normalization or choice of units), we can proceed as follows. Using Equation 17, we can solve for $H_n$ and substitute this solution in Equation 16. The resulting equation can then be solved for $\{A_n\}$ using information on $\{L_n, w_n\}$. To show that such a solution exists and is unique, we can again use the mathematical theorems that guarantee solutions to these type of equations [e.g., those of Fujimoto & Krause (1985) for discrete space or Zabreyko et al. (1975) for continuous space]. As in the previous section, a solution exists if, after substituting for $H_n$, the exponent of the $A_i$ term inside the sum is smaller than the exponent of the $A_n$ term outside the sum. In our example, this is guaranteed if $\sigma (1 - \alpha) > 1$. We can then recover $\{H_n\}$ using Equation 17.

Having recovered the unobserved productivities $\{A_n\}$, these can be used, together with the parameterization of trade costs ($d_{ni}$) and observed wages ($w_n$) in the trade shares (Equation 9), to generate predictions for unobserved bilateral trade shares ($\pi_{ni}$) in the equilibrium observed in the data. Related arguments hold if the researcher directly observes bilateral trade shares ($\pi_{ni}$) instead of having to assume values for bilateral trade costs ($d_{ni}$). In this case, unobserved quality-adjusted land supplies ($H_n$), productivities ($A_n$), and bilateral trade costs ($d_{ni}$) can be recovered from the observed data (up to a normalization or choice of units).

These arguments imply that the model is exactly identified, in the sense that it has the same number of degrees of freedom (unobserved location characteristics in the form of the quality-adjusted land supplies $H_n$ and productivities $A_n$ for each location) as observed endogenous variables (population $L_n$ and wages $w_n$). Therefore, the model's ability to explain the observed data cannot be used as a test of the model because the unobserved quality-adjusted land supplies and productivities are free parameters that can be adjusted so as to ensure that the model exactly matches the data. Note also that, because $H_n$ denotes quality-adjusted land, it cannot be directly compared to observations on total land supply in a location. Doing so would simply yield a measure of the average quality of land in that region (which is related to its residential amenities). Furthermore, the observed data on wages and populations cannot be used to estimate the model's structural parameters, $\alpha$ and $\sigma$. Given any value for these parameters, quality-adjusted land supplies and productivities can be adjusted so as to ensure that the model exactly rationalizes the data. Therefore, any change in the value of the structural parameters can be offset by changes in these unobserved location characteristics such that the model continues to explain the data.

Nonetheless, the model's ability to exactly explain the data implies that it provides a framework that can be used to decompose the observed variation in endogenous variables (e.g., population and wages) into the contribution of different exogenous determinants (e.g., trade costs and exogenous determinants of productivity and quality-adjusted land). Note that the recovered exogenous fundamentals do not explain the variation in endogenous variables but simply account for it according to the model. Furthermore, additional data not used for the quantification of the model or exogenous shocks to the economy (e.g., natural experiments from history) can be used to provide overidentification checks (tests of the model's external validity) or to estimate the model's structural parameters. We consider several empirical studies in Section 5 that have sought to provide such overidentification checks or undertake such structural estimation.

## 3.7. Counterfactuals

In this section, we show how our quantitative spatial model can be used to undertake counterfactuals for the effects of public policy interventions, such as transport infrastructure improvements. We show that these counterfactuals can be undertaken using the observed values of the endogenous variables of the model in an initial equilibrium without having to solve for the unobserved location characteristics, as in the work of Dekle et al. (2007). We denote the (unknown) value of variables in the counterfactual equilibrium with a prime ($x'$) and the relative value of variables in the counterfactual and observed equilibria by a hat ($\hat{x} = x'/x$). We suppose that the researcher observes population ($L_n$), wages ($w_n$), and trade shares ($\pi_{ni}$) in the initial equilibrium and can parameterize the change in bilateral trade costs as a result of the transport infrastructure improvement ($\hat{d}_{ni}$). From the trade share (Equation 9), price index (Equation 10), income equal to expenditure (Equation 11), land market clearing (Equation 12), and population mobility (Equation 13), we obtain the following system of equations that can be used to solve for the counterfactual changes in wages, trade shares, and population shares ($\{\hat{w}, \hat{\pi}_{ni}, \hat{\lambda}_n\}$) given only the observed wages, trade shares, and population shares in the initial equilibrium ($\{w, \pi_{ni}, \lambda_n\}$):

$$\hat{w}_i \hat{\lambda}_i (w_i \lambda_i) = \sum_{n \in N} \hat{\pi}_{ni} \hat{w}_n \hat{\lambda}_n \pi_{ni} (w_n \lambda_n), \qquad 18.$$

$$\hat{\pi}_{ni} \pi_{ni} = \frac{(\hat{d}_{ni} \hat{w}_i)^{1-\sigma} \hat{L}_i \pi_{ni}}{\sum_{k \in N} (\hat{d}_{nk} \hat{w}_k)^{1-\sigma} \hat{L}_k \pi_{nk}}, \qquad 19.$$

$$\hat{\lambda}_n \lambda_n = \frac{\hat{\pi}_{nn}^{-\frac{\alpha}{\sigma(1-\alpha)-1}} \lambda_n}{\sum_{k \in N} \hat{\pi}_{kk}^{-\frac{\alpha}{\sigma(1-\alpha)-1}} \lambda_k}. \qquad 20.$$

Given our assumptions of $\sigma(1-\alpha) > 1$ and symmetric trade costs, there exists a unique general equilibrium in the model, which ensures that these counterfactuals yield determinate predictions for the impact of the transport infrastructure improvement or another public policy intervention on the spatial equilibrium distribution of economic activity.

## 3.8. Welfare

A further implication of this class of quantitative spatial models is that the welfare effects of public policy interventions that change trade costs can be expressed solely in terms of empirically observable sufficient statistics. Consider a transport infrastructure improvement that reduces trade costs between an initial equilibrium (indexed by 0) and a subsequent equilibrium (indexed by 1). Perfect population mobility implies that the transport infrastructure improvement leads to reallocations of population across locations until real wages are equalized. Using the population mobility condition (Equation 14), the changes in the domestic trade share ($\pi_{nn}$) and population ($L_n$) for any one location are sufficient statistics for the welfare impact of the transport infrastructure improvement on all locations:

$$\frac{\bar{V}^1}{\bar{V}^0} = \left(\frac{\pi_{nn}^0}{\pi_{nn}^1}\right)^{\frac{\alpha}{\sigma-1}} \left(\frac{\lambda_n^0}{\lambda_n^1}\right)^{\frac{\sigma(1-\alpha)-1}{\sigma-1}}. \qquad 21.$$

Given our assumption of $\sigma(1-\alpha) > 1$, a larger reduction in a location's domestic trade share must be offset by a larger increase in its population to preserve real wage equalization. Intuitively, if the transport infrastructure improvement decreases trade costs for one location more than for other locations (and thus reduces the domestic trade share of this location), the resulting upwards

pressure on its real wage induces a population inflow until the price of the immobile factor land is bid up to restore real wage equalization. This implication is a direct analog of the result found by the international trade literature that the domestic trade share is a sufficient statistic for the welfare gains from trade in a class of trade models (see Arkolakis et al. 2012). In an economic geography model in which an immobile factor of production, such as land, is used residentially or commercially, changes in the distribution of mobile factors of production across locations must also be taken into account (see Caliendo et al. 2014, Redding 2016).

### 3.9. Quantitative Illustration

In this section, we outline a quantitative illustration of the model for which accompanying MATLAB code is available. We show how the quantitative model can be used to evaluate the impact of trade frictions (both between countries and between regions within countries) on the spatial distribution of economic activity and welfare. We consider a model economy on a $30 \times 30$ latitude and longitude grid. We assume that this economy consists of two countries, one occupying the western half of the grid (West) and the other occupying the eastern half (East). We assume that labor is perfectly mobile across locations within each country but perfectly immobile across countries. Following Donaldson (2016), we compute a measure of lowest-cost-route effective distance, which minimizes the cumulative cost of traveling between each origin and destination. In particular, this measure assigns a cost (or weight) to traveling across each point $n$ on the grid of $\delta_n$. For a pair of neighboring points $n$ and $i$, the effective distance for an orthogonal link between these locations (moving vertically or horizontally across the grid) is $\text{dist}_{ni} = (\delta_n + \delta_i)/2$. In contrast, the effective distance for a diagonal link between these locations (moving diagonally across the grid) is $\text{dist}_{ni} = \{[2(\delta_n + \delta_i)^2]^{0.5}\}/2$. For any origin and destination, the effective distance is the sum across all of the links between neighboring locations along the lowest-cost route. We assume that the cost (weight) is the same for each location on the grid ($\delta_n = \delta_i = \delta$) and normalize this common distance weight to one.

We allow productivity to differ randomly across locations. For each location, we draw a realization for productivity $\{A_n\}$ from an independent standard log normal distribution. **Figure 1** displays the realization of productivities for the 900 locations in our grid. In this realization, there are two clusters of high-productivity areas, one northwest of the border and another southeast of it. Other, more isolated, high-productivity areas are also evident. Of course, in this example the location of high-productivity areas is purely random.

