

NBER WORKING PAPER SERIES

QUANTITATIVE SPATIAL ECONOMICS

Stephen J. Redding  
Esteban Rossi-Hansberg

Working Paper 22655  
<http://www.nber.org/papers/w22655>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2016

This paper is prepared for the Annual Review of Economics. We thank the IES and Princeton for research support. We are grateful to co-authors and colleagues for insightful comments and discussion, including Lorenzo Caliendo, Klaus Desmet, Dave Donaldson, Pablo Fajgelbaum, Gene Grossman, Ferdinando Monte, Eduardo Morales, Henry Overman, Daniel Sturm, Tony Venables, and Nikolaus Wolf. Responsibility for any views, errors and omissions lies with the authors alone. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w22655.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Stephen J. Redding and Esteban Rossi-Hansberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Quantitative Spatial Economics  
Stephen J. Redding and Esteban Rossi-Hansberg  
NBER Working Paper No. 22655  
September 2016  
JEL No. F10,F14,R12,R23,R41

### **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 recently, the theoretical literature on economic geography had focused on stylized settings that could not easily be taken to the data. This paper 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. Yet 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.

Stephen J. Redding  
Department of Economics  
and Woodrow Wilson School  
Princeton University  
Fisher Hall  
Princeton, NJ 08544  
and NBER  
reddings@princeton.edu

Esteban Rossi-Hansberg  
Princeton University  
Department of Economics  
Fisher Hall  
Princeton, NJ 08544-1021  
and NBER  
erossi@princeton.edu

# 1 Introduction

Economic activity is highly unevenly distributed across space, as reflected in 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 strength of the agglomeration and dispersion forces that underlie these concentrations of economic activity is 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. Yet at the same time these models are sufficiently tractable as to enable quantitative counterfactuals to evaluate empirically-meaningful policies and counterfactual scenarios. In this paper, 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. A distinguishing feature of this field relative to the study of international trade is that economic agents are typically assumed to be geographically mobile. Early theoretical research on “new economic geography” (as synthesized in 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 forces. 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 like a few 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 (the Lucas Critique). Furthermore, it was unclear the extent to which theoretical results for stylized spatial settings would generalize qualitatively and quantitatively to more realistic environments.<sup>1</sup>

Following the introduction of quantitative models of international trade (in particular Eaton & Kortum, 2002), more recent 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 instead aims 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, there is a whole series of predictions for which these models are isomorphic to one another (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, through accommodating many regions and a rich geography of trade costs, these models provide microfoundations for central features of the data. Second, through allowing for many regions that can differ in their productivity and amenities, as well as potentially a number of other characteristics, these models are sufficiently rich as to be able 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 contri-

---

<sup>1</sup>For reviews of the earlier theoretical and empirical literature on new economic geography, see for example Overman et al. (2003) and Redding (2010, 2011).

butions of each of them. Inevitably, this analysis is conditional on the assumed model, and different models in general 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 and 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. A necessary assumption for these exercises to be valid is that the identified structural fundamentals are stable and invariant to the analyzed policy interventions.<sup>2</sup> 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.<sup>3</sup> These interactions and general equilibrium effects are typically not identified in reduced-form difference-in-difference approaches, because differencing between the treatment and control group eliminates any effect that is common to both groups. Hence, a key implication of this analysis is that locations are not independent observations in a cross-section regression, but are rather systematically linked to one another through trade, commuting and migrations flows. Not recognizing this inter-dependence in reduced-form empirical analysis can lead to significant biases and substantial heterogeneity in treatment effects that threatens the external validity of the results (as shown for example in 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 the 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 hence 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: if a railroad is built between these cities in

---

<sup>2</sup>For a general review of structural estimation approaches in urban economies, see Holmes & Sieg (2014).

<sup>3</sup>Quantitative spatial models share some features with the earlier Computable General Equilibrium (CGE) literature in international trade, as reviewed in Shoven and Whalley (2005). Two key differences are the ability to obtain sharp analytical results and transparent counterfactuals, in addition to the focus on economic geography rather than international trade.

this country at this time, what is the quantitative effect on these particular regions, sectors and factors of production? This specificity not only can 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.

On top of the quantitative evaluation of specific counterfactuals and policy exercises, the existing research on quantitative spatial models has yielded two main general sets of insights relative to 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 is 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. Following the division of Germany after the end of the Second World War, this mechanism can account for the observed decline in the relative size of West German cities close to the new border with East Germany of around one third (Redding & Sturm 2008). Removing all railroads constructed up to 1890 would have reduced the value of agricultural land in the United States by around 60 percent, with limited potential for mitigating these losses through feasible extensions to the canal network (Donaldson & Hornbeck 2016). Second, canonical models of urban economics (as in Fujita & Ogawa 1982, and Lucas & Rossi-Hansberg 2002) can account quantitatively for the observed gradients of economic activity within cities (as in Ahlfeldt et al. 2015). The estimated parameter values imply substantial and highly localized agglomeration externalities, both for 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 and 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 for the counterfactual dynamic response of the economy to global migration and trade policy changes, as well as global shocks such as climate change (e.g. Desmet & Rossi-Hansberg

2015, Desmet et al. 2016 and Nagy 2016).

The remainder of the paper 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 are linked with one another 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 implicitly or explicitly makes 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 the location of economic agents relative to one another is of consequence: 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 the ensuing sections of the paper, 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.

**(1) Preferences:** Assumptions about preferences play a central role in shaping consumers' location decisions. Five main sets of assumptions about preferences can be distinguished.

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

**(B) Single versus multiple sectors:** To preserve analytical tractability theoretical models of economic

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

**(C) Exogenous amenities (e.g scenic views) and/or endogenous amenities (e.g. crime):** Early new economic geography models such as Krugman (1991a) 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 characteristics that makes a location a more or less desirable place of residence. A number of studies have also argued that in order to match the response of the local economy to external shocks, it is important to allow for endogenous as well as exogenous amenities, as in Ahlfeldt et al. (2015) and Diamond (2016) among others.

**(D) Fixed local factors in utility (residential land):** The presence of such fixed factors in utility acts as a congestion or dispersion force (as in Helpman 1998 and more recently Monte et al. 2015).

**(E) 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, no arbitrage 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 (as in Artuc et al. 2010, Busso et al. 2013, Grogger & Hanson 2011, and Kennan & Walker 2011, among others). 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, albeit they have different predictions for wages (see for example Galle et al. 2015).

**(2) Production Technology:** Assumptions about production technology critically influence firms' location decisions. Four main sets of selections concerning the production technology can be distinguished.

**(A) 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 pro-

cess 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 (as in Armington 1969) and Ricardian technology differences (as in Eaton & Kortum 2002) can provide alternative mechanisms for specialization from the love of variety and increasing returns to scale in new economic geography models.

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

**(C) Input-output linkages:** Input-output linkages play a key role in determining how productivity shocks in a particular sector and region spread through the wider economy and shape local multipliers (the extent to which an increase in expenditure in one sector leads to more than proportionate increases in overall expenditure through increased demand for other sectors). Such input-output linkages provide an additional mechanism for agglomeration (as in 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).

**(D) Fixed local factors in production (commercial land):** The presence of such fixed factors again acts as a congestion force (as in Rossi-Hansberg 2005 and Ahlfeldt et al. 2015).

**(3) Technology for Trading Goods:** The first mechanism that explains why the location of agents relative to one another is consequential in quantitative spatial models is the costs of trading goods. Four main sets of decisions concerning the technology for trading goods can be delineated.

**(A) Variable versus fixed trade costs:** A widespread assumption used for analytical tractability is iceberg variable transport costs, whereby  $d_{ni} > 1$  units of a good must be shipped from location  $i$  to location  $n \neq i$  in order for one unit to arrive (so that some of each unit "melts" in transit).<sup>4</sup> Combining assumptions about the functional form of trade costs with those for preferences and production technology generates predictions for bilateral trade. A strong empirical feature of bilateral trade data that any plausible quantitative spatial model arguably should explain is the gravity equation, in which bilateral trade increases with exporter and importer size and declines with geographical distance (see for example the

---

<sup>4</sup>The spatial economics literature has devoted less attention to the distinction between fixed versus variable trade costs than for example in the recent literature on heterogeneous firms in international trade (an exception is Baldwin & Okubo 2005).

survey by Head & Mayer 2014).

