TRADE AND THE TOPOGRAPHY OF THE SPATIAL ECONOMY*

TREB ALLEN AND COSTAS ARKOLAKIS

We develop a general equilibrium framework to determine the spatial distribution of economic activity on any surface with (nearly) any geography. Combining the gravity structure of trade with labor mobility, we provide conditions for the existence, uniqueness, and stability of a spatial economic equilibrium and derive a simple set of equations that govern the relationship between economic activity and the geography of the surface. We then use the framework to estimate the topography of trade costs, productivities and amenities in the United States. We find that geographic location accounts for at least twenty percent of the spatial variation in U.S. income. Finally, we calculate that the construction of the interstate highway system increased welfare by 1.1 to 1.4 percent, which is substantially larger than its cost. *JEL* Codes: R12, F10, R13, R40.

I. INTRODUCTION

There exists an enormous disparity in economic activity across space. For example, in 2000, the population density in McLeod County, MN, was 26 persons/km^2 and the payroll per capita was \$13,543, whereas in Mercer County, NJ, the population density was 369 persons/km^2 and the payroll per capita was \$20,795 (MPC 2011b). Many explanations for this disparity focus on the characteristics of a location that affect either the productivity or the amenity value of living there (e.g., climate, natural resources, institutions).[1] These explanations ignore the role of geographical location: if the local characteristics of McLeod County were identical to those of Mercer County, such explanations would imply that the two locations should have the same

*We thank George Alessandria, Lorenzo Caliendo, Arnaud Costinot, Dave Donaldson, Gilles Duranton, Jonathan Eaton, Michal Fabinger, Pablo Fajgelbaum, Gene Grossman, Gordon Hanson, Johannes Horner, Sam Kortum, Kiminori Matsuyama, Stephen Redding, Andres Rodriguez-Clare, Esteban Rossi-Hansberg, Antonios Stampoulis, and Jonathan Vogel for their helpful suggestions and comments. We are indebted to Xiangliang Li for excellent research assistance. This article started while Allen and Arkolakis were visitors at the International Economics Section at Princeton University, the hospitality of whom we gratefully acknowledge. All errors are our own.

1. The literature examining the factors contributing to the productivity of a location is immense. See, for example, Sachs (2001), Acemoglu, Johnson, and Robinson (2002).

The Quarterly Journal of Economics (2014), 1085–1139. doi:10.1093/qje/qju016.
Advance Access publication on May 9, 2014.

economic activity. In contrast, the theoretical literature in spatial economics developed over the past few decades emphasizes that because trade over space is costly, geographical location plays an important role by affecting how remote a location is from economic activity elsewhere.

How much of the observed spatial disparity in economic activity is due to geographic location? Unfortunately, the simplicity of the spatial structure postulated in most spatial economic models has restricted their direct applicability to a set of stylized examples. In this article, we resolve this tension between theory and data by developing a new framework that allows us to determine the equilibrium spatial distribution of economic activity on any surface with (nearly) any geography. With this framework, we perform a quantitative empirical analysis to estimate the fraction of spatial inequality in incomes in the United States that is due to variation in trade costs arising from differences in geographic location.

Our theoretical framework relies on an economic and a geographic component, which are distinct but mutually compatible. The economic component combines the gravity structure of international trade with labor mobility to determine the equilibrium distribution of economic activity on a space with any continuous topography of exogenous productivity and amenity differences and any continuous bilateral iceberg trade costs.[2] To incorporate the possibility of productivity or congestion externalities, we allow for the overall productivity and amenity in a location to endogenously depend on its population (local "spillovers"). Given this setup, we show that the equilibrium conditions can be formulated as a set of integral equations, which allows us to apply a set of conventional mathematical theorems to characterize conditions for the existence, uniqueness, and stability of a spatial economic equilibrium. In turn, this equilibrium system yields simple relationships between the endogenous economic outcomes and the underlying geography of the surface and highlights the role that spillovers play in determining the equilibrium spatial distribution of economic activity.