For simplicity, we assume that each location has the same quality-adjusted land area ($H_n$) of 100 km². We choose central values for the model's parameters based on the existing empirical literature. First, we set the share of land in residential consumption expenditure ($1 - \alpha$) to 25%, which is in line with the housing expenditure share of Davis & Ortalo-Magne (2011). Second, we set the elasticity of substitution ($\sigma$) equal to 5, which implies an elasticity of trade flows with respect to trade costs of $\sigma - 1 = 4$, in line with the estimates of Simonovska & Waugh (2014). Third, we assume that trade costs are a constant elasticity function of effective distance ($d_{ni} = \text{dist}_{ni}^{\phi}$), which implies an elasticity of trade flows with respect to effective distance of $-(\sigma - 1)\phi$ [because trade flows depend on $d_{ni}^{-(\sigma-1)} = \text{dist}_{ni}^{-(\sigma-1)\phi}$]. We choose the parameter $\phi$ to match the elasticity of trade flows with respect to distance in gravity equations using interregional trade data of $-(\sigma - 1)\phi = -1.5$, which, for our assumed value for $\sigma - 1$, implies $\phi = 0.375$.

In addition to these geographical frictions from transport costs, we consider two forms of economic frictions to trade between locations. First, we assume a proportional internal tax on trade with other locations of 100% ($\tau^{\text{in}} = 2$), which is paid whenever a good flows from one
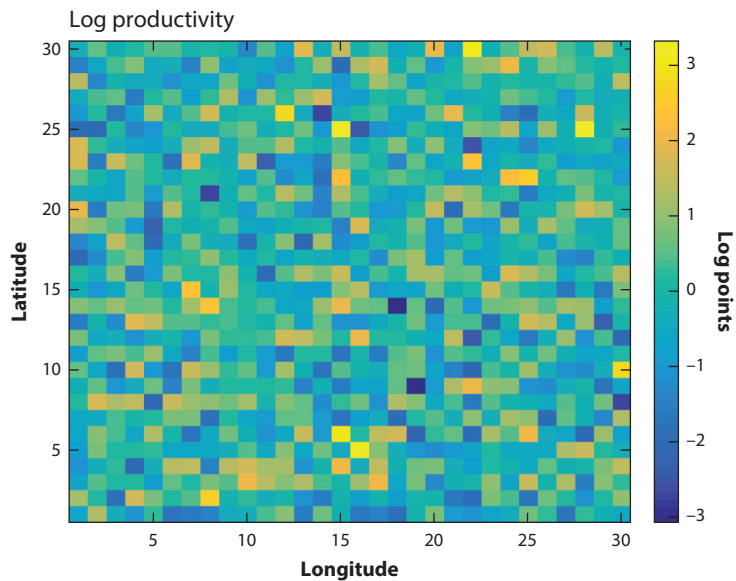
**Figure 1**

Realization of random local productivity for the 900 locations in our grid. Cold (*blue*) colors correspond to lower values, and hot (*yellow*) colors correspond to higher values.

location to another. Second, we assume a proportional external tax on trade between the two countries of 100% ($\tau^{out} = 2$), which is paid whenever a good crosses the border between the two countries (between latitude 15 and 16). For simplicity, we assume that the revenue from both taxes is wasted, and thus both correspond to real resource costs. In **Figure 2**, we display the log level of economic activity across locations in the initial equilibrium with both taxes. Areas of high productivity have large population concentrations (**Figure 2a**), high wages (**Figure 2b**), and high land prices (**Figure 2c**). As expected, the log of the price index (**Figure 2d**) is a smooth surface with the gradient governed by trade costs. Prices are lower in areas that produce a large variety of goods, for example, at the two large cities close to the border. The largest agglomerations in this economy are most clearly appreciated in this panel. Panel *d* also clearly exhibits the border effect created by the tariff between the two countries.

In **Figure 3**, we display the log relative changes in population (log $\hat{L}$), wages (log $\hat{w}$), land prices (log $\hat{r}$), and price indices (log $\hat{P}$) as a result of the removal of the proportional tax on trade between the two countries. As trade costs between the two countries fall, economic activity reallocates toward the border between them. The areas that benefit the most are the ones close to but on the opposite side of the border from the large cities. These locations can now trade more cheaply with the large market in those cities and thus experience the largest increases in population (**Figure 3a**), wages (**Figure 3b**), and land rents (**Figure 3c**) and the largest reductions in the price index (**Figure 3d**). In contrast, the largest agglomerations lose relative to these up and coming locations. In the first row of **Table 1**, we report the resulting impact on the common level of welfare across locations within each country. We find that this external trade liberalization raises welfare in West and East by around 0.2% and 0.3%, respectively.

In **Figure 4**, we present an alternative counterfactual experiment where we remove all internal trade costs but leave international trade costs as in the initial equilibrium. The figure presents
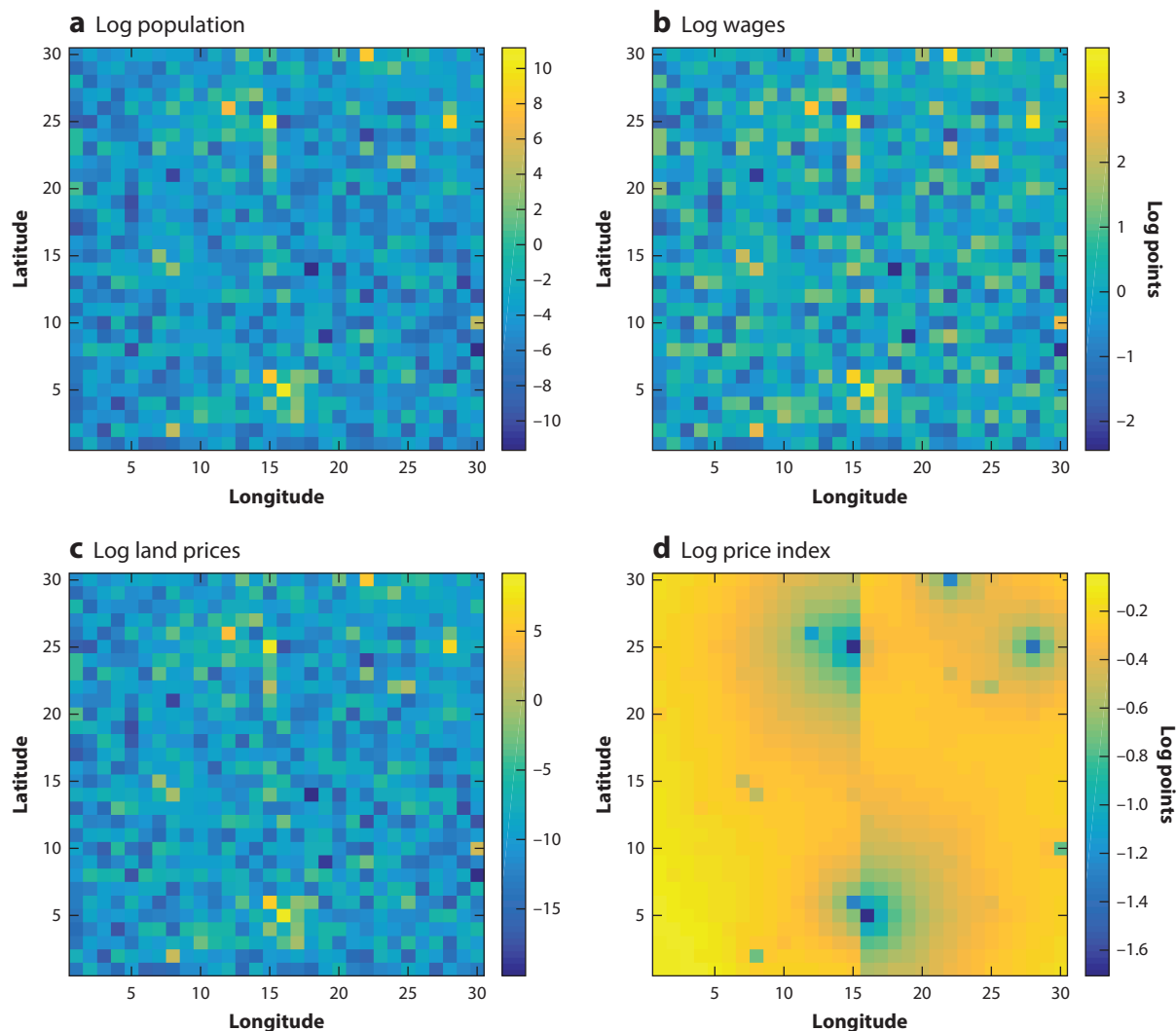
**a** Log population

**b** Log wages

**c** Log land prices

**d** Log price index

**Figure 2**

Initial equilibrium, showing (*a*) log population, (*b*) log wages, (*c*) log land prices, and (*d*) log price index. Cold (*blue*) colors correspond to lower values, and hot (*yellow*) colors correspond to higher values.

relative changes with respect to the initial equilibrium. The implications of an internal reduction in trade costs are clearly quite different than the ones from an external trade cost reduction. The main effect of the internal liberalization is to reduce the size of the two large cities in favor of rural areas, thereby making economic activity more dispersed (**Figure 4*a***). As trade costs decline, the home market effect reducing local price indexes in large cities weakens, so that prices fall everywhere but less so in areas with larger populations (**Figure 4*d***). Wages and land rents also fall in large agglomerations, whereas they increase in all other regions (**Figure 4*b,c***). The second row of **Table 1** reports the welfare impact of the removal of the internal tax on trade with other locations. We find that this internal trade liberalization raises welfare in West and East by around 1.4% and 2.3%, respectively, more than five times the effects of the external trade liberalization.