**(B) Asymmetric versus symmetric transport costs:** Whether transport costs are symmetric or asymmetric (whether  $d_{ni} = d_{in}$ ) has implications both for the characterization of equilibrium and patterns of trade and income (see Waugh 2010, and Allen et al. 2015). While transport costs are necessarily symmetric if they depend solely on geographical distance, departures from symmetry can arise from a variety of geographic and economic factors (e.g. land gradient and trade volumes).

**(C) 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 Geographical Information System (GIS) data and software, an important advance has been 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 “Dijkstra” or “Fast Marching,” as in Allen & Arkolakis (2014), Ahlfeldt et al. (2015), Donaldson (2016), Donaldson & Hornbeck (2015), Desmet et al. (2016), and Nagy (2016).

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

**(4) Technology for Idea Flows:** The second mechanism that explains why the location of agents relative to one another is of relevance in quantitative spatial models is frictions in idea flows. Three main sets of specifications for the technology for idea flows have been considered.

**(A) Knowledge externalities and diffusion:** An externality arises whenever an economic agent takes an action that affects another economic agent and this effect is not internalized when evaluating the cost and benefits of the action. Such externalities for idea flows 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 is when 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 (as in Fujita & Ogawa 1982 or Lucas & Rossi-Hansberg, 2002). This reduced-form specification can be derived from alternative micro-foundations. The standard classification of these microfoundations is due to Marshall (1920) and distinguishes between knowledge spillovers, externalities due to thick labor markets, and backward and forward linkages. More recently, Duranton & Puga (2004) proposed 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 and 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 and Ahlfeldt et al. 2015). A key decision in setting up spatial models is whether these externalities are present only within the spatial unit of analysis (as in Allen & Arkolakis, 2014) or also across them as in Rossi-Hansberg (2005).

**(B) Innovation:** A second choice relates to 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 them and hence the speed with which these ideas diffuse to other agents. Whereas most research in economic geography is static and concerned with the spatial distribution of economic activity at a point in time, 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. But tractable quantitative models of spatial innovation, and the corresponding evolution of economic activity, have been recently developed in Desmet and 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 their tractability is that a competitive market for land, together with local diffusion of technology, imply that future returns from an innovation are fully capitalized in land rents.

**(C) Transferability of ideas:** A third choice 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 of transferring that blueprint to other locations, as in Fajgelbaum et al. (2015).

**(5) Technology for the movement of people:** The third mechanism that explains the importance of the spatial location of agents relative to one another in quantitative spatial models is frictions in the movement of people. Here four main sets of assumptions can be distinguished.

**(A) Migration costs:** A first choice relates to frictions for the migration of people. Such frictions provide an alternative explanation for real wage differences across locations to the idiosyncratic differences in preferences discussed above. This raises the question of the extent to which observed urban-rural wage differentials within countries reflect migration frictions, non-random selection of worker productivity, the cost of land and other non-traded goods, and amenity differences (see for example Bryan & Morten, 2015 and Young, 2013). 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 in Artuc et al. (2010), Caliendo et al. (2015) and Morten & Oliviera (2015). Desmet et al. (2016) measures the

origin and destination moving costs that rationalize the observed net population flows across regions in the world using a dynamic spatial model.

**(B) Commuting:** A second choice 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 and Lucas 2000). A key contribution of more recent research has been to allow for non-monocentric patterns of economic activity within cities, in which 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, and Brinkman 2016). In models of systems of cities, the efficiency of the commuting technologies within each city is an important determinant alongside amenities and productivities in shaping the distribution of city sizes (see Desmet & Rossi-Hansberg 2013, and Behrens et al. 2014). In the local labor markets literature, a large literature has examined the impact of local shocks and policy interventions on local employment (see Kline & Moretti 2014a, Moretti 2011b and Neumark & Simpson 2014 for reviews). But relatively little attention has been devoted to commuting and the resulting distinction between employment by workplace and 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).

**(C) Skills/heterogeneity:** For both migration and commuting decisions, a third modelling decision 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, and 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 for example Forthringham & O’Kelly 1989). Whether agents are assumed to have common or idiosyncratic preferences/productivities, an additional 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 that 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 (as in Davis & Dingel 2015, Gaubert 2015, and Redding & Sturm 2016).

**(D) 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 the “fundamental law of highway congestion” (suggested by Downs 1962), according to which increased provisions of

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

**(6) Endowments:** The above choices about preferences, production technology, and the technologies for the movement of goods, ideas and people need to be combined with choices on the endowments of the economy.

**(A) Population and skills:** A minimal endowment is homogenous 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.

**(B) 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 (as in Rossi-Hansberg 2005 or 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, group of countries, or the global economy as a whole? Is space ordered along one dimension (say latitude) or two-dimensional? Data are typically available for discrete spatial units. How disaggregated are these units: points on a latitude and longitude grid, city blocks, municipalities, counties, commuting zones, metropolitan areas, states/provinces, regions or countries as a whole? Clearly these two decisions are interrelated as the choice of spatial scope may limit the level of spatial disaggregation of the units for which data are available.

**(C) Capital and infrastructure:** Other mobile factors of production can be introduced, such as physical capital that is used in a construction sector (as in Combes et al. 2014, Epple et al. 2010 and Ahlfeldt et al. 2015). Incorporating physical production capital that fully depreciates every period is also simple as in Desmet & Rossi-Hansberg (2013). However, incorporating local capital investments over time 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.<sup>5</sup> More generally, depending on the assumptions made about the technologies for 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.

---

<sup>5</sup>The closest frameworks are the models following Desmet & Rossi-Hansberg (2014) that model the local accumulation of technology over time. Still, 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).

**(7) Equilibrium:** Given the above assumptions about preferences, production technology, endowments, and the technologies for the movement of goods, ideas and people, a final remaining set of choices concerns the equilibrium conditions of the model.

**(A) 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 Armington (1969) and Eaton & Kortum (2002), typically assume perfect competition (see for example Allen & Arkolakis 2014 and 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 for example Helpman 1998, Redding 2016 and Monte et al. 2015). One reason is that internal increasing returns to scale requires the assumption of imperfect competition, as otherwise by 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.

**(B) 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 face the choice of 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 that is 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 non-traded, some of the equilibrium conditions will hold within each country separately.

**(C) Land ownership and the distribution of rents:** If land is used for either residential or production purposes it will generate rents to its owners. Hence 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 that 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 is particularly important, since many changes in policy, or productivity effects of innovation, will be ultimately 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 entails the difficulty of 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 it makes agents heterogenous as a function of their

location history. The literature has devised three main ways of incorporating land rents in the analysis, and hence 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 (as in Desmet & Rossi-Hansberg 2014). The second is to distribute land rents locally to current residents (as in Redding 2016). This option generates inefficiencies since moving across locations imposes an externality on the rents received by other agents. One can also combine both to account for trade deficits as in Caliendo et al. (2014). Finally, one can assume the presence of local immobile landlords that consume all their land income locally. Then, local consumption will simply add up to total labor income (as in Monte et al. 2015).

**(D) Trade balance:** In any spatial model one has to take a stand on the spatial unit for which trade is balanced. In quantitative trade models many times this is a country, although clearly country trade accounts exhibit long-lasting and persistent trade imbalances (see for example Reyes-Heroles 2016). The decision is even more relevant when one focuses on smaller spatial units like state, commuting zones, counties, or even zip codes or census tracts. The narrower the spatial unit, the less likely it is that trade needs to be balanced for each location since 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. Still, 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 (as in Allen & Arkolakis 2014). Yet another possibility is to model them and 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 (as in 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.

## 2.1 Criteria for Menu Choice

Having outlined a menu of modules or building blocks for quantitative spatial models, we now discuss some of the possible criteria for choosing from this menu.

**(A) Tractability:** This first criterion includes both analytical and computational tractability. Traditionally theoretical models of economic geography 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 and 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 and 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 non-linear equations for large numbers of locations over realistic computational time periods.<sup>6</sup>

**(B) 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 has to take a stand on what components of the model are invariant to this intervention. Are productivity and amenities exogenous? Or are there agglomeration externalities for productivity and amenities, such that only a component 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 expenditure on world markets? Answers to questions such as these will influence the selection of building blocks from the menu. Only when the assumed structural parameters and locational fundamentals are indeed constant will the analysis not be subject to the Lucas Critique.