The geographic component provides a micro-foundation for the bilateral trade costs. We suppose that there exists a

2. The idea of analyzing economic activity on a surface has a long tradition, see Beckmann (1952) and Beckmann and Puu (1985, 1990), and has also been recently used in Krugman and Venables (1995) and Mossay and Picard (2011).

topography of instantaneous trade costs over a surface. The bilateral trade costs are then equal to the accumulation of these instantaneous trade costs over the least-cost route. We use methods from differential geometry to characterize the bilateral trade costs between any two points in space. Combining the economic and geographic frameworks, we provide stylized examples where the equilibrium can be described as the solution of a differential equation and derive closed-form solutions to the equilibrium distribution of population for some simple geographies (e.g., the line).[3]

Combining the economic and geographic components, we analyze the real-world distribution of economic activity throughout the continental United States. We begin by estimating the underlying geography—the bilateral trade costs, productivities, and amenities—of the United States. To estimate the bilateral trade costs, we combine detailed geographic information on the rail, road, and water networks with mode-specific bilateral trade shares to infer the relative cost of travel using different modes of transportation. The procedure is greatly facilitated by the "fast marching method" algorithm borrowed from computational physics, which allows us to efficiently compute the trade cost along the least-cost route from all locations to all other locations. Given the trade costs, we then identify the unique topography of composite productivities and amenities that exactly match the observed spatial distribution of wages and population given the structure of the model.

We then perform two exercises using the estimated geography of the United States. First, we estimate what fraction of the observed variation in income can be explained by geographic location. Because the model yields a log-linear relationship between the income of a location and its productivity, amenity, and price index (which is a sufficient statistic for geographic location), we can apply a standard decomposition technique to determine how much of the observed variation in income the price index can explain. The decomposition implies that at least 20% of the spatial variation in income across the United States in 2000 can be explained by geographic location alone. Second, we examine the effect of removing the Interstate Highway System. We estimate

3. Analytical characterization of the equilibrium distribution of economic activity across space is provided in a case-by-case basis in Matsuyama (1999) and Fabinger (2011) when labor is immobile and in Fujita and Thisse (2013) for various examples when labor is mobile.

that without the Interstate Highway System, welfare would decline by 1.1% to 1.4%, suggesting that the benefits of the Interstate Highway System substantially outweigh the costs.

Our framework departs from the seminal economic geography model of Krugman (1991) (which is extensively analyzed in Fujita, Krugman, and Venables 1999) in two important ways. First, we dispense with the assumption of a homogeneous freely traded good, thereby allowing nominal wages to vary across space. Second, we depart from the tradition of a monopolistic competition structure, instead using a perfect competition Armington setup with differentiated varieties as in Anderson (1979) and Anderson and Van Wincoop (2003).

Unlike much of this literature, rather than taking a stand on the source of production or congestion externalities, we incorporate such spillovers by simply assuming that productivity and amenities may depend in part on the local population.[4] While ad hoc, this assumption allows us to show that for particular strengths of spillovers, our model becomes isomorphic to many other spatial economic models, including the free entry monopolistic competition setup similar to the one considered by Krugman (1980, 1991) and the fixed amenity framework of Helpman (1998) and Redding and Sturm (2008). By showing how spillovers affect the existence, uniqueness, and stability of the equilibrium as well as the relationship between equilibrium economic activity and the underlying geography, our framework provides a link between these previously distinct spatial theories.

Our model is also related to a large literature on urban development based on the framework of Roback (1982), as in Kline and Moretti (2014) and Diamond (2012). These papers assume free labor and capital mobility and costless trade of a homogeneous commodity. Although our model relies on differentiated goods to provide a dispersion force, it turns out that when trade is costless, the equilibrium conditions for our model is equivalent to versions of the Roback (1982) model; hence, our framework can be interpreted as an extension of the Roback (1982) framework to a world with costly trade.

While there has been much empirical work examining the implication of space for the allocation of people (Davis and

4. Unlike Rossi-Hansberg (2005), we restrict such spillovers to be local. For the examination of micro-foundations of spillovers, see Lucas and Rossi-Hansberg (2003), Duranton and Puga (2004), and Rossi-Hansberg and Wright (2007).