**a**  Log relative population

**b**  Log relative wages

**c**  Log relative land rents
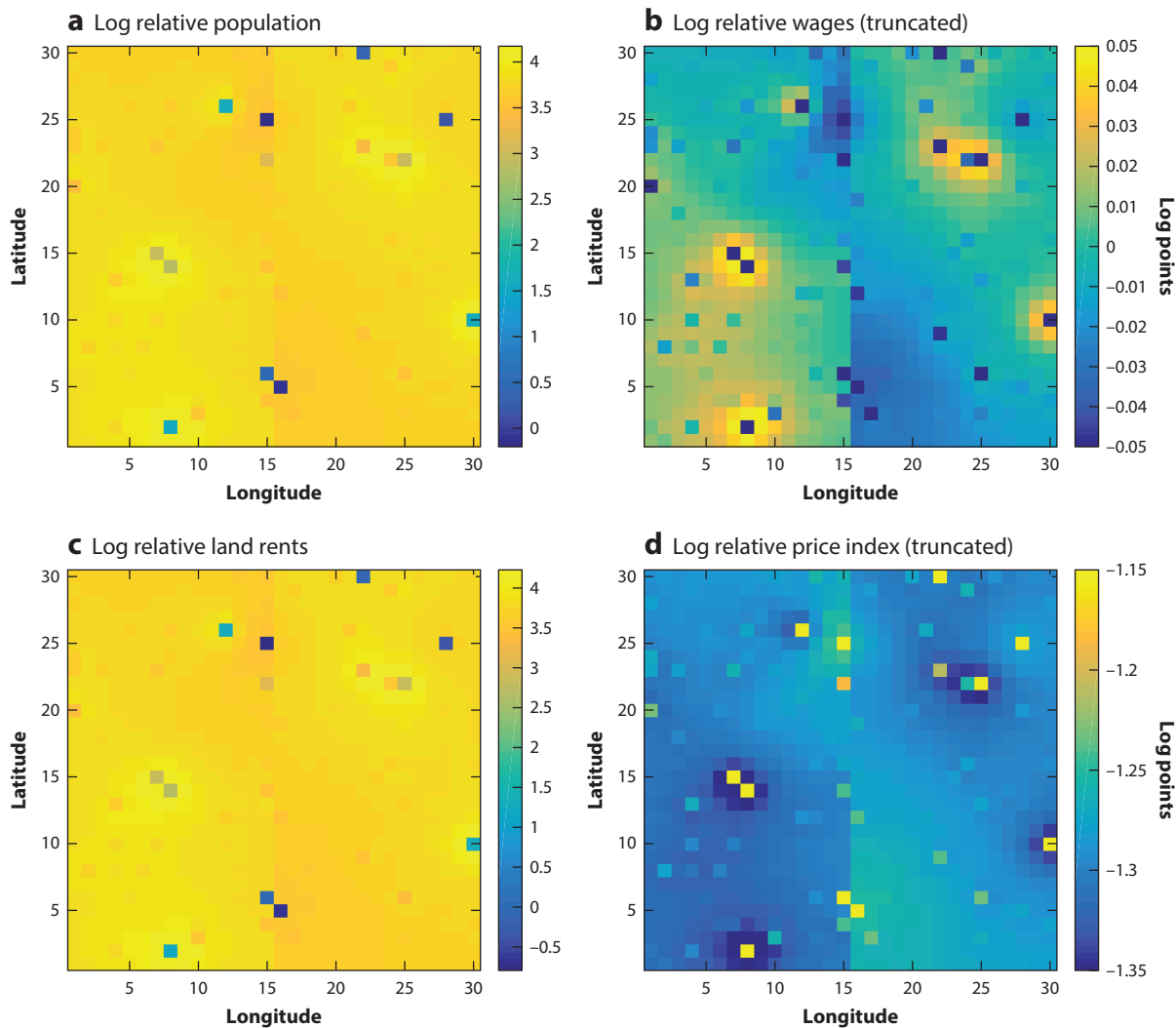
**d**  Log relative price index
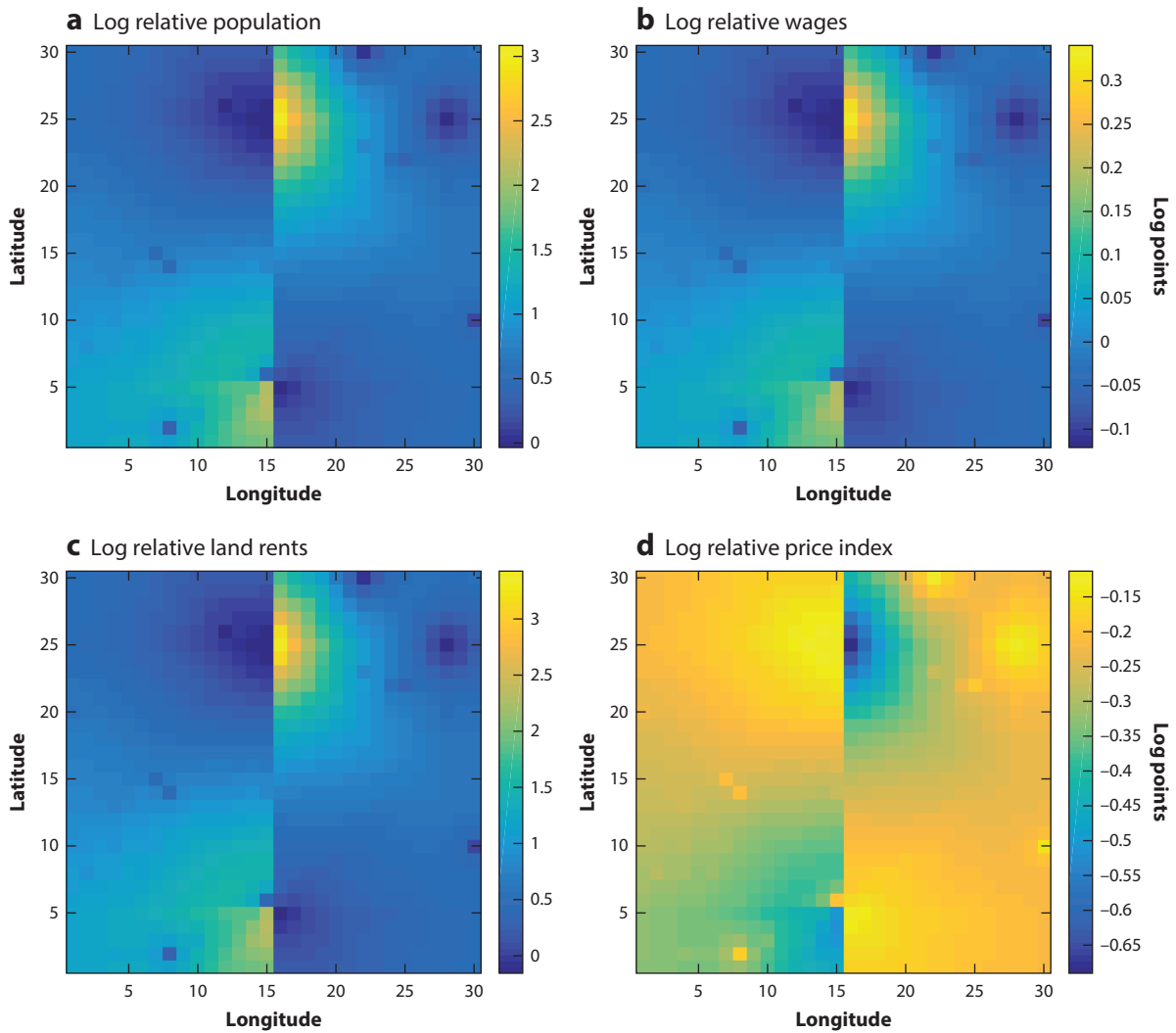


**Figure 3**

External liberalization, showing the ratio of counterfactual to initial values, $\hat{x} = x'/x$: (*a*) log relative population, (*b*) log relative wages, (*c*) log relative land rents, and (*d*) log relative price index. Cold (*blue*) colors correspond to lower values, and hot (*yellow*) colors correspond to higher values.

**Table 1**  Welfare gains from external and internal liberalization

|  | Welfare gains $(\hat{V} = V'/V)$ | |
|---|---|---|
|  | West | East |
| External liberalization | 0.2% | 0.3% |
| Internal liberalization | 1.4% | 2.3% |

**Figure 4**

Internal liberalization, showing the ratio of counterfactual to initial values, $\hat{x} = x'/x$: (*a*) log relative population, (*b*) log relative wages (truncated), (*c*) log relative land rents, and (*d*) log relative price index (truncated). Cold (*blue*) colors correspond to lower values, and hot (*yellow*) colors correspond to higher values.

Intuitively, trade is much larger between regions than between countries, highlighting the greater importance of internal trade frictions relative to external trade frictions.

Although our quantitative analysis in this section is inevitably stylized, particularly because we started from a random productivity distribution, it highlights the power and flexibility of this class of quantitative spatial models. Simply adding population data for a region would allow us to obtain real-world productivity estimates and would make this exercise quite informative about the region's economy and policy options. Although the particular quantitative model considered in this section abstracts from a number of real-world features, such as multiple traded sectors and input–output linkages, it could be further enriched to incorporate these and other salient features of the economy.