**(C) 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 is 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 to the movement 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, changes in wages). Sometimes different types of data can be substitutes for one another. For example, quantitative models typically can 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). Does 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?

---

<sup>6</sup>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 129,600 locations in the whole world.

### 3 A Quantitative Spatial Model

In this section, we outline a canonical quantitative spatial model that corresponds to a multi-region 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) Preferences:** (A) Love of variety; (B) Single traded sector; (C) No amenities; (D) Residential land use; (E) Common preferences; **(2) Production Technology:** (A) Increasing returns to scale; (B) Exogenous productivity; (C) No input-output linkages; (D) No commercial land use; **(3) Technology for Trading Goods:** (A) Iceberg variable trade costs; (B) Symmetric trade costs; (C) Economic and Geographic Frictions; (D) No non-traded goods besides residential land use; **(4) Technology for the Movement of Ideas:** (A) No knowledge externalities or diffusion; (B) No innovation; (C) No transferability of ideas; **(5) Technology for the Movement of People:** (A) Perfectly costless migration; (B) No commuting; (C) Single worker type with no heterogeneity; (D) No congestion in transportation; **(6) Endowments:** (A) Homogenous labor; (B) Exogenous land endowments in regions within a single country; (C) No capital; **(7) Equilibrium:** (A) Monopolistic competition; (B) General equilibrium with a single country; (C) Land rents redistributed to residents; (D) Trade is balanced in each location. This model has been widely used in empirical work, including 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 hence 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  $d_{nn} = 1$ .<sup>7</sup>

#### 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 as in<sup>8</sup>

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

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 with dual price index ( $P_n$ ) given

<sup>7</sup>Although we interpret the locations here as regions, another interpretation could be as cities, as in the literature on systems of cities following Henderson (1974).

<sup>8</sup>For empirical evidence using U.S. data in support of the constant housing expenditure share implied by the Cobb-Douglas functional form, see Davis & Ortalo-Magne (2011).

by

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

### 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

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

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

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

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

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

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

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

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:

$$M_i = \frac{L_i}{\sigma F}. \quad (7)$$

### 3.3 Price Indices and Expenditure Shares

Using equilibrium prices (4) and labor market clearing (7), the price index dual to the consumption index (2) can be expressed as:

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

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

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

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 Anderson & Wincoop (2003). Together (8) and (9) imply that each location’s price index can be again written in terms of its trade share with itself, so

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

### 3.4 Income and Population Mobility

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

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

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

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

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

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

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

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

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,

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

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 (9), price index (8), and population mobility condition (13). Under the assumption that trade costs are symmetric ( $d_{ni} = d_{in}$ ), one can follow the arguments in 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:

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

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

$$\begin{aligned} \tilde{\sigma} &\equiv \frac{\sigma-1}{2\sigma-1}, & \gamma_1 &\equiv \frac{\sigma(1-\alpha)}{\alpha}, \\ \gamma_2 &\equiv 1 + \frac{\sigma}{\sigma-1} - \frac{(\sigma-1)(1-\alpha)}{\alpha}. \end{aligned}$$

Wages in turn are implicitly determined by

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

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 (16) in Fujimoto & Krause (1985), to show that there exists a unique vector  $L_n$  that satisfies (16) as long as  $\gamma_2/\gamma_1 \in (0, 1]$ . Hence, 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  $\gamma_2/\gamma_1 \in (0, 1)$  one can also guarantee that a solution to (16) can be found by iteration from any initial distribution of populations.<sup>9</sup>

The parameter restrictions to 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, this expands the measure of varieties produced there, 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 the agglomeration force. The higher the share of land ( $1-\alpha$ ), the stronger the dispersion force. For parameter values for which  $\sigma(1-\alpha) > 1$ , the dispersion force dominates the agglomeration force, and there exists a unique equilibrium distribution of economic activity.

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. While this is a convenient property of the model

<sup>9</sup>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 in Zabreyko et al. (1975). Hence if  $\gamma_2/\gamma_1 \in (0, 1)$  a solution exists, is unique, and can be found by iteration.

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 1991a), and assuming  $\sigma(1 - \alpha) > 1$  excludes this possibility. Hence, for the range of parameters where  $\sigma(1 - \alpha) > 1$  the model cannot generate agglomerations when space is perfectly homogenous, it can only generate agglomeration 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 in 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 determined 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 have focused 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. Arkolakis & Allen (2014) demonstrates 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. Arkolakis & Allen (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

We now 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 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 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 use again the mathematical theorems that guarantee solutions to these type of equations (for example 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  $A_i$  inside the sum is smaller than the one 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 (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).

An implication of these arguments is 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, since 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 (and hence have no standard errors). Note also that since  $H_n$  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 a 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 (such as 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 below that have sought to provide such overidentification checks and/or undertake such structural estimation to provide evidence in support of this class of quantitative spatial models.

### 3.7 Counterfactuals

We now 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 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 (9), price index (10), income equals expenditure (11), land market clearing (12), and population mobility (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 ( $\{w, \pi_{ni}, \lambda_n\}$ ):

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

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

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

Under 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 (14), the class of models considered here has the property that the change 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:

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

Under 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 hence reduces its domestic trade share), 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 analogue of the result in 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 also need to be taken into account (see Caliendo et al. 2014 and Redding 2016).

### 3.9 Quantitative Illustration

We close this section with 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 of which occupies the Western half of the grid (West), and another which takes up the Eastern half of the grid (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 the lowest-cost-route effective distance, which minimizes the cumulative cost of travelling between each origin and destination. In particular, this measure assigns a cost (or weight) to travelling across each point  $n$  on the grid of  $\delta_n$ . For a pair of neighboring points  $n$  and  $i$ , the effective distance for a orthogonal link between these locations (moving vertically or horizontally across the grid) is  $dist_{ni} = (\delta_n + \delta_i) / 2$ . In contrast, the effective distance for a diagonal link between these locations (moving diagonally across the grid) is  $dist_{ni} = \left( \left( 2(\delta_n + \delta_i)^2 \right)^{0.5} \right) / 2$ . 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 Figure 1, and all other figures, blue

(cold) colors correspond to lower values and yellow (hot) colors correspond to higher values. In this realization there are two clusters of high productivity areas, one North-West of the border and another one South-East 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 kilometers squared. 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 percent, which is in line with the housing expenditure share in 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  $\sigma - 1 = 4$  that is line with the estimates in Simonovska & Waugh (2014). Third, we assume that trade costs are a constant elasticity function of effective distance ( $d_{ni} = dist_{ni}^\phi$ ), which implies an elasticity of trade flows with respect to effective distance of  $(\sigma - 1)\phi$  (since trade flows depend on  $d_{ni}^{-(\sigma-1)} = dist_{ni}^{-(\sigma-1)\phi}$ ). We choose the parameter  $\phi$  to match the elasticity of trade flows with respect to distance in gravity equations using inter-regional 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 percent ( $\tau^{\text{in}} = 2$ ), which is paid whenever a good flows from one location to another. Second, we assume a proportional external tax on trade between the two countries of 100 percent ( $\tau^{\text{out}} = 2$ ), 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 hence 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. Panels A, B, C, show that areas of high productivity have large population concentrations, high wages, and high land prices respectively. Panel D shows the log of the Price Index. As expected, this is a smooth surface with 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 exhibits clearly the border effect created by the tariff between both 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 towards the border between them. The areas that benefit the most are the ones close, but on the opposite side of the border, of the large cities. These locations can now trade more cheaply with the large market in those cities and hence experience the largest increases in population (Panel A), wages (Panel B) and land rents (Panel C) and the largest reductions in the price index (Panel D). 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 percent 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 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 (Panel 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 (Panel D). Wages and land rents also fall in large agglomerations, while they increase in all other regions (Panel B and 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 3.4 percent respectively, more than five times the effects of the external trade liberalization. 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 since 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. While the particular quantitative model considered here 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) Preferences:** (A) Homogeneous good; (B) Single traded sector; (C) Endogenous amenities; (D) Residential land use; (E) Idiosyncratic preferences; **(2) Production Technology:** (A) Constant returns to scale; (B) Endogenous productivity; (C) No input-output linkages; (D) Commercial land use; **(3) Technology for Trading Goods:** (A)-(C) No trade costs; (D) No non-traded goods besides residential and commercial land use; **(4) Technology for the Movement of Ideas:** (A) Knowledge externalities; (B) No innovation or dynamics; (C) Perfect transferability of ideas; **(5) Technology for the Movement of People:** (A) Perfectly costless migration; (B) Costly commuting; (C) Single worker type with heterogeneity; (D) No congestion in transportation; **(6) Endowments:** (A) Homogenous labor; (B) Blocks within a single city; (C) No capital ; **(7) Equilibrium:**