Weinstein 2002, 2008) and wages (Hanson 2005; Breinlich 2006; Head and Mayer 2006; Amiti and Cameron 2007), there has been little empirical application of the extensive body of theoretical research on economic geography. In recent exceptions to this rule, Redding and Sturm (2008), Redding (2012), and Ahlfeldt et al. (2012) use a quantitative framework to analyze the spatial distribution of economic activity. This article follows in their tradition and develops a number of tools to facilitate future quantitative analysis of economic geography.

Finally, our empirical work is related to a recent literature estimating the impact of large changes to transportation infrastructure on economic output. Donaldson (2012) and Pérez-Cervantes (2012) consider the impact of railroads in India and the United States, respectively, when labor is immobile, whereas Donaldson and Hornbeck (2012) consider the effect of the construction of the railroad network in the United States when labor is mobile. Though we show how such transportation networks can be incorporated in our framework, we can also incorporate geographical characteristics that do not have obvious network representations.

The remainder of the article is organized as follows. The next section presents the theoretical framework and the third section presents the empirical analysis. The last section concludes.

II. Theoretical Framework

This section describes our theoretical framework. It comprises three subsections. We first present the economic component of the framework, where we describe the equilibrium distribution of economic activity in a space with arbitrary trade costs. Second, we present the geographic component of the framework, where we define and characterize geographic trade costs that arise from moving goods across a surface. Finally, we combine the economic and geographic components to characterize the equilibrium distribution of economic activity for several geographies.

II.A. Economic Component

In this subsection, we present the economic component of our framework and characterize the existence, uniqueness, and stability of a spatial equilibrium.

1. Setup. The world consists of a continuum of locations $i \in S$, where S is a closed bounded set of a finite dimensional Euclidean space with the Euclidean norm as its metric.[5] Each location $i \in S$ produces a unique differentiated variety of a good. Trade is costly: trade costs are of the iceberg form and are described by the function $T : S \times S \rightarrow [1, \infty)$, where $T(i,j)$ is the quantity of a good needed to be shipped from location i in order for a unit of a good to arrive in location j. We normalize $T(i,i) = 1$ for all locations.

The world is inhabited by a measure $\bar{L}$ of workers who are freely mobile across locations and derive utility from the consumption of differentiated varieties and the local amenity. In particular, we assume that workers have identical constant elasticity of substitution (CES) preferences over the continuum of differentiated varieties, so that the total welfare in location $i \in S$, $W(i)$, can be written as:

$$W(i) = \left(\int_{s \in S} q(s,i)^{\frac{\sigma-1}{\sigma}} ds \right)^{\frac{\sigma}{\sigma-1}} u(i),$$

where $q(s,i)$ is the per capita quantity of the variety produced in location s and consumed in location i, $\sigma \in (1, \infty)$ is the elasticity of substitution between goods ω, and $u(i)$ is the local amenity.[6]

Labor is the only factor of production. Each worker provides a unit of labor inelastically in the location where she lives, for which she is compensated with a wage. A worker in location i produces $A(i)$ units of a good, where $A(i)$ is the local productivity. Production is assumed to be perfectly competitive. We define the functions $L : S \rightarrow \mathbb{R}_{+}$ and $w : S \rightarrow \mathbb{R}_{++}$ to be the density of workers and their wage, respectively.

To allow for the possibility of productivity or congestion externalities, both productivity and amenities may depend on

5. The continuum of locations is not important for much of what follows. In particular, as we discuss later, Theorems 1 and 2 generalize for the case of discrete number of locations; however, Proposition 1 is only true with a continuum of locations since it relies on the fact that a change in the population in one location does not affect the price index.