# 4. A QUANTITATIVE URBAN MODEL

In this section, we show that the same quantitative methods used to analyze the distribution of economic activity across regions in the previous section can be used to study the internal structure of economic activity within cities. We outline a canonical quantitative urban model following Lucas & Rossi-Hansberg (2002) and Ahlfeldt et al. (2015), in which the agglomeration forces are external economies and the dispersion forces are an inelastic supply of land and commuting costs. To incorporate these features, we select the following items from the menu of building blocks outlined above:

1. From Section 2.1, we select homogeneous goods, single traded sector, endogenous amenities, residential land use, and idiosyncratic preferences.
2. From Section 2.2, we select constant returns to scale, endogenous productivity, no input–output linkages, and commercial land use.
3. From Section 2.3, we select no trade costs and no nontraded goods besides residential and commercial land use.
4. From Section 2.4, we select knowledge externalities, no innovation or dynamics, and perfect transferability of ideas.
5. From Section 2.5, we select perfectly costless migration, costly commuting, single worker type with heterogeneity, and no congestion in transportation.
6. From Section 2.6, we select homogeneous labor, blocks within a single city, and no capital.
7. From Section 2.7, we select perfect competition, equilibrium within a city given prices and utility in an outside economy, absentee landlords, and trade balanced within the city.

We show how the model can be used to quantify the role of productivity and amenities in determining internal city structure and used to evaluate the impact of transport infrastructure improvements.

We consider a city embedded within a larger economy that provides a reservation level of utility ($\bar{U}$). The city consists of a set of discrete blocks indexed by $n, i = \{1, \ldots, S\}$. Each block has a supply of floor space ($H_i$) that depends on geographical land area ($K_i$) and the density of development ($\varphi_i$). There is a single final good that is costlessly traded within the city and to the larger economy and is chosen as the numeraire ($p_i = 1$ for all $i$).[8] Markets are perfectly competitive. The final good is produced from labor and commercial floor space according to a Cobb-Douglas production technology with the following unit cost function,

$$1 = \frac{1}{A_i} w_i^\alpha q_i^{1-\alpha}, \qquad 0 < \alpha < 1, \qquad \qquad 22.$$

where $w_i$ denotes the wage and $q_i$ is the price of commercial floor space. Productivity ($A_i$) in each location can depend on production externalities (e.g., knowledge spillovers) and production fundamentals (e.g., access to natural water). Production externalities are modeled as depending on the travel time–weighted sum of workplace employment density in surrounding blocks, so that

$$A_i = a_i \Upsilon_i^\mu, \qquad \Upsilon_i \equiv \sum_{s=1}^{S} e^{-\delta \tau_{is}} \left( \frac{L_{Ms}}{K_s} \right), \qquad \qquad 23.$$

where $L_{Ms}/K_s$ is workplace employment density per unit of geographical land area, production externalities decline with travel time ($\tau_{is}$) through the iceberg factor $e^{-\delta \tau_{is}} \in (0, 1]$, $\delta$

---

[8]Allen et al. (2016) consider a setting in which final goods are differentiated by origin, trade is costly, and the city corresponds to the entire economy, so that the level of utility ($\bar{U}$) is endogenously determined.

determines their rate of spatial decay, and $\mu$ controls their relative importance in determining overall productivity.

Workers decide whether or not to move to the city before observing idiosyncratic utility shocks for each possible pair of residence and employment locations within the city. If a worker decides to move to the city, she observes these realizations for idiosyncratic utility and picks the pair of residence and employment locations within the city that maximizes her utility. The preferences of a worker $o$ residing in location $n$ and working in location $i$ are defined over consumption of the final good; consumption of residential floor space; residential amenities ($B_n$), which capture common characteristics that make a block a more or less attractive place to live (e.g., leafy streets and scenic views); the disutility of commuting from residence $n$ to workplace $i$ ($\kappa_{ni}$); and the idiosyncratic preference shock ($z_{nio}$), which captures the fact that individual workers can have idiosyncratic reasons for living and working in different parts of the city. Preferences are assumed to take the Cobb-Douglas form with the following indirect utility function:

$$u_{nio} = \frac{z_{nio} B_n w_i}{\kappa_{ni} Q_n^{1-\beta}}, \qquad\qquad 24.$$

where $w_i$ is the wage at the worker's place of employment $i$, $Q_n$ is the residential price of floor space at the worker's place of residence $n$, and the iceberg commuting cost $\kappa_{ni} = e^{\kappa \tau_{ni}} \in [1, \infty)$ increases with the travel time ($\tau_{ni}$) between $n$ and $i$.

Residential amenities ($B_n$) in each location can depend on residential externalities (e.g., crime or local public goods) and residential fundamentals (e.g., scenic views). Residential externalities are modeled symmetrically to production externalities depending on the travel time–weighted sum of residential employment density in surrounding blocks. Thus,

$$B_n = b_n \Omega_n^{\eta}, \qquad \Omega_n \equiv \sum_{r=1}^{S} e^{-\rho \tau_{nr}} \left( \frac{L_{Rr}}{K_r} \right), \qquad\qquad 25.$$

where $L_{Rr}/K_r$ is residence employment density per unit of land area, residential externalities decline with travel time ($\tau_{nr}$) through the iceberg factor $e^{-\rho \tau_{nr}} \in (0, 1]$, $\rho$ determines their rate of spatial decay, and $\eta$ controls their relative importance in overall residential amenities.

Idiosyncratic preferences for each pair of residence location $n$ and workplace location $i$ are drawn from an independent Fréchet distribution,

$$F(z_{nio}) = e^{-T_n E_i z_{nio}^{-\epsilon}}, \qquad T_n, E_i > 0, \ \epsilon > 1, \qquad\qquad 26.$$

where the scale parameter $T_n > 0$ determines the average utility derived from living in location $n$, the scale parameter $E_i$ determines the average utility derived from working in location $i$, and the shape parameter $\epsilon > 1$ controls the dispersion of idiosyncratic utility.

Using the Fréchet distribution for idiosyncratic preferences, one finds that the probability that a worker chooses to live in location $n$ and work in location $i$ depends on the desirability of $n$ as a residence (amenities $B_n$, the Fréchet scale parameter $T_n$, and residential floor prices $Q_n$), the attractiveness of $i$ as an employment location (the Fréchet scale parameter $E_i$ and the wage $w_i$), and the cost of commuting between $n$ and $i$ (as determined by the iceberg commuting cost $\kappa_{ni}$), relative to the attractiveness of all other possible pairs of residence $r$ and employment $s$, namely,

$$\lambda_{ni} = \frac{T_n E_i \left( \kappa_{ni} Q_n^{1-\beta} \right)^{-\epsilon} (B_n w_i)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s (\kappa_{rs} Q_r^{1-\beta})^{-\epsilon} (B_r w_s)^{\epsilon}} \equiv \frac{\Phi_{ni}}{\Phi}. \qquad\qquad 27.$$

Therefore, bilateral commuting exhibits a gravity equation where the flow of commuters between locations $n$ and $i$ depends on both bilateral resistance (bilateral commuting costs $\kappa_{ni}$) and multilateral resistance (commuting costs between all other locations $r$ and $s$).

Population mobility implies that the expected utility from moving to the city must equal the expected utility in the larger economy ($\bar{U}$). Using the Fréchet distribution for idiosyncratic preferences, this population mobility condition implies that

$$\mathbb{E}[u] = \gamma \left[ \sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s \left( \kappa_{rs} Q_r^{1-\beta} \right)^{-\epsilon} (B_r w_s)^{\epsilon} \right]^{1/\epsilon} = \bar{U}, \qquad 28.$$

where $\mathbb{E}$ is the expectations operator and the expectation is taken over the distribution for the idiosyncratic component of utility; $\gamma = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$, and $\Gamma(\cdot)$ is the Gamma function.

Commuting market clearing requires that the measure of workers employed in location $i$ ($L_{Mi}$) equals the sum across locations $n$ of the measure of workers residing in $n$ ($L_{Rn}$) times the probability of commuting to $i$ conditional on residing in $n$, which can be written as

$$L_{Mi} = \sum_{n=1}^{S} \lambda_{ni|n} L_{Rn}, \qquad \lambda_{ni|n} = \frac{E_i \left( w_i / \kappa_{ni} \right)^{\epsilon}}{\sum_{s=1}^{S} E_s \left( w_s / \kappa_{ns} \right)^{\epsilon}}, \qquad 29.$$

where $\lambda_{ni|n}$ is the probability of commuting to location $i$ conditional on residing in location $n$; labor market clearing implies $L_{Mi} = \sum_{i=1}^{S} \sum_{n=1}^{S} \lambda_{ni} L$ and $L_{Ri} = \sum_{i=1}^{S} \sum_{n=1}^{S} \lambda_{ni} L$, where $L$ is the total measure of workers that choose to live in the city. Expected residential income conditional on living in location $n$ ($\bar{v}_n$) differs from the wage in location $n$ because of commuting to work in other locations $i$:

$$\bar{v}_n = \sum_{i \in N} \lambda_{ni|n} w_i. \qquad 30.$$

The no-arbitrage condition across alternative land uses implies that floor space is used entirely commercially ($q_i > \xi_i Q_i$), used entirely residentially ($q_i < \xi_i Q_i$), or allocated to both uses if the commercial price of floor price ($q_i$) equals the residential price of floor space ($Q_i$) net of the tax equivalent of land use regulations ($q_i = \xi_i Q_i$). The observed price of floor space ($\mathbb{Q}_i$) in the data is assumed to equal the maximum of the prices of commercial and residential floor space. Therefore, for blocks that are incompletely specialized in commercial and residential activity, observed floor prices equal commercial floor prices ($\mathbb{Q}_i = q_i = \xi_i Q_i$). Similarly, for blocks that are completely specialized in commercial activity, observed floor prices again equal commercial floor prices ($\mathbb{Q}_i = q_i$). In contrast, for blocks that are completely specialized in residential activity, observed floor prices equal residential floor prices ($\mathbb{Q}_i = Q_i$).