(A) Perfect competition; (B) Equilibrium within a city given prices and utility in an outside economy; (C) Absentee landlords; (D) Trade is 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  $n, i = \{1, \dots, S\}$ . Each block has a supply of floor space ( $H_n$ ) that depends on geographical land area ( $K_n$ ) and the density of development ( $\varphi_n$ ). There is a single final good which is costlessly traded within the city and to the larger economy and is chosen as the numeraire ( $p_i = 1$  for all  $i$ ).<sup>10</sup> 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,

$$1 = \frac{1}{A_i} w_i^\alpha q_i^{1-\alpha}, \quad 0 < \alpha < 1, \quad (22)$$

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

$$A_j = a_j Y_j^\mu, \quad Y_j \equiv \sum_{s=1}^S e^{-\delta \tau_{js}} \left( \frac{L_{Ms}}{K_s} \right), \quad (23)$$

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

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

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  $i$ .

<sup>10</sup>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.

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

$$B_i = b_i \Omega_i^\eta, \quad \Omega_i \equiv \sum_{r=1}^S e^{-\rho \tau_{ir}} \left( \frac{L_{Rr}}{K_r} \right), \quad (25)$$

where  $L_{Rr}/K_r$  is residence employment density per unit of land area; residential externalities decline with travel time ( $\tau_{ir}$ ) through the iceberg factor  $e^{-\rho \tau_{ir}} \in (0, 1]$ ;  $\rho$  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,

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

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, 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,

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

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

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

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

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

where  $\lambda_{ni|n}$  is the probability of commuting to location  $i$  conditional on residing in location  $n$ ; labor market clearing implies  $L_{Mi} = \sum_{n=1}^S \lambda_{ni} L$  and  $L_{Ri} = \sum_{n=1}^S \lambda_{ni} L$ , 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  $i$ :

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

No-arbitrage across alternative land uses implies that floor space is either 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 ( $Q_i$ ) 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 ( $Q_i = q_i = \xi_i Q_i$ ). Similarly, for blocks that are completely specialized in commercial activity, observed floor prices again equal commercial floor prices ( $Q_i = q_i$ ). In contrast, for blocks that are completely specialized in residential activity, observed floor prices equal residential floor prices ( $Q_i = Q_i$ ).

As for the quantitative spatial model in Section 3, analytical results can be provided for 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 hence the congestion forces of commuting costs and an inelastic supply of land ensure the existence of a unique equilibrium, as shown in 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 in 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 in Section 3, the model 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, such as using the initial values from the observed equilibrium 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) develops 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

A first line of empirical research has examined a key implication of quantitative spatial models that both wages and population depend on market access.<sup>11</sup> We illustrate this prediction in the context of the quantitative spatial model developed in Section 3 above. Using CES demand, profit maximization (4) and zero profits (6), the free-on-board price ( $p_i$ ) charged for each variety by a firm in each location  $i$  must

---

<sup>11</sup>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 Davis and Weinstein (1999, 2003) and Costinot et al. (2016).

be low enough in order to sell the quantity  $\bar{x}_i$  and cover the firm's fixed production costs, so

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

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

$$w_i = \xi A_i^{\frac{\sigma-1}{\sigma}} (FMA_i)^{\frac{1}{\sigma}}, \quad FMA_i \equiv \sum_{n \in N} (w_n L_n) (P_n)^{\sigma-1} (d_{ni})^{1-\sigma}, \quad (32)$$

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 (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 ( $CMA_n$ ):

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

Using data on a cross-section of countries, Redding & Venables (2004) finds 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 Barthelme (2016) all report instrumental variables 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 liberalizations as a natural experiment that changes the relative market access of regions (as in Hanson 1996, 1997).<sup>12</sup> 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. Here 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 locations.<sup>13</sup> 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. It finds that

<sup>12</sup>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.

<sup>13</sup>See the review in Redding & Turner (2014).

there is a strong and statistically significant estimated effect of railroads on real income levels, but 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 model. In a similar vein, Donaldson & Hornbeck (2016) investigates the impact of the expansion of the railroad network on the agricultural sector and shows 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) uses the division of Germany after the Second World War 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. 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 the disease environment. Third, the change in market access following German division is much larger than 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 pre-division characteristics of cities.<sup>14</sup>

In the quantitative model from Section 3, the treatment effect of division on border cities is shaped by 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) undertakes 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 empirical literature finds that wages, land prices, productivity, employment and employment growth are positively correlated with population density.<sup>15</sup> In their survey of this empirical literature, Rosenthal & Strange (2004) reports that the elasticity of productivity with respect to the density of economic activity is typically estimated to lie within the range of 3-8 percent.<sup>16</sup> However, establishing that

---

<sup>14</sup>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 Zonenrandgebiet (ZRG) place-based policy, Ehrlich & Seidel (2015) finds even larger effects of 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, Brühlhart, et al. (2012) finds substantial increases in both wages and employment for Austrian municipalities within 50 kilometers of the former Iron Curtain. Using the economic separation of Japan and Korea after the Second World War and implementing the same empirical specification as in Redding & Sturm (2008), Kentaro (2008) finds a similar pattern of market access effects.

<sup>15</sup>See the survey in Moretti (2011b).

<sup>16</sup>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 Rosenthal & Strange (2004).

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) provides evidence on agglomeration spillovers by comparing changes in total factor productivity (TFP) 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) estimates substantial and highly localized housing externalities. Land prices in neighborhoods targeted for revitalization rose by 2-5 percent at an annual rate about those in a control neighborhood. With every 1,000 feet of distance, housing externalities are estimated to decline by around one half.<sup>17</sup>

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

Using the division of Berlin following the Second World War 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 estimates the parameters of the quantitative spatial model in Section 4. Following the city’s division, there is 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, while the city’s reunification leads 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 and residence employment are  $\mu = 0.07$  and  $\eta = 0.15$  respectively. Undertaking counterfactuals for the impact of division and reunification, the special case of the model without agglomeration forces ( $\mu = \eta = 0$ ) is shown to be unable to account quantitatively for the observed

---

<sup>17</sup>Using 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.

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.<sup>18</sup>

Further quantitative evidence on the role of residential externalities in influencing the spatial distribution of economic activity is provided in Diamond (2015). From 1980 to 2000, the rise in the U.S. college-high school graduate wage gap 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 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 suggested by the increase in the college wage gap alone.

While 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, most of the empirical literature on the evolution of the distribution of economic set out to look for evidence of path dependence. Namely, evidence that temporary shocks can affect the distribution of economic activity permanently. Davis & Weinstein (2002, 2008) used Japanese war-time bombing as such an exogenous temporary shock, and found little evidence of path dependence for either the distribution of population as a whole or employment in individual industries.<sup>19</sup> Subsequent studies have provided a number of apparent examples of path dependence using a variety of alternative sources of exogenous variation. Redding et al. (2010) finds path dependence for the location of Germany's air hub using the natural experiment of Germany's division and reunification.<sup>20</sup> Bleakley & Lin (2012) finds permanent effects of a temporary historical advantage on the spatial distribution of population using variation from portage sites in the United States. Hornbeck & Keniston (2014) finds long-lived effects of the Boston fire through the potential for large-scale rebuilding using plot-level data on land values. Michaels & Rauch (2016) reports 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, infrastructure, and better institutions, dynamic agglomeration effects can lead to main-

---

<sup>18</sup>For a quantitative analysis of the role of urban property rights in influencing internal city structure in a developing-country context, see Bernard et al. (2016).

<sup>19</sup>Other research exploiting war-time bombing as an exogenous shock includes Bosker et al. (2007, 2008) and Miguel & Roland (2011).