6. While the model attains a nontrivial solution even for $\sigma \in (0, 1)$, we focus on the case where $\sigma > 1$ so that the elasticity of trade flows to trade costs is negative.

the density of workers. In particular, we assume that overall (or composite) productivity in location i can be written as:

$$A(i) = \overline{A}(i)L(i)^{\alpha}, \tag{1}$$

where $\overline{A}(i)$ is the exogenous component of productivity inherent to location i and $\alpha \in \mathbb{R}$ determines the extent to which productivity is affected by the population density. Similarly, we assume that the overall amenity in location i can be written as:

$$u(i) = \overline{u}(i)L(i)^{\beta}, \tag{2}$$

where $\overline{u}(i)$ is the exogenous utility derived from living in location i inherent to the location and $\beta \in \mathbb{R}$ determines the extent to which amenities are affected by the population density. In what follows, we refer to α and β as governing the strength of productivity and amenity spillovers, respectively. Although we make no theoretical restrictions regarding α or β, in what follows we focus on the empirically relevant cases of $\alpha \geq 0$ and $\beta \leq 0$. It is important to note that these spillovers are assumed to be local in nature (i.e., they do not affect the productivity or amenities in nearby regions).

In Online Appendix A.2, we show how particular productivity and amenity spillovers make our framework isomorphic to other spatial economic models. In particular, if $\alpha = \frac{1}{\sigma-1}$, our model is isomorphic to a monopolistically competitive framework with differentiated varieties and free entry, where the number of varieties produced in a location is proportional to its population. The productivity spillover can be interpreted in this sense as an agglomeration externality caused by more entry in markets with a larger size, as in the standard geography setup of Krugman (1991).

Similarly, if $\alpha = \frac{1}{\sigma-1}$ and $\beta = -\frac{1-\delta}{\delta}$, our model is isomorphic to the Helpman (1998) and Redding (2012) framework with $1-\delta$ being the budget share spent on an immobile factor, for example, land or housing. In this case, the value of β is negative, capturing the inelastic supply of land or housing and the resulting congestion externality through their increased prices. Intuitively, the amenity spillover can be interpreted as capturing the disutility of higher housing prices. The model is also isomorphic to a model in which land is a factor of production if $\alpha = \delta - 1$, where δ is the share of labor in the Cobb-Douglas production function, in which case the productivity spillover can be interpreted as also capturing the diminishing returns to labor in the production function. Finally, our model is isomorphic to one in which workers have heterogeneous preferences (drawn from an extreme value distribution) for living in different locations, so that the amenity spillover can be interpreted as the extent to which workers differ in their locational preferences. Notice that the isomorphisms we discussed before regard trade flows, wages, population, and welfare, but not necessarily other aspects of these models. Independent of their interpretation, the degree of this agglomeration and dispersion externalities are crucial to guarantee uniqueness and existence of a spatial equilibrium, as we discuss in detail later.

We define the geography of S to be the set of functions $\overline{A}$, $\overline{u}$, and T, where $\overline{A}$ and $\overline{u}$ make up the local characteristics and T makes up the geographic location. S is said to have a regular geography if $\overline{A}$, $\overline{u}$, and T are continuous and bounded above and below by strictly positive numbers. We define the distribution of economic activity to be the set of functions w and L, where we normalize $\int_S w(s)\,ds = 1$. Finally, the topography of the spatial economy is the complete set of functions comprising the geography and the distribution of economic activity of S.

2. Gravity. We first determine bilateral trade flows as a function of the geography of the surface, the wages, and the labor supply. The function $X(i,j)$ expresses the value of bilateral trade flows from location i to location j, where $X : S \times S \to \mathbb{R}_+$. Using the CES assumption and the fact that with perfect competition the final price of the good produced in location i and sold in location j is equal to the marginal production and shipping cost, $\frac{w(i)}{A(i)} T(i,j)$, the value of location j's imports from location i can be expressed as:

$$X(i,j) = \left(\frac{T(i,j)w(i)}{A(i)P(j)}\right)^{1-\sigma} w(j)L(j), \tag{3}$$

where $P(j)$ is the CES price index with

$$P(j)^{1-\sigma} = \int_S T(s,j)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma}\, ds. \tag{4}$$

3. Equilibrium. The CES assumption implies that the welfare of living in a particular location can be written as an indirect function of the real wage and the overall amenity value:

$$W(i) = \frac{w(i)}{P(i)} u(i). \tag{5}$$

Welfare is said to be equalized if for all $i \in S$ there exists a $W > 0$ such that $W(i) \leq W$, with equality if $L(i) > 0$. That is, welfare is equalized if the welfare of living in every inhabited location is the same and the welfare of living in every uninhabited location is no greater than the welfare of the inhabited locations.