As was the case for the quantitative spatial model in Section 3, analytical results can be provided for the model considered here, including the existence and uniqueness of equilibrium, the ability to invert the model to recover unobserved fundamentals from observed endogenous variables, and the potential to undertake model-based counterfactuals. In the absence of production and residential externalities ($\mu = \eta = 0$), there are no agglomeration forces in the model, and thus the congestion forces of commuting costs and an inelastic supply of land ensure the existence of a unique equilibrium, as shown by Ahlfeldt et al. (2015). In the presence of production and residential externalities ($\mu \neq 0$ or $\eta \neq 0$), there is the potential for multiple equilibria in the model, depending on the strength of these agglomeration forces relative to the exogenous differences in characteristics across locations. This potential multiplicity implies that the mapping from the model's parameters and exogenous location characteristics to its endogenous variables is not unique. Nonetheless, given sufficient data on these endogenous variables, some observed location characteristics, and model parameters, the mapping to the remaining unobserved location

characteristics can be unique. In this case, the unobserved location characteristics (such as production and residential fundamentals) again correspond to structural residuals that exactly rationalize the observed data as an equilibrium of the model, as shown by Ahlfeldt et al. (2015). Intuitively, given sufficient data, the equilibrium conditions of the model, such as utility maximization, profit maximization, and market clearing, can be used to solve for unique values of fundamentals consistent with the observed equilibrium.

Using a similar approach as that in Section 3, the model in this section can be used to undertake counterfactuals for the impact of public policy interventions, such as transport infrastructure improvements, on the spatial distribution of economic activity. For parameter values for which the model has a unique equilibrium, these counterfactuals yield determinate predictions for the impact of the public policy intervention on the spatial distribution of economic activity. For parameter values for which the model has multiple equilibria, counterfactuals can be undertaken assuming an equilibrium selection rule; for example, the initial values from the observed equilibrium can be used to select the closest counterfactual equilibrium.

Although for simplicity we have developed separate quantitative models of goods trade across cities or regions (Section 3) and commuting within a city (Section 4), both sources of spatial linkages are likely to be important in practice. To incorporate both sources of linkages, Monte et al. (2015) develop a unified quantitative model in which a system of regions are connected in both goods markets, through trade, and factor markets, through migration and commuting. Within this unified framework, the effect of changes in the local economic environment on employment depends critically on the ability to attract both migrants and commuters. Although a large literature on local labor markets has sought to estimate a representative local employment elasticity, a key implication of this framework is that the local employment elasticity is heterogeneous across locations. Therefore, an elasticity estimated in one context need not be generalizable to another context.

## 5. QUANTITATIVE EVIDENCE

We have shown how general equilibrium spatial models are typically exactly identified and can be quantified to rationalize the observed data as an equilibrium of the model. We now turn to the empirical literature that has used additional data, assumptions, or sources of variation to provide evidence on the mechanisms in these models, to test their quantitative predictions, and to structurally estimate their parameters. We first discuss the empirical evidence on the role of market access in determining the spatial distribution of economic activity across countries and regions. We next turn to the empirical evidence on productivity and the density of economic activity. Finally, we consider the empirical literature on path dependence and the dynamics of the spatial distribution of economic activity over time.

### 5.1. Market Access

One line of empirical research has examined a key implication of quantitative spatial models, namely, that both wages and population depend on market access.[9] We illustrate this prediction in the context of the quantitative spatial model developed in Section 3. Using CES demand, one

---

[9]A related empirical literature has sought to test the home market effect prediction of models of love of variety, increasing returns to scale, and transport costs that an increase in expenditure on a good should lead to a more than proportionate increase in production of that good, as in the work of Davis & Weinstein (1999, 2003) and Costinot et al. (2016).

finds that profit maximization (Equation 4) and zero profits (Equation 6) imply that the free-on-board price ($p_i$) charged for each variety by a firm in each location $i$ must be low enough to sell the quantity $\bar{x}_i$ and cover the firm's fixed production costs, so that

$$\left(\frac{\sigma}{\sigma - 1}\frac{w_i}{A_i}\right)^{\sigma} = \frac{1}{\bar{x}_i}\sum_{n \in N}(w_n L_n)(P_n)^{\sigma-1}(d_{ni})^{1-\sigma}. \qquad 31.$$

We define the weighted sum of market demands faced by firms as firm market access (FMA$_i$), following Redding & Venables (2004), such that

$$w_i = \xi A_i^{\frac{\sigma-1}{\sigma}}(\text{FMA}_i)^{\frac{1}{\sigma}}, \qquad \text{FMA}_i \equiv \sum_{n \in N}(w_n L_n)(P_n)^{\sigma-1}(d_{ni})^{1-\sigma}, \qquad 32.$$

where $\xi \equiv [F(\sigma - 1)]^{-1/\sigma}(\sigma - 1)/\sigma$ collects together earlier constants. Thus, wages are increasing in both productivity ($A_i$) and firm market access (FMA$_i$).

Market access also affects the price index (Equation 8), which depends on consumers' access to tradeable varieties, as captured by the measure of varieties ($M_i$) and their free-on-board prices ($p_i$) in each location $i$, together with the trade costs of shipping the varieties from locations $i$ to $n$ ($d_{ni}$). We summarize this access to tradeable varieties using the concept of consumer market access (CMA$_n$):

$$P_n = (\text{CMA}_n)^{\frac{1}{1-\sigma}}, \qquad \text{CMA}_n \equiv \sum_{i \in N}M_i(p_i d_{ni})^{1-\sigma}. \qquad 33.$$

Using data on a cross section of countries, Redding & Venables (2004) find a strong correlation between wages and these measures of market access. Using data on counties within the United States, Hanson (2005) finds a similarly strong relationship between wages and market access. However, establishing that these relationships are causal is more challenging. For example, Redding & Venables (2004), Hanson (2005), and Bartelme (2016) all report instrumental variable specifications, but it is difficult to definitively establish that the exclusion restriction of the instruments only affecting wages through market access is satisfied. One line of research has used trade liberalization as a natural experiment that changes the relative market access of regions (see, e.g., Hanson 1996, 1997).[10] Although this evidence from trade liberalization has strengthened the case for a causal interpretation of the role of market access, a remaining concern is that trade liberalization reforms could be endogenous to domestic political economy concerns.

Another line of research has sought to use transport infrastructure as a source of variation in market access. In this case, the key endogeneity concern is that transport infrastructure is unlikely to be randomly assigned. Therefore, a growing reduced-form literature has sought exogenous sources of variation in transport infrastructure, including from routes planned for strategic reasons, historical exploration routes, and inconsequential places that are connected to transport infrastructure merely because they lie along the route between two other locations (for a review, see Redding & Turner 2015). A smaller number of studies have sought to estimate structurally the impact of transport infrastructure improvements within a quantitative spatial model. Donaldson (2016) undertakes a quantitative evaluation of the construction of India's vast railroad network and finds that there is a strong and statistically significant estimated effect of railroads on real income levels but that this effect becomes statistically insignificant after controlling for the model's sufficient statistic of a region's own trade share. This pattern of results is consistent with the view that the estimated effects of railroads are in fact operating through the market access mechanism in the

---

[10]Other studies using trade liberalization as a source of variation in market access include Overman & Winters (2006) for the United Kingdom, Tirado et al. (2002) for early twentieth-century Spain, and Wolf (2007) for early twentieth-century Poland.

model. In a similar vein, Donaldson & Hornbeck (2016) investigate the impact of the expansion of the railroad network on the agricultural sector and show that the overall impact on each location can be captured in terms of its market access. Increases in market access from the expansion in the railroad network from 1870 to 1890 are found to substantially increase county agricultural land values.

To provide further evidence of a causal role for market access, Redding & Sturm (2008) use the division of Germany after World War II as a natural experiment that provides plausibly exogenous variation in market access. The division of Germany has a number of attractive features for isolating the role played by market access across West German cities depending on their proximity to the new border with East Germany. First, in contrast to cross-country studies, there is no obvious variation in institutions across cities within West Germany. Second, there are no obvious changes in natural advantage, such as access to navigable rivers or coasts, climatic conditions, or disease environment. Third, the change in market access following German division is much larger than what is typically observed in other contexts, and the effects can be observed over a long period of time. Fourth, the drawing of the border dividing Germany into East and West Germany was based on military considerations that are unlikely to be correlated with the predivision characteristics of cities.[11]

In the quantitative model from Section 3, the treatment effect of division on border cities relative to those further from the new border depends on two parameter combinations that capture (*a*) the strength of agglomeration and dispersion forces [$\sigma(1-\alpha)$] and (*b*) the elasticity of trade with respect to distance [$(\sigma-1)\phi$, where $\phi$ is the elasticity of trade costs with respect to distance]. Redding & Sturm (2008) undertake a quantitative analysis of the model and show that, for plausible values of these parameter combinations that satisfy the condition $\sigma(1-\alpha) > 1$ for a unique equilibrium, the model can account quantitatively for both the average treatment effect of division and the larger treatment effect for small than for large cities.