<sup>20</sup>For structural estimations of the determinants of the location of particular economic activities, see Holmes (2005) for head-quarter location choices and Holmes (2011) for the expansion of Walmart's distribution and retail network.

tained 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 in Desmet & Rossi-Hansberg (2015) and Desmet et al. (2016). The latter paper 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, the paper finds a correlation between predicted country population levels in 1950 and 2000 as high as 0.69. In these models, the equilibrium is unique, and so the evolution of the distribution of economic activity 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.<sup>21</sup>

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

## 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 non-tradable. 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

---

<sup>21</sup>See Desmet & Henderson (2015) for a review of the theory and empirics of the evolution of the distribution of economic activity within countries.

economic analysis of the geography of firm and worker networks remains under-explored. We expect much progress along these and other dimensions over the coming decades.

## References

- [1] Ahlfeldt GM, Redding SJ, Sturm DM, Wolf N. 2015. The Economics of Density: Evidence from the Berlin Wall. *Econometrica* 83(6): 2127-2189.
- [2] Allen T, Arkolakis C. 2014. Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics* 129(3): 1085-1140.
- [3] Allen T, Arkolakis C, Takahashi Y. 2015. *Universal Gravity*. NBER Working Paper 20787.
- [4] Allen T, Arkolakis C, Li X. 2016. *Optimal City Structure*. Yale University, mimeograph.
- [5] Alonso W. 1964. *Location and Land Use*. Cambridge MA: Harvard.
- [6] Anderson ML. 2014. Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion. *American Economic Review* 104(9): 2763-2796.
- [7] Anderson JE, van Wincoop E. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* 93(1): 170-192.
- [8] Arkolakis C, Costinot A, Rodriguez-Clare A. 2012. New Trade Models, Same Old Gains. *American Economic Review* 102(1): 94-130.
- [9] Arkolakis C, Ramondo N, Rodriguez-Clare A, Yeaple S. 2014. *Innovation and Production in the Global Economy*. NBER Working Paper 18792.
- [10] Armington PS. 1969. A Theory of Demand for Products Distinguished by Place of Production. *IMF Staff Papers* 16(1): 159-178.
- [11] Artuç ES, Chaudhuri S, McLaren J. 2010. Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review*. 100(3): 1008-1045.
- [12] Arzaghi M., Henderson JV. 2008. Networking Off Madison Avenue. *Review of Economic Studies*: 75(4): 1011-1038.
- [13] Autor DH, Palmer CJ, Pathak PA. 2014. Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts. *Journal of Political Economy* 122(3): 661-717.
- [14] Baldwin R, Forslid R, Martin P, Ottaviano GIP, Robert-Nicoud F. 2003. *Economic Geography and Public Policy*. Princeton: Princeton University Press.

- [15] Baldwin R, Okubo T. 2005. Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting. *Journal of Economic Geography*. 6(3): 323-346.
- [16] Bartelme D. 2016. *Trade Costs and Economic Geography: Evidence from the US*. University of Michigan, mimeograph.
- [17] Behrens K, Mion G, Murata Y, Südekum J. 2014. *Spatial Frictions*. DICE Discussion Papers 160, Heinrich-Heine-Universität Düsseldorf, Düsseldorf Institute for Competition Economics (DICE).
- [18] Bernard L, Bird J, Venables AJ. 2016. *The Urban Land Market: A Computable Equilibrium Model Applied to Kampala City*. University of Oxford, mimeograph.
- [19] Bleakley H, Lin J. 2012. Portage and Path Dependence. *Quarterly Journal of Economics*. 127: 587-644.
- [20] Bosker M, Brakman S, Garretsen H, Schramm M. 2007. Looking for Multiple Equilibria when Geography Matters: German City Growth and the WWII Shock. *Journal of Urban Economics* 61(1): 152-169.
- [21] Bosker M, Brakman S, Garretsen H, Schramm M. 2008. A Century of Shocks: The Evolution of the German City Size Distribution 1925-1999. *Regional Science and Urban Economics* 38(4): 330-347.
- [22] Brinkman J. 2016. Congestion, Agglomeration, and the Structure of Cities. *Journal of Urban Economics* 94: 13-31.
- [23] Brülhart M, Carrère C, Trionfetti F. 2012. How Wages and Employment Adjust to Trade Liberalization: Quasi-Experimental Evidence from Austria. *Journal of International Economics* 86: 68-81.
- [24] Bryan G, Morten M. 2015. *Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia*. Stanford University, mimeograph.
- [25] Busso M, Gregory J, Kline P. 2013. Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review* 103(2): 897-947.
- [26] Caliendo L, Parro F, Rossi-Hansberg E, Sarte PD. 2014. *The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy*. NBER Working Paper 20168.
- [27] Caliendo L, Dvorkin M, Parro F. 2015. *Trade and Labor Market Dynamics*. NBER Working Paper 21149.
- [28] Combes P, Duranton G, Gobillon L. 2016. *The Costs of Agglomeration: Land Prices in French Cities*. Wharton, mimeograph.
- [29] Combes P, Duranton G, Gobillon L, Roux S. 2010. Estimating Agglomeration Economies With History, Geology, and Worker Effects. In *Agglomeration Economies*, ed. E. L. Glaeser. Chicago: Chicago University Press.

- [30] Comin D, Dmitriev M, Rossi-Hansberg E. 2013. *The Spatial Diffusion of Technology*. Princeton University, mimeograph.
- [31] Coşar AK, Fajgelbaum, PD. 2016. Internal Geography, International Trade, and Regional Specialization. *American Economic Journal: Microeconomics*. 8(1): 24-56.
- [32] 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*. Stanford University, mimeograph.
- [33] Davis DR, Dingel JI. 2015. *The Comparative Advantage of Cities*. Columbia University, mimeograph.
- [34] Davis DR, Weinstein DE. 1999. Economic Geography and Regional Production Structure: An Empirical Investigation. *European Economic Review* 43: 379-407.
- [35] Davis DR, Weinstein DE. 2002. Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review* 92(5): 1269-1289.
- [36] Davis DR, Weinstein DE. 2003. Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment. *Journal of International Economics* 59(1): 1-23.
- [37] Davis DR, Weinstein DE. 2008. A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science* 48(1): 29-65.
- [38] Davis MA, Ortalo-Magné F. 2011. Housing Expenditures, Wages, Rents. *Review of Economic Dynamics* 14(2): 248-261.
- [39] Dekle R, Eaton J, Kortum S. 2007. Unbalanced Trade. *American Economic Review*, 97(2), 351-355.
- [40] Desmet K, Henderson JV. 2015. The Geography of Development Within Countries. In *Handbook of Regional and Urban Economics*, ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, Amsterdam: Elsevier, 1457-1517.
- [41] Desmet K, Rossi-Hansberg E. 2013. Urban Accounting and Welfare. *American Economic Review* 103(6): 2296-2327.
- [42] Desmet K, Rossi-Hansberg E. 2014. Spatial Development. *American Economic Review* 104(4): 1211-1243.
- [43] Desmet K, Rossi-Hansberg E. 2015. On the Spatial Economic Impact of Global Warming. *Journal of Urban Economics* 88: 16-37.
- [44] Desmet K, Nagy DK, Rossi-Hansberg, E. 2016. *The Geography of Development*. Princeton University, mimeograph.
- [45] Diamond R. 2016. The Determinants and Welfare Implications of US Workers? Diverging Location Choices by Skill: 1980-2000. *American Economic Review* 106(3): 479-524.