Markets are said to clear if the income is equal to the value of goods sold in all locations, that is, for all $i \in S$:

$$w(i)L(i) = \int_S X(i, s)\, ds. \tag{6}$$

Given a regular geography with parameters σ, α, and β, we define a *spatial equilibrium* as a distribution of economic activity such that (i) markets clear, (ii) welfare is equalized, and (iii) the aggregate labor market clears:

$$\int_S L(s)\, ds = \bar{L}. \tag{7}$$

In what follows, we pay particular attention to spatial equilibria with the following two features. A spatial equilibrium is said to be regular if w and L are continuous and every location is inhabited, that is, for all $i \in S$, $L(i) > 0$. A spatial equilibrium is said to be point-wise locally stable if $\frac{dW(i)}{dL(i)} < 0$ for all $i \in S$. Intuitively, a point-wise locally stable equilibrium is one where no small number of workers can increase their welfare by moving to another location.[7]

4. Existence, Uniqueness, and Stability. We now discuss sufficient conditions for the existence and uniqueness of regular spatial equilibria. Using equation (3) to substitute for trade flows

7. This concept of stability is an adaptation of the one first introduced by Krugman (1991) to a continuum of locations.

and the indirect utility function (5), we can write the market clearing condition (6) for all $i \in S$ as:

$$L(i)w(i)^{\sigma} = \int_S W(s)^{1-\sigma} T(i,s)^{1-\sigma} A(i)^{\sigma-1} u(s)^{\sigma-1} L(s) w(s)^{\sigma}\, ds. \tag{8}$$

Combining the indirect utility function (5) with the price index (4) yields:

$$w(i)^{1-\sigma} = \int_S W(i)^{1-\sigma} T(s,i)^{1-\sigma} A(s)^{\sigma-1} u(i)^{\sigma-1} w(s)^{1-\sigma}\, ds. \tag{9}$$

When there are no productivity or amenity spillovers (i.e., $\alpha = \beta = 0$ so that $A(i) = \overline{A}(i)$ and $u(i) = \overline{u}(i)$) and welfare is equalized so that $W(i) = W$ for all $i \in S$, equations (8) and (9) are linear operators whose eigenfunctions are $L(i)w(i)^{\sigma}$ and $w(i)^{1-\sigma}$ and whose eigenvalues are $W^{\sigma-1}$, respectively. Note that the kernels of the two equations are transposes of each other. These two results allow us to prove the following theorem:

THEOREM 1. *Consider a regular geography with exogenous productivity and amenities. Then:*

(i) *there exists a unique spatial equilibrium and this equilibrium is regular; and*
(ii) *this equilibrium can be computed as the uniform limit of a simple iterative procedure.*

Proof. See Online Appendix A.1.1. ■

Equations (8) and (9) can be viewed as a linear system of equations for which extensions of standard results in linear algebra guarantee the existence and uniqueness of a positive solution. Part (ii) of Theorem 1 guarantees that the equilibrium wages and population can be calculated quickly without the need of a good prior guess.

When there are productivity or amenity spillovers and welfare is equalized, substituting equations (1) and (2) into equations (8) and (9) yields the following two equations:

$$L(i)^{1-\alpha(\sigma-1)} w(i)^{\sigma} = W^{1-\sigma} \int_S T(i,s)^{1-\sigma} \overline{A}(i)^{\sigma-1} \overline{u}(s)^{\sigma-1} L(s)^{1+\beta(\sigma-1)} w(s)^{\sigma}\, ds, \tag{10}$$

$$w(i)^{1-\sigma}L(i)^{\beta(1-\sigma)} = W^{1-\sigma}\int_S T(s,i)^{1-\sigma}\overline{A}(s)^{\sigma-1}\overline{u}(i)^{\sigma-1}w(s)^{1-\sigma}L(s)^{\alpha(\sigma-1)}\,ds. \tag{11}$$