## 5.2. Productivity and Density

A large body of empirical literature finds that wages, land prices, productivity, employment, and employment growth are positively correlated with population density (for a survey, see Moretti 2011a). In their survey of this empirical literature, Rosenthal & Strange (2004) report that the elasticity of productivity with respect to the density of economic activity is typically estimated to lie within the range of 3–8%.[12] However, establishing that this correlation is indeed causal remains challenging, and a relatively small number of studies have sought exogenous sources of variation in the surrounding concentration of economic activity. For example, Rosenthal & Strange (2008) and Combes et al. (2010) use geology as an instrument for population density, exploiting the idea that tall buildings are easy to construct where solid bedrock is accessible. Greenstone et al. (2010) provide evidence on agglomeration spillovers by comparing changes in total factor productivity

---

[11]Further evidence in support of the role of market access has been provided by a number of subsequent studies. Using detailed data on whether West German municipalities qualified for the place-based Zonenrandgebiet (ZRG) policy, Ehrlich & Seidel (2015) find even larger market access effects after conditioning on ZRG qualification. Using the opening of Central and Eastern European markets after the fall of the Iron Curtain in 1990, Brülhart et al. (2012) find substantial increases in both wages and employment for Austrian municipalities within 50 km of the former Iron Curtain. Using the economic separation of Japan and Korea after World War II and implementing the same empirical specification as that of Redding & Sturm (2008), Nakajima (2008) finds a similar pattern of market access effects.

[12]In a recent meta-analysis of estimates of urban agglomeration economies, Melo et al. (2009) report a mean estimate of 0.058 across 729 estimates from 34 studies, consistent with the results of Rosenthal & Strange (2004).

among incumbent plants in winning counties that attracted a large manufacturing plant and losing counties that were the new plant's runner-up choice.

Several recent studies have used exogenous variation from natural experiments to examine whether estimated agglomeration economies are consistent with the predictions of quantitative spatial models. Combining data from an urban revitalization program in Richmond, Virginia, between 1999 and 2004 with a structural model of residential externalities, Rossi-Hansberg et al. (2010) estimate substantial and highly localized housing externalities. Land prices in neighborhoods targeted for revitalization rose by 2–5% more per annum than those in a control neighborhood. With every 1,000 ft of distance from the subsidized area, housing externalities are estimated to decline by around one-half.[13]

Using the Tennessee Valley Authority (TVA) as a natural experiment and a structural model, Kline & Moretti (2014a) provide evidence on the long-run effects of one of the most ambitious regional development programs in US history. Using as controls authorities that were proposed but never approved by Congress, Kline & Moretti (2014a) found that the TVA led to large gains in agricultural employment that were eventually reversed when the program's subsidies ended. In contrast, gains in manufacturing employment were found to intensify well after federal transfers had lapsed, consistent with agglomeration economies in manufacturing.

Using the division of Berlin following World War II and its reunification after the fall of the Iron Curtain as an exogenous source of variation in the surrounding density of economic activity, Ahlfeldt et al. (2015) structurally estimate the parameters of the quantitative spatial model explained in Section 4. Following the city's division, there was a reorientation of the gradient in land prices and employment in West Berlin away from the main pre-war concentration of economic activity in East Berlin, whereas the city's reunification led to a reemergence of these gradients. The model's parameters are identified from the assumption that the systematic change in the pattern of economic activity in West Berlin following division and reunification is explained by the mechanisms of the model (changes in commuting access and production and residential externalities) rather than by systematic changes in the pattern of structural residuals (production and residential fundamentals).

Both productivity and amenities are found to exhibit substantial and highly localized agglomeration externalities. The estimated elasticities of productivity and amenities with respect to the surrounding densities of workplace employment and residence employment are $\mu = 0.07$ and $\eta = 0.15$, respectively. Undertaking counterfactuals for the impact of division and reunification, one finds that the special case of the model without agglomeration forces ($\mu = \eta = 0$) is unable to account quantitatively for the observed reallocations of economic activity in the data. In contrast, for the estimated values of production and residential externalities, the model is successful in matching the observed impacts of division and reunification, both qualitatively and quantitatively.

Other research, such as that of Bernard et al. (2016), has examined the role of urban property rights in influencing internal city structure in a developing-country context. Further quantitative evidence on the role of residential externalities in influencing the spatial distribution of economic activity is provided by Diamond (2015). From 1980 to 2000, the rise in the US wage gap between college and high-school graduates coincided with increased geographic sorting as college graduates concentrated in high-wage, high-rent cities. A structural spatial equilibrium model is used to evaluate the causes and welfare consequences of this increased skill sorting. Although local labor demand changes fundamentally caused the increased skill sorting, it was further fueled by

---

[13]Examining the end of rent control in Cambridge, Massachusetts, Autor et al. (2014) provide reduced-form evidence of substantial and, again, highly localized housing market spillovers.

endogenous increases in amenities (residential externalities) within higher-skill cities. Changes in cities' wages, rents, and endogenous amenities increased inequality between high-school and college graduates by more than would be suggested by the increase in the college wage gap alone.

Several of the above studies find evidence of residential externalities; the microeconomic determinants of these externalities and their implications remain interesting areas for further research.

## 5.3. The Evolution of the Spatial Distribution of Economic Activity

Motivated in part by models with multiple equilibria, one line of empirical research on changes in the spatial distribution of economic activity over time has set out to look for evidence of path dependence, defined as temporary shocks permanently affecting the distribution of economic activity. Davis & Weinstein (2002, 2008) use Japanese war-time bombing as such an exogenous temporary shock and find little evidence of path dependence for either the distribution of population as a whole or employment in individual industries (for other research exploiting war-time bombing as an exogenous shock, see, e.g., Bosker et al. 2007, 2008; Miguel & Roland 2011). Subsequent studies have provided a number of apparent examples of path dependence using a variety of alternative sources of exogenous variation. Redding et al. (2011) find path dependence for the location of Germany's air hub using the natural experiment of Germany's division and reunification.[14] Bleakley & Lin (2012) find, using variation from portage sites in the United States, that a temporary historical advantage had permanent effects on the spatial distribution of population. Hornbeck & Keniston (2014) find long-lived effects of the Boston fire through the potential for large-scale rebuilding using plot-level data on land values. Michaels & Rauch (2016) report results consistent with path dependence using data on Roman cities.

Although we now have strong evidence of path dependence in the distribution of economic activity, the interpretation of these empirical results remains open to question. In particular, findings of path dependence do not necessarily imply multiple equilibria. If historical advantages lead to investments in local technology and infrastructure and better institutions, dynamic agglomeration effects can lead to maintained investments in these locations. These investments can serve as new sources of local advantages even after the original historical advantage has become obsolete or irrelevant. This is the case in the dynamic models of Desmet & Rossi-Hansberg (2015) and Desmet et al. (2016). The latter article shows that these dynamic agglomeration effects have significant predictive power over long periods. In particular, using only information for the year 2000 and running the dynamics backwards, Desmet et al. (2016) find a correlation between actual and predicted country population levels in 1950 as high as 0.74. In these models, the equilibrium is unique, and so the evolution of the distribution of economic activity is uniquely determined by initial conditions. In that sense, any shock or change in fundamentals in a given period will affect the future evolution and the balanced-growth path of the world economy. Empirical and theoretical research on the dynamic evolution of the distribution of economic activity is reviewed by Desmet & Henderson (2015).

Clearly, more theoretical and empirical work on the role of shocks and initial conditions in determining the evolution of economic activity is needed. Perhaps this will be one of the more fruitful areas for future research on quantitative spatial economics.

---

[14]For structural estimations of the determinants of the location of particular economic activities, readers are referred to Holmes (2005) for headquarter location choices and Holmes (2011) for the expansion of Walmart's distribution and retail network.

## 6. CONCLUSIONS

The development of quantitative spatial models has been a major breakthrough in our understanding of the determinants of the spatial distribution of economic activity. Relative to earlier theoretical research, these models connect in an empirically meaningful way with the observed data, thereby permitting the quantification of key theoretical mechanisms and specificity in evaluating the counterfactual effects of policy interventions in particular settings. These models have also delivered both methodological and substantive general insights. The methodological insights include a list of components that can be used to incorporate a variety of agglomeration and congestion mechanisms in a unified spatial framework. Within this framework, they also include the conditions for the existence and uniqueness of equilibrium, the conditions under which these models can be inverted to separate out the contributions of physical and economic geography, and methods for undertaking counterfactuals to evaluate comparative statics with respect to changes in the model's parameters. The substantive insights include the empirical relevance of market access, the strength of agglomeration and dispersion forces within cities, and the importance of the dynamic forces shaping the evolution of the spatial distribution of economic activity over time.