- [46] Donaldson D. 2016. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure? *American Economic Review*, forthcoming.
- [47] Donaldson D, Hornbeck R. 2016. Railroads and American Economic Growth: A 'Market Access' Approach. *Quarterly Journal of Economics*, forthcoming.
- [48] Downs A. 1962. The Law of Peak-Hour Expressway Congestion. *Traffic Quarterly*. 16(3): 393-409.
- [49] Duranton G, Puga D. 2004. Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics*, ed. J. V. Henderson and J.-F. Thisse, Amsterdam: Elsevier. Vol. 4: 2063-2117.
- [50] Duranton G, Turner MA. 2011. The Fundamental Law of Road Congestion: Evidence from U.S. Cities. *American Economic Review* 101(6): 2616-2652.
- [51] Eaton B, Kortum S. 2002. Technology, Geography, and Trade. *Econometrica* 70(5): 1741-1779.
- [52] Ehrlich M, Seidel T. 2015. *The Persistent Effects of Place-Based Policy: Evidence from the West German Zonenrangeland*. University of Bern, mimeograph.
- [53] Ellison G, Glaeser EL, Kerr WR. 2010. What Causes Industry Agglomeration? Evidence from Co-agglomeration Patterns. *American Economic Review* 100(3): 1195-1213.
- [54] Epple D, Gordon B, Sieg H. 2010. A New Approach to Estimating the Production Function for Housing. *American Economic Review* 100(3): 905-924.
- [55] Fajgelbaum PD, Morales E, Suárez Serrato JC, Zidar O. 2015. *State Taxes and Spatial Misallocation*. NBER Working Paper 21760.
- [56] Fajgelbaum PD, Redding SJ. 2014. *External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914*. NBER Working Paper 20217.
- [57] Fortheringham S, O'Kelly M. 1989. *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer.
- [58] Fujimoto T, Krause U. 1985. Strong Ergodicity for Strictly Increasing Nonlinear Operators. *Linear Algebra and its Applications* 71(1): 101-112.
- [59] Fujita M, Ogawa H. 1982. Multiple Equilibria and Structural Transformation of Non-Monocentric Urban Configurations. *Regional Science and Urban Economics* 12(2): 161-196.
- [60] Fujita M, Thisse JF. 2002. *The Economics of Agglomeration*. Cambridge: Cambridge University Press.
- [61] Fujita M, Krugman P, Venables AJ. 1999. *The Spatial Economy: Cities, Regions and International Trade*. Cambridge: MIT Press.

- [62] Galle S, Rodriguez-Clare A, Yi M. 2015. *Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade*. University of California, Berkeley, mimeograph.
- [63] Gaubert C. 2015. *Firm Sorting and Agglomeration*. University of California at Berkeley, mimeograph.
- [64] Greenstone M, Hornbeck R, Moretti E. 2010. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy* 118(3): 536-598.
- [65] Grogger J, Hanson G. 2011. Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics* 95(1): 42-57.
- [66] Hanson GH. 1996. Localization Economies, Vertical Organization, and Trade. *American Economic Review* 86(5): 1266-1278.
- [67] Hanson GH. 1997. Increasing Returns, Trade, and the Regional Structure of Wages. *Economic Journal* 107: 113-133.
- [68] Hanson GH. 2005. Market Potential, Increasing Returns, and Geographic Concentration. *Journal of International Economics* 67(1): 1-24.
- [69] Head K, Mayer T. 2014. Gravity Equations: Workhorse, Toolkit, and Cookbook. In *Handbook of International Economics*. ed. G Gopinath, E Helpman, K. Rogoff. Amsterdam: Elsevier. Vol. 4: 131-195.
- [70] Helpman E. 1998. The Size of Regions. In *Topics in Public Economics: Theoretical and Applied Analysis*. ed. D Pines, E Sadka, I Zilcha, Cambridge: Cambridge University Press.
- [71] Henderson JV. 1974. The Sizes and Types of Cities. *American Economic Review* 64(4): 640-56.
- [72] Holmes TJ. 2005. The Location of Sales Offices and the Attraction of Cities. *Journal of Political Economy* 113(3): 551-581.
- [73] Holmes TJ. 2011. The Diffusion of Walmart and the Economies of Density. *Econometrica* 79(1): 253-302.
- [74] Holmes TJ, Sieg H. 2015. Structural Estimation in Urban Economics. In *Handbook of Regional and Urban Economics*. ed. G Duranton, JV Henderson, WC Strange. Amsterdam: Elsevier. Vol. 5. 69-114.
- [75] Hornbeck R, Keniston D. 2014. *Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872*. Chicago Booth School of Business, mimeograph.
- [76] Jaffe AB, Trajtenberg M, Henderson R. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108(3): 577-598.

- [77] Kennan J, Walker JR. 2011. The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, 79(1): 211-251.
- [78] Kentaro N. 2008. Economic Division and Spatial Relocation: The Case of Postwar Japan. *Journal of the Japanese and International Economies* 22(3): 383-400.
- [79] Kline P, Moretti E. 2014a. People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Policies. *Annual Review of Economics* 6: 629-662.
- [80] Kline P, Moretti E. 2014b. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence From the Tennessee Valley Authority. *Quarterly Journal of Economics* 129(1): 275-331.
- [81] Krugman PR. 1991a. Increasing Returns and Economic Geography. *Journal of Political Economy* 99(3): 483-499.
- [82] Krugman PR. 1991b. *Geography and Trade*. Cambridge MA: MIT Press.
- [83] Krugman PR, Venables AJ. 1995. Globalisation and the Inequality of Nations. *Quarterly Journal of Economics*. 60: 857-880.
- [84] Lucas RE. 2000. Externalities and Cities. *Review of Economic Dynamics* 4(2): 245-274.
- [85] Lucas RE, Rossi-Hansberg E. 2002. On the Internal Structure of Cities. *Econometrica* 70(4): 1445-76.
- [86] Marshall A. 1920. *Principles of Economics*. London: Macmillan.
- [87] McFadden D. 1974. The Measurement of Urban Travel Demand. *Journal of Public Economics* 3(4): 303-328.
- [88] Melo PC, Graham DJ, Noland RB. 2009. A Meta-Analysis of Estimates of Urban Agglomeration Economies. *Regional Science and Urban Economics*. 39(3): 332-342.
- [89] Michaels G, Rauch F. 2016. Resetting the Urban Network: 117-2012. *Economic Journal*. In press.
- [90] Miguel E, Roland G. 2011. The Long-run Impact of Bombing Vietnam. *Journal of Development Economics* 96: 1-15.
- [91] Mills ES. 1967. An Aggregative Model of Resource Allocation in a Metropolitan Centre. *American Economic Review* 57(2): 197-210.
- [92] Monte F. 2015. The Local Incidence of Trade Shocks. Georgetown University, mimeograph.
- [93] Monte F, Redding SJ, Rossi-Hansberg E. 2015. *Commuting, Migration and Local Employment Elasticities*. NBER Working Paper 21706.

- [94] Moretti E. 2011a. Local Multipliers. *American Economic Review* 100(2): 1-7.
- [95] Moretti E. 2011b. Local Labor Markets. In *Handbook of Labor Economics*. ed. D. Card, O Ashenfelter. Amsterdam: Elsevier. Vol. 4B: 1238-1303.
- [96] Morten M, Oliveira J. 2016. *Paving the Way to Development: Costly Migration and Labor Market Integration*. Stanford University, mimeograph.
- [97] Muth R. 1969. *Cities and Housing*. Chicago: University of Chicago Press.
- [98] Nagy DK. 2016. City Location and Economic Development. Princeton University, mimeograph.
- [99] Neumark D, Simpson H. 2014. Place-based Policies. In *Handbook of Regional and Urban Economics*, ed. G Duranton, W Strange, JV Henderson, Amsterdam: Elsevier North-Holland. Volume 5.
- [100] 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, Oxford: Basil Blackwell. 353-87.
- [101] Overman HG, Winters LA. 2006. *Trade Shocks and Industrial Location: The Impact of EEC accession on the UK*. CEP Discussion Paper 588. London School of Economics.
- [102] Puga D. 1999. The Rise and Fall of Regional Inequalities. *European Economic Review* 43: 303-334.
- [103] Redding SJ. 2010. The Empirics of New Economic Geography. *Journal of Regional Science*, 50th Anniversary Special Issue, 50(1): 297-311.
- [104] 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. London: Palgrave Macmillan. Chapter 16.
- [105] Redding SJ. 2016. Goods Trade, Factor Mobility and Welfare. *Journal of International Economics*, 101: 148-167.
- [106] Redding SJ, Sturm DM. 2008. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review* 98(5): 1766-1797.
- [107] Redding SJ, Sturm DM. 2016. *Estimating Neighborhood Effects: Evidence from War-time Destruction in London*. Princeton University, mimeograph.
- [108] Redding SJ, Sturm DM, Wolf N. 2011. History and Industrial Location: Evidence from German Airports. *Review of Economics and Statistics* 93(3): 814-831.
- [109] Redding, SJ, Turner M. 2015. Transportation Costs and the Spatial Organization of Economic Activity. In *Handbook of Urban and Regional Economics*, ed. G Duranton, JV Henderson , W Strange. Vol. 5: 1339-1398.