Equations (10) and (11) are a system of two nonlinear integral equations; such systems have only recently begun to be studied in the mathematics literature (see, e.g., Yang and O'Regan 2005). However, when bilateral trade costs are symmetric, that is, $T(i,s) = T(s,i)$ for all i, $s \in S$, it turns out that the system can be written as a single nonlinear integral equation, which will allow us to provide a simple characterization of the equilibrium system.[8] To see this, suppose that

$$L(i)A(i)^{1-\sigma}w(i)^{\sigma} = \phi w(i)^{1-\sigma}u(i)^{1-\sigma}, \tag{12}$$

where $\phi > 0$ is some scalar. Given equations (1) and (2) governing the strength of spillovers, it is straightforward to show that if equation (12) holds, then any functions $w(i)$ and $L(i)$ satisfying equation (10) will also satisfy (11) (and vice versa). We prove in the subsequent theorem that for any regular equilibrium, equation (12) is the unique relationship between $L(i)$ and $w(i)$ such that equations (10) and (11) hold.

Substituting equations (12), (1), and (2) into either equation (10) or (11) yields (after some algebra):

$$\begin{aligned} L(i)^{\tilde{\sigma}\gamma_1} &= \overline{u}(i)^{(1-\tilde{\sigma})(\sigma-1)}\overline{A}(i)^{\tilde{\sigma}(\sigma-1)}W^{1-\sigma} \\ &\times \int_S T(s,i)^{1-\sigma}\overline{A}(s)^{(1-\tilde{\sigma})(\sigma-1)}\overline{u}(s)^{\tilde{\sigma}(\sigma-1)}\left(L(s)^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}}ds, \end{aligned} \tag{13}$$

where

$$\gamma_1 \equiv 1 - \alpha(\sigma-1) - \beta\sigma,$$

$$\gamma_2 \equiv 1 + \alpha\sigma + (\sigma-1)\beta,$$

and $\tilde{\sigma} \equiv \frac{\sigma-1}{2\sigma-1}$.

8. This method of reducing a system of nonlinear equations into a single nonlinear equation when trade costs are symmetric can also be applied more generally to prove the existence and uniqueness of the equilibrium of trade models where welfare does not necessarily equalize; see Allen et al. (2014). Indeed, the method also works if trade costs are "quasi-symmetric," that is, if $T(i,s) = T_A(i)T_B(s)\tilde{T}(i,s)$, for any functions $T_A : S \to \mathbb{R}_{++}$ and $T_B : S \to \mathbb{R}_{++}$, where $\tilde{T}(i,s) = \tilde{T}(s,i)$ for all $i, s \in S$.

Note that equation (13) characterizes the equilibrium distribution of labor as a function only of the underlying geography of the surface; wages, in particular, do not enter. Equation (13) is a nonlinear integral equation known as a homogeneous Hammerstein equation of the second kind (see, e.g., Polyanin and Manzhirov 2008, p. 807). If equation (13) has a solution for $L(i)$ and $W^{1-\sigma}$ then equilibrium wages can be determined from equation (12) using the aggregate labor clearing condition to determine the scalar ϕ. The next theorem discusses the conditions for existence and uniqueness of spatial equilibria.

THEOREM 2. *Consider a regular geography with overall productivity and amenity functions specified in equations (1) and (2), respectively, and assume that iceberg trade costs are symmetric and parameters are such that* $\gamma_1 \neq 0$*. Then:*

(i) *there exists a regular spatial equilibrium;*
(ii) *if* $\gamma_1 > 0$*, all equilibria are regular;*
(iii) *if* $\frac{\gamma_2}{\gamma_1} \in [-1, 1]$*, the spatial equilibrium is unique; and if* $\frac{\gamma_2}{\gamma_1} \in (-1, 1]$*, it can be computed as the uniform limit of a simple iterative procedure.*

Proof. See Online Appendix A.1.2. ■

Note that $\frac{\gamma_2}{\gamma_1} \in [-1, 1]$ implies $\gamma_1 > 0$, so that part (iii) holds only if part (ii) holds as well. It is straightforward to show that if $\gamma_1 = 0$ there is (generically) no regular spatial equilibrium satisfying equations (10) and (11). Finally, the following proposition characterizes when a spatial equilibria is point-wise locally stable.