This literature on quantitative spatial models has already achieved much. Nonetheless, there remain many areas where further research is needed. First, most research has continued to be concerned with the production and trade of goods, whereas much economic activity today is concentrated in services, whether tradable or nontradable. Second, most of the main frameworks in the literature are static and abstract from the effect of spatial frictions on the evolution of the spatial distribution of economic activity and growth. Third, although there have been several influential studies of the sorting of heterogeneous workers and firms across geographic space, there remains scope for further work. Fourth, the economic analysis of the geography of firm and worker networks remains underexplored. We expect much progress in these and other dimensions over the coming decades.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

## LITERATURE CITED

Ahlfeldt GM, Redding SJ, Sturm DM, Wolf N. 2015. The economics of density: evidence from the Berlin Wall. *Econometrica* 83(6):2127–89

Albouy D. 2016. *Are big cities bad places to live? Estimating quality of life across metropolitan areas*. Unpublished manuscript, Univ. Illinois, Urbana-Champaign

Allen T, Arkolakis C. 2014. Trade and the topography of the spatial economy. *Q. J. Econ.* 129(3):1085–140

Allen T, Arkolakis C, Li X. 2016. *Optimal city structure*. Unpublished manuscript, Yale Univ., New Haven, NJ

Allen T, Arkolakis C, Takahashi Y. 2015. *Universal gravity*. NBER Work. Pap. 20787

Alonso W. 1964. *Location and Land Use*. Cambridge, MA: Harvard Univ. Press

Anderson JE, van Wincoop E. 2003. Gravity with gravitas: a solution to the border puzzle. *Am. Econ. Rev.* 93(1):170–92

Anderson ML. 2014. Subways, strikes, and slowdowns: the impacts of public transit on traffic congestion. *Am. Econ. Rev.* 104(9):2763–96

Arkolakis C, Costinot A, Rodriguez-Clare A. 2012. New trade models, same old gains. *Am. Econ. Rev.* 102(1):94–130

Arkolakis C, Ramondo N, Rodriguez-Clare A, Yeaple S. 2014. *Innovation and production in the global economy*. NBER Work. Pap. 18792

Armington PS. 1969. A theory of demand for products distinguished by place of production. *IMF Staff Pap.* 16(1):159–78

Artuç ES, Chaudhuri S, McLaren J. 2010. Trade shocks and labor adjustment: a structural empirical approach. *Am. Econ. Rev.* 100(3):1008–45

Arzaghi M, Henderson JV. 2008. Networking off Madison Avenue. *Rev. Econ. Stud.* 75(4):1011–38

Autor DH, Palmer CJ, Pathak PA. 2014. Housing market spillovers: evidence from the end of rent control in Cambridge, Massachusetts. *J. Polit. Econ.* 122(3):661–717

Baldwin R, Forslid R, Martin P, Ottaviano GIP, Robert-Nicoud F. 2003. *Economic Geography and Public Policy*. Princeton, NJ: Princeton Univ. Press

Baldwin R, Okubo T. 2005. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *J. Econ. Geogr.* 6(3):323–46

Bartelme D. 2016. *Trade costs and economic geography: evidence from the US*. Unpublished manuscript, Univ. Mich., Ann Arbor

Behrens K, Mion G, Murata Y, Südekum J. 2014. *Spatial frictions*. Düsseldorf Inst. Compet. Econ. Discuss. Pap. 160, Heinrich-Heine-Univ., Düsseldorf, Ger.

Bernard L, Bird J, Venables AJ. 2016. *The urban land market: a computable equilibrium model applied to Kampala City*. Unpublished manuscript, Oxford Univ., Oxford, UK

Bleakley H, Lin J. 2012. Portage and path dependence. *Q. J. Econ.* 127:587–644

Bosker M, Brakman S, Garretsen H, Schramm M. 2007. Looking for multiple equilibria when geography matters: German city growth and the WWII shock. *J. Urban Econ.* 61(1):152–69

Bosker M, Brakman S, Garretsen H, Schramm M. 2008. A century of shocks: the evolution of the German city size distribution 1925–1999. *Reg. Sci. Urban Econ.* 38(4):330–47

Brinkman J. 2016. Congestion, agglomeration, and the structure of cities. *J. Urban Econ.* 94:13–31

Brülhart M, Carrère C, Trionfetti F. 2012. How wages and employment adjust to trade liberalization: quasi-experimental evidence from Austria. *J. Int. Econ.* 86:68–81

Bryan G, Morten M. 2015. *Economic development and the spatial allocation of labor: evidence from Indonesia*. Unpublished manuscript, Stanford Univ., Stanford, CA

Busso M, Gregory J, Kline P. 2013. Assessing the incidence and efficiency of a prominent place based policy. *Am. Econ. Rev.* 103(2):897–947

Caliendo L, Dvorkin M, Parro F. 2015. *Trade and labor market dynamics*. NBER Work. Pap. 21149

Caliendo L, Parro F, Rossi-Hansberg E, Sarte PD. 2014. *The impact of regional and sectoral productivity changes on the US economy*. NBER Work. Pap. 20168

Combes P, Duranton G, Gobillon L. 2016. *The costs of agglomeration: land prices in French cities*. Unpublished manuscript, Wharton Sch., Univ. Pa., Philadelphia

Combes P, Duranton G, Gobillon L, Roux S. 2010. Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economies*, ed. EL Glaeser, pp. 16–66. Chicago: Chicago Univ. Press

Comin D, Dmitriev M, Rossi-Hansberg E. 2013. *The spatial diffusion of technology*. Unpublished manuscript, Princeton Univ., Princeton, NJ

Coşar AK, Fajgelbaum PD. 2016. Internal geography, international trade, and regional specialization. *Am. Econ. J. Microecon.* 8(1):24–56

Costinot A, Donaldson D, Kyle M, Williams H. 2016. *The more we die the more we sell: a simple test of the home-market effect*. Unpublished manuscript, Stanford Univ., Stanford, CA

Davis DR, Dingel JI. 2015. *The comparative advantage of cities*. Unpublished manuscript, Columbia Univ., New York

Davis DR, Weinstein DE. 1999. Economic geography and regional production structure: an empirical investigation. *Eur. Econ. Rev.* 43:379–407

Davis DR, Weinstein DE. 2002. Bones, bombs, and break points: the geography of economic activity. *Am. Econ. Rev.* 92(5):1269–89

Davis DR, Weinstein DE. 2003. Market access, economic geography and comparative advantage: an empirical assessment. *J. Int. Econ.* 59(1):1–23

Davis DR, Weinstein DE. 2008. A search for multiple equilibria in urban industrial structure. *J. Reg. Sci.* 48(1):29–65

Davis MA, Ortalo-Magné F. 2011. Housing expenditures, wages, rents. *Rev. Econ. Dyn.* 14(2):248–61

Dekle R, Eaton J, Kortum S. 2007. Unbalanced trade. *Am. Econ. Rev.* 97(2):351–55

Desmet K, Henderson JV. 2015. The geography of development within countries. See Duranton et al. 2015, pp. 1457–517

Desmet K, Nagy DK, Rossi-Hansberg E. 2016. The geography of development. *J. Polit. Econ.* In press

Desmet K, Rossi-Hansberg E. 2013. Urban accounting and welfare. *Am. Econ. Rev.* 103(6):2296–327

Desmet K, Rossi-Hansberg E. 2014. Spatial development. *Am. Econ. Rev.* 104(4):1211–43

Desmet K, Rossi-Hansberg E. 2015. On the spatial economic impact of global warming. *J. Urban Econ.* 88:16–37

Diamond R. 2016. The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. *Am. Econ. Rev.* 106(3):479–524

Donaldson D. 2016. Railroads of the Raj: estimating the impact of transportation infrastructure. *Am. Econ. Rev.* In press

Donaldson D, Hornbeck R. 2016. Railroads and American economic growth: a "market access" approach. *Q. J. Econ.* 131(2):799–858

Downs A. 1962. The law of peak-hour expressway congestion. *Traffic Q.* 16(3):393–409

Duranton G, Henderson JV, Strange WC, eds. 2015. *Handbook of Regional and Urban Economics*, Vol. 5: *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier

Duranton G, Puga D. 2004. Micro-foundations of urban agglomeration economies. See Henderson & Thisse 2004, pp. 2063–117

Duranton G, Turner MA. 2011. The fundamental law of road congestion: evidence from U.S. cities. *Am. Econ. Rev.* 101(6):2616–52

Eaton J, Kortum S. 2002. Technology, geography, and trade. *Econometrica* 70(5):1741–79

Ehrlich M, Seidel T. 2015. *The persistent effects of place-based policy: evidence from the West German Zonenrangebiet*. Unpublished manuscript, Univ. Bern, Switz.