- [110] Redding SJ, Venables A. 2004. Economic Geography and International Inequality. *Journal of International Economics* 62(1): 53-82.
- [111] Reyes-Heroles R. 2016. *The Role of Trade Costs in the Surge of Trade Imbalances*. Princeton University, mimeograph.
- [112] Roback J. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90: 1257-1278.
- [113] Rosen S. 1979. Wages-based Indexes of Urban Quality of Life. In *Current Issues in Urban Economics*, ed. P. Mieszkowski, M. Straszheim. Baltimore: John Hopkins Univ. Press.
- [114] Rosenthal SS, Strange WC. 2004. Evidence on the Nature and Sources of Agglomeration Economics. In *Handbook of Regional and Urban Economics*, ed. JV Henderson, JF Thisse. Amsterdam: Elsevier North Holland. Vol. 4.
- [115] Rosenthal SS, Strange WC. 2008. The Attenuation of Human Capital Spillovers. *Journal of Urban Economics* 64(2): 373-389.
- [116] Rossi-Hansberg E. 2005. A Spatial Theory of Trade. *American Economic Review* 95(5): 1464-1491.
- [117] Rossi-Hansberg E, Sarte P, Owens RI. 2010. Housing Externalities. *Journal of Political Economy* 118(3): 485-535.
- [118] Shoven JB, Whalley J. 2005. *Applying General Equilibrium*. Cambridge: Cambridge University Press.
- [119] Simonovska I, Waugh ME. 2014. The Elasticity of Trade: Estimates and Evidence. *Journal of International Economics* 92(1): 34-50.
- [120] Tirado DA, Paluzie E, Pons J. 2002. Economic Integration and Industrial Location: The Case of Spain before WWI. *Journal of Economic Geography* 2(3): 343-363.
- [121] Waugh M. 2010. International Trade and Income Differences. *American Economic Review* 100(5): 2093-2124.
- [122] Wolf N. 2007. Endowments vs. Market Potential: What Explains the Relocation of Industry after the Polish Reunification in 1918? *Explorations in Economic History* 44(1): 22-42.
- [123] Young A. 2013. Inequality, the Urban-Rural Gap, and Migration. *Quarterly Journal of Economics* 128: 1727-1785.
- [124] Zabreyko P, Koshelev A, Krasnoselskii M, Mikhlin S, Rakovshchik L, Stetisenko V. 1975. *Integral Equations: A Reference Text*. Leyden: Noordhoff International Publishing.

Figure 1: Realization of Random Local Productivity

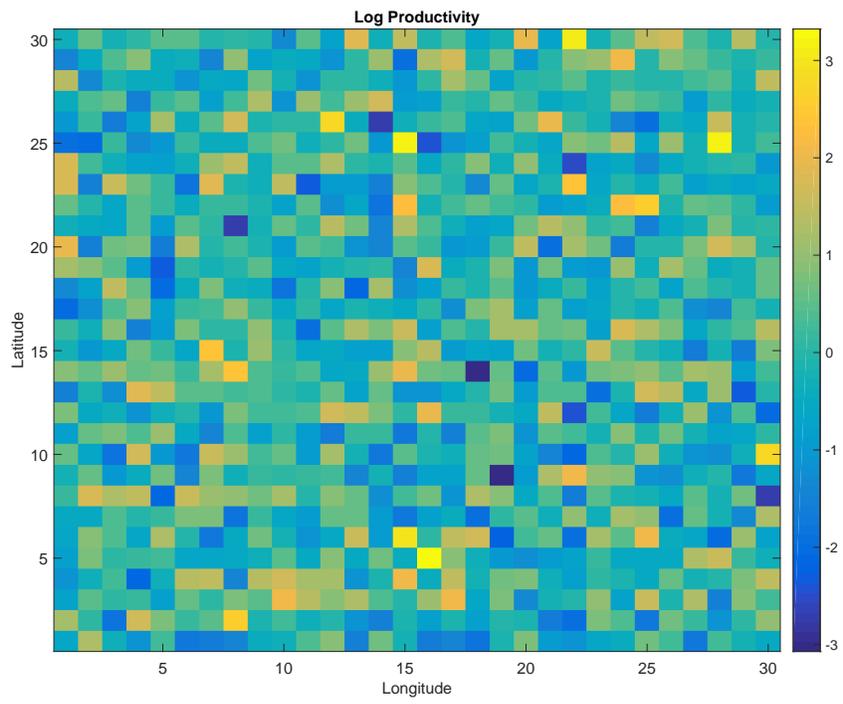


Figure 2: Initial Equilibrium

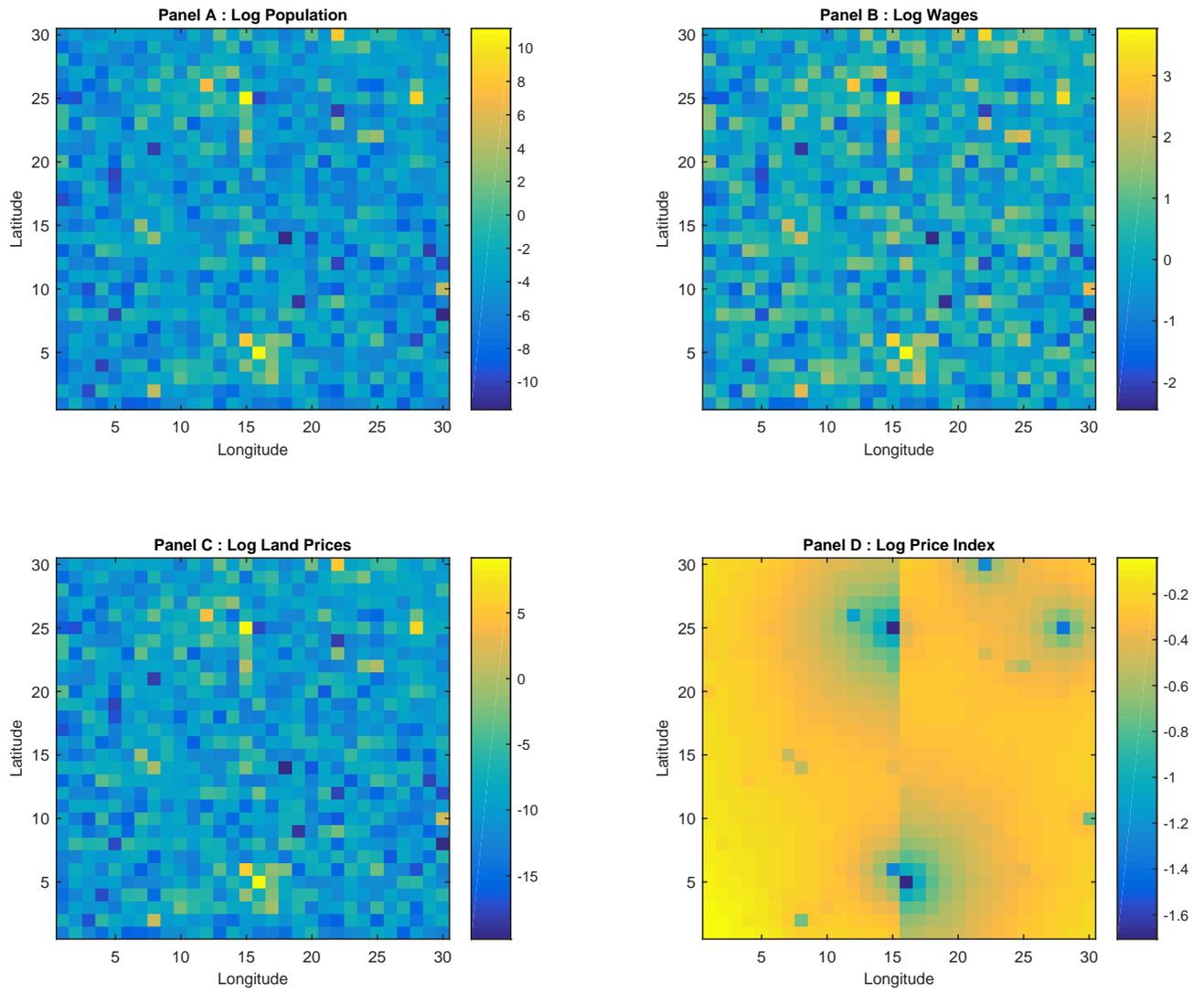


Figure 3: External Liberalization (Ratio of Counterfactual to Initial Values,  $\hat{x} = x'/x$ )