Ellison G, Glaeser EL, Kerr WR. 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *Am. Econ. Rev.* 100(3):1195–213

Epple D, Gordon B, Sieg H. 2010. A new approach to estimating the production function for housing. *Am. Econ. Rev.* 100(3):905–24

Fajgelbaum PD, Morales E, Suárez Serrato JC, Zidar O. 2015. *State taxes and spatial misallocation*. NBER Work. Pap. 21760

Fajgelbaum PD, Redding SJ. 2014. *External integration, structural transformation and economic development: evidence from Argentina 1870–1914*. NBER Work. Pap. 20217

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

Fujimoto T, Krause U. 1985. Strong ergodicity for strictly increasing nonlinear operators. *Linear Algebra Appl.* 71(1):101–12

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

Fujita M, Ogawa H. 1982. Multiple equilibria and structural transformation of non-monocentric urban configurations. *Reg. Sci. Urban Econ.* 12(2):161–96

Fujita M, Thisse JF. 2002. *The Economics of Agglomeration*. Cambridge, UK: Cambridge Univ. Press

Galle S, Rodriguez-Clare A, Yi M. 2015. *Slicing the pie: quantifying the aggregate and distributional consequences of trade*. Unpublished manuscript, Univ. Calif., Berkeley

Gaubert C. 2015. *Firm sorting and agglomeration*. Unpublished manuscript, Univ. Calif., Berkeley

Greenstone M, Hornbeck R, Moretti E. 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *J. Polit. Econ.* 118(3):536–98

Grogger J, Hanson G. 2011. Income maximization and the selection and sorting of international migrants. *J. Dev. Econ.* 95(1):42–57

Hanson GH. 1996. Localization economies, vertical organization, and trade. *Am. Econ. Rev.* 86(5):1266–78

Hanson GH. 1997. Increasing returns, trade, and the regional structure of wages. *Econ. J.* 107:113–33

Hanson GH. 2005. Market potential, increasing returns, and geographic concentration. *J. Int. Econ.* 67(1):1–24

Head K, Mayer T. 2014. Gravity equations: workhorse, toolkit, and cookbook. In *Handbook of International Economics*, Vol. 4, ed. G Gopinath, E Helpman, K Rogoff, pp. 131–95. Amsterdam: Elsevier

Helpman E. 1998. The size of regions. In *Topics in Public Economics: Theoretical and Applied Analysis*, ed. D Pines, E Sadka, I Zilcha, pp. 33–54. Cambridge, UK: Cambridge Univ. Press

Henderson JV. 1974. The sizes and types of cities. *Am. Econ. Rev.* 64(4):640–56

Henderson JV, Thisse J-F, eds. 2004. *Handbook of Regional and Urban Economics*, Vol. 4: *Cities and Geography*. Amsterdam: Elsevier

Holmes TJ. 2005. The location of sales offices and the attraction of cities. *J. Polit. Econ.* 113(3):551–81

Holmes TJ. 2011. The diffusion of Walmart and the economies of density. *Econometrica* 79(1):253–302

Holmes TJ, Sieg H. 2015. Structural estimation in urban economics. See Duranton et al. 2015, pp. 69–114

Hornbeck R, Keniston D. 2014. *Creative destruction: barriers to urban growth and the Great Boston Fire of 1872*. Unpublished manuscript, Chicago Booth Sch. Bus., Univ. Chicago

Jaffe AB, Trajtenberg M, Henderson R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* 108(3):577–98

Kennan J, Walker JR. 2011. The effect of expected income on individual migration decisions. *Econometrica* 79(1):211–51

Kline P, Moretti E. 2014a. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *Q. J. Econ.* 129(1):275–331

Kline P, Moretti E. 2014b. People, places, and public policy: some simple welfare economics of local economic development policies. *Annu. Rev. Econ.* 6:629–62

Krugman PR. 1991a. *Geography and Trade*. Cambridge, MA: MIT Press

Krugman PR. 1991b. Increasing returns and economic geography. *J. Polit. Econ.* 99(3):483–99

Krugman PR, Venables AJ. 1995. Globalisation and the inequality of nations. *Q. J. Econ.* 60:857–80

Lucas RE. 2000. Externalities and cities. *Rev. Econ. Dyn.* 4(2):245–74

Lucas RE, Rossi-Hansberg E. 2002. On the internal structure of cities. *Econometrica* 70(4):1445–76

Marshall A. 1920. *Principles of Economics*. London: Macmillan

McFadden D. 1974. The measurement of urban travel demand. *J. Public Econ.* 3(4):303–28

Melo PC, Graham DJ, Noland RB. 2009. A meta-analysis of estimates of urban agglomeration economies. *Reg. Sci. Urban Econ.* 39(3):332–42

Michaels G, Rauch F. 2016. Resetting the urban network: 117–2012. *Econ. J.* In press

Miguel E, Roland G. 2011. The long-run impact of bombing Vietnam. *J. Dev. Econ.* 96:1–15

Mills ES. 1967. An aggregative model of resource allocation in a metropolitan centre. *Am. Econ. Rev.* 57(2):197–210

Monte F. 2015. *The local incidence of trade shocks*. Unpublished manuscript, Georgetown Univ., Washington, DC

Monte F, Redding SJ, Rossi-Hansberg E. 2015. *Commuting, migration and local employment elasticities*. NBER Work. Pap. 21706

Moretti E. 2011a. Local labor markets. In *Handbook of Labor Economics*, Vol. 4B, ed. D Card, O Ashenfelter, pp. 1238–303. Amsterdam: Elsevier

Moretti E. 2011b. Local multipliers. *Am. Econ. Rev.* 100(2):1–7

Morten M, Oliveira J. 2016. *Paving the way to development: costly migration and labor market integration*. Unpublished manuscript, Stanford Univ., Stanford, CA

Muth R. 1969. *Cities and Housing*. Chicago: Univ. Chicago Press

Nagy DK. 2016. *City location and economic development*. Unpublished manuscript, Princeton Univ., Princeton, NJ

Nakajima K. 2008. Economic division and spatial relocation: the case of postwar Japan. *J. Jpn. Int. Econ.* 22(3):383–400

Neumark D, Simpson H. 2014. Place-based policies. See Duranton et al. 2015, pp. 1197–1287

Overman HG, Redding SJ, Venables AJ. 2003. The economic geography of trade, production and income: a survey of empirics. In *Handbook of International Trade*, ed. E Kwan-Choi, J Harrigan, pp. 353–87. Oxford, UK: Basil Blackwell

Overman HG, Winters LA. 2006. *Trade shocks and industrial location: the impact of EEC accession on the UK*. Discuss. Pap. 588, Centre Econ. Perform., London Sch. Econ., London

Puga D. 1999. The rise and fall of regional inequalities. *Eur. Econ. Rev.* 43:303–34

Redding SJ. 2010. The empirics of new economic geography. *J. Reg. Sci.* 50(1):297–311

Redding SJ. 2011. Economic geography: a review of the theoretical and empirical literature. In *Palgrave Handbook of International Trade*, ed. D Bernhofen, R Falvey, D Greenaway, U Kreickemeier, pp. 497–531. London: Palgrave Macmillan

Redding SJ. 2016. Goods trade, factor mobility and welfare. *J. Int. Econ.* 101:148–67

Redding SJ, Sturm DM. 2008. The costs of remoteness: evidence from German division and reunification. *Am. Econ. Rev.* 98(5):1766–97

Redding SJ, Sturm DM. 2016. *Estimating neighborhood effects: evidence from war-time destruction in London*. Unpublished manuscript, Princeton Univ., Princeton, NJ

Redding SJ, Sturm DM, Wolf N. 2011. History and industrial location: evidence from German airports. *Rev. Econ. Stat.* 93(3):814–31

Redding SJ, Turner M. 2015. Transportation costs and the spatial organization of economic activity. In *Handbook of Urban and Regional Economics*, Vol. 5, ed. G Duranton, JV Henderson, W Strange, pp. 1339–98. Amsterdam: Elsevier

Redding SJ, Venables A. 2004. Economic geography and international inequality. *J. Int. Econ.* 62(1):53–82

Reyes-Heroles R. 2016. *The role of trade costs in the surge of trade imbalances*. Unpublished manuscript, Princeton Univ., Princeton, NJ

Roback J. 1982. Wages, rents, and the quality of life. *J. Polit. Econ.* 90:1257–78

Rosen S. 1979. Wages-based indexes of urban quality of life. In *Current Issues in Urban Economics*, ed. P Mieszkowski, M Straszheim, pp. 74–104. Baltimore: Johns Hopkins Univ. Press

Rosenthal SS, Strange WC. 2004. Evidence on the nature and sources of agglomeration economics. See Henderson & Thisse 2004, pp. 2119–71

Rosenthal SS, Strange WC. 2008. The attenuation of human capital spillovers. *J. Urban Econ.* 64(2):373–89

Rossi-Hansberg E. 2005. A spatial theory of trade. *Am. Econ. Rev.* 95(5):1464–91

Rossi-Hansberg E, Sarte P, Owens RI. 2010. Housing externalities. *J. Polit. Econ.* 118(3):485–535

Shoven JB, Whalley J. 2005. *Applying General Equilibrium*. Cambridge, UK: Cambridge Univ. Press

Simonovska I, Waugh ME. 2014. The elasticity of trade: estimates and evidence. *J. Int. Econ.* 92(1):34–50

Tirado DA, Paluzie E, Pons J. 2002. Economic integration and industrial location: the case of Spain before WWI. *J. Econ. Geogr.* 2(3):343–63

Waugh M. 2010. International trade and income differences. *Am. Econ. Rev.* 100(5):2093–124

Wolf N. 2007. Endowments versus market potential: What explains the relocation of industry after the Polish Reunification in 1918? *Explor. Econ. Hist.* 44(1):22–42

Young A. 2013. Inequality, the urban-rural gap, and migration. *Q. J. Econ.* 128:1727–85

Zabreyko P, Koshelev A, Krasnoselskii M, Mikhlin S, Rakovshchik L, Stetísenko V. 1975. *Integral Equations: A Reference Text*. Leiden, Neth.: Noordhoff Int. Publ.

# Contents

**Indexes**

**Errata**

An online log of corrections to *Annual Review of Economics* articles may be found at
http://www.annualreviews.org/errata/economics