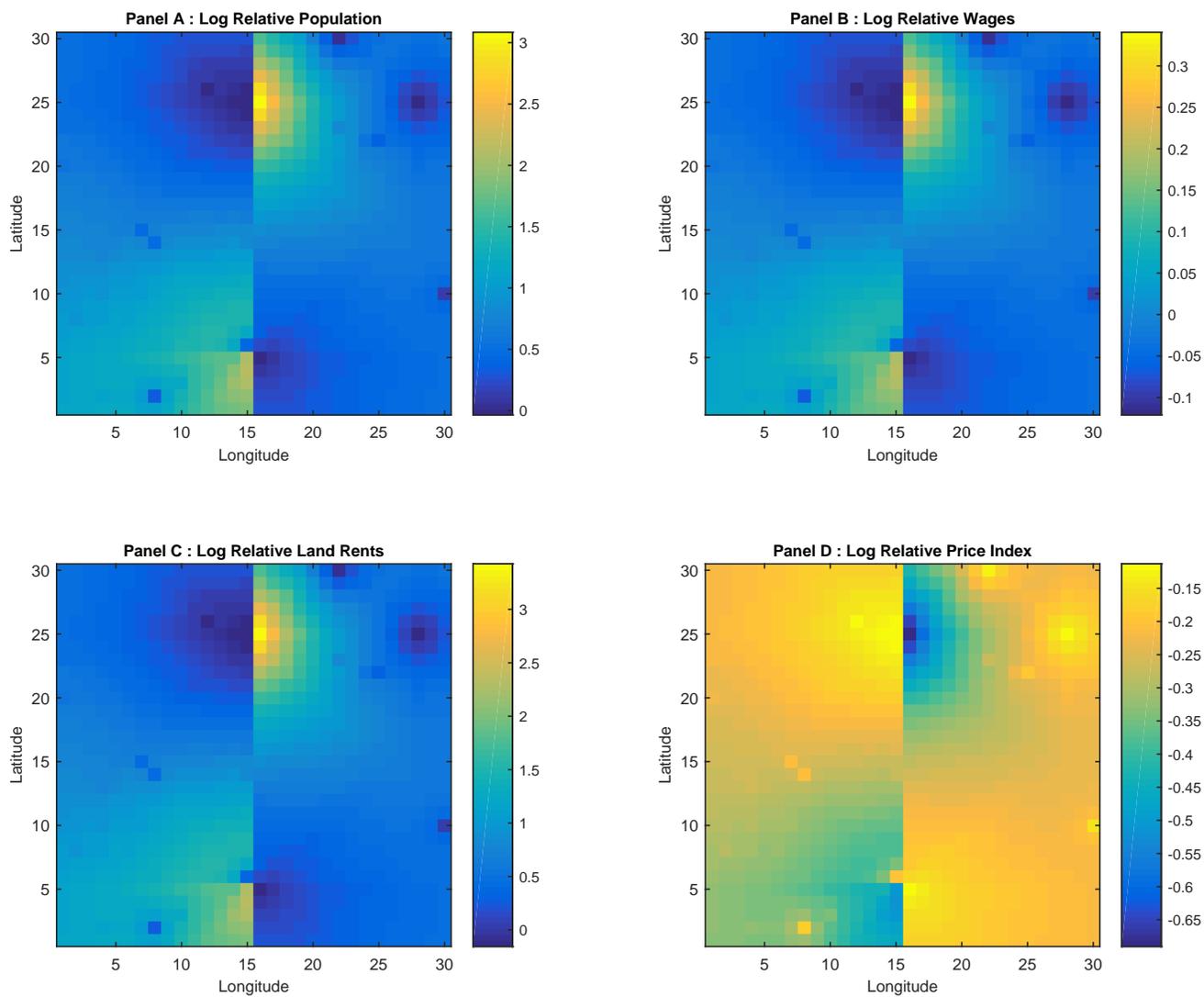
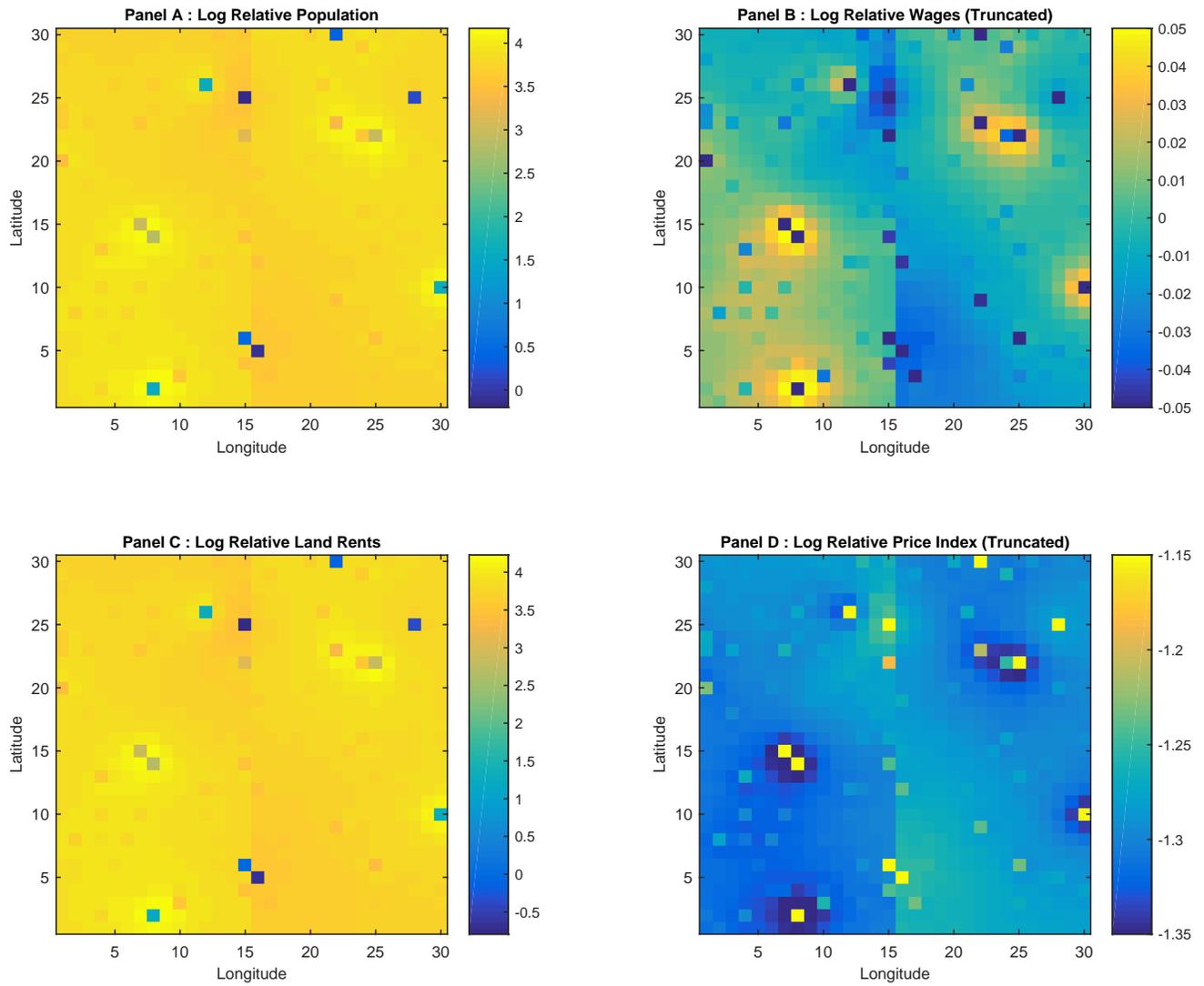


Figure 4: Internal Liberalization (Ratio of Counterfactual to Initial Values,  $\hat{x} = x'/x$ )



	Welfare Gain West ( $\hat{V} = V'/V$ )	Welfare Gain East ( $\hat{V} = V'/V$ )
External Liberalization	0.2%	0.3%
Internal Liberalization	1.4%	2.3%

Table 1: Welfare Gains from External and Internal Liberalization