PROPOSITION 1. *Consider a regular geography with overall productivity and amenity functions specified in equations (1) and (2), respectively, and parameters such that* $\gamma_1 \neq 0$*. Then if* $\gamma_1 < 0$*, no regular equilibria is point-wise locally stable, and if* $\gamma_1 > 0$*, all equilibria are point-wise locally stable.*

Proof. See Online Appendix A.1.4. ■

To get intuition for this result notice that when markets clear, the welfare of living in a location can be written as:

$$W(i) = \frac{\left(\int_S T(i,s)^{1-\sigma} P(s)^{\sigma-1} w(s) L(s)\, ds\right)^{\frac{1}{\sigma}}}{P(i)} \overline{A}(i)^{\frac{\sigma-1}{\sigma}} \overline{u}(i) L(i)^{-\frac{\gamma_1}{\sigma}}. \tag{14}$$

The parameter γ_1 is the partial elasticity of welfare with respect to the population in a location. Expression (14) shows that if a small number of workers moves to a location, the welfare in that location will decrease if and only if $\gamma_1 > 0$.[9]

Figure I depicts the ranges of $\alpha \geq 0$ and $\beta \leq 0$ and the different cases of equilibrium uniqueness and stability with $\sigma = 9$ (a complete characterization for $\alpha, \beta \in \mathbb{R}$ is presented in Online Appendix A). The graph is divided in four regions with sufficient conditions on α and β for uniqueness and stability. Focusing on the range where $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$, we see that $\frac{\gamma_2}{\gamma_1} \in [-1, 1]$ if and only if $\alpha + \beta \leq 0$, so there is a unique stable equilibrium regardless of the economic geography as long as dispersion forces are at least as strong as agglomeration forces. When $\alpha + \beta > 0$ but is small, there exists an equilibrium that is stable (since $\gamma_1 > 0$) but it need not be unique (since $\frac{\gamma_2}{\gamma_1} > 1$). We provide specific examples of the possible multiple equilibria later. However if $\alpha + \beta$ increases enough so that $\gamma_1 \leq 0$, the agglomeration forces are sufficiently strong that they can induce complete concentration in a single location, that is, a black hole. Black holes are the only possible equilibria when $\gamma_1 = 0$; however, if $\gamma_1 < 0$, regular equilibria also exist (although they are not point-wise locally stable).[10]

The existence and uniqueness results of Theorems 1 and 2 generalize for a discrete number of locations, as we discuss in Online Appendix A.1.3, in which case the set S is finite or countable.[11] However, with a discrete number of locations, stability has to be analyzed in a case-by-case basis as in Fujita, Krugman, and Venables (1999) because a change in the population in one location will affect the price index.

9. The fact that the competitive equilibrium exists and is stable means that the spatial impossibility result of Starrett (1978) does not apply in our case. The difference arises from the fact that in our model, the production set of firms differs across locations because of the Armington assumption. It differs since in the Armington model, as in the monopolistic competition models (e.g., Krugman 1979; Helpman and Krugman 1985), varieties are location-specific.

10. Notice that if $\alpha \geq 0$ and $\beta \leq 0$ and $\gamma_1 < 0$, the condition for uniqueness is not satisfied. However, there do exist alternative configurations of $\alpha < 0$ and $\beta > 0$ in which there is a unique point-wise locally unstable equilibrium; see Online Appendix A.

11. Continuity extends to the discrete topology in a trivial way since any function in a discrete topology is continuous. If S is finite or countable, a Lebesgue integral can be considered and $\int_S$ is formally equivalent to $\sum_S$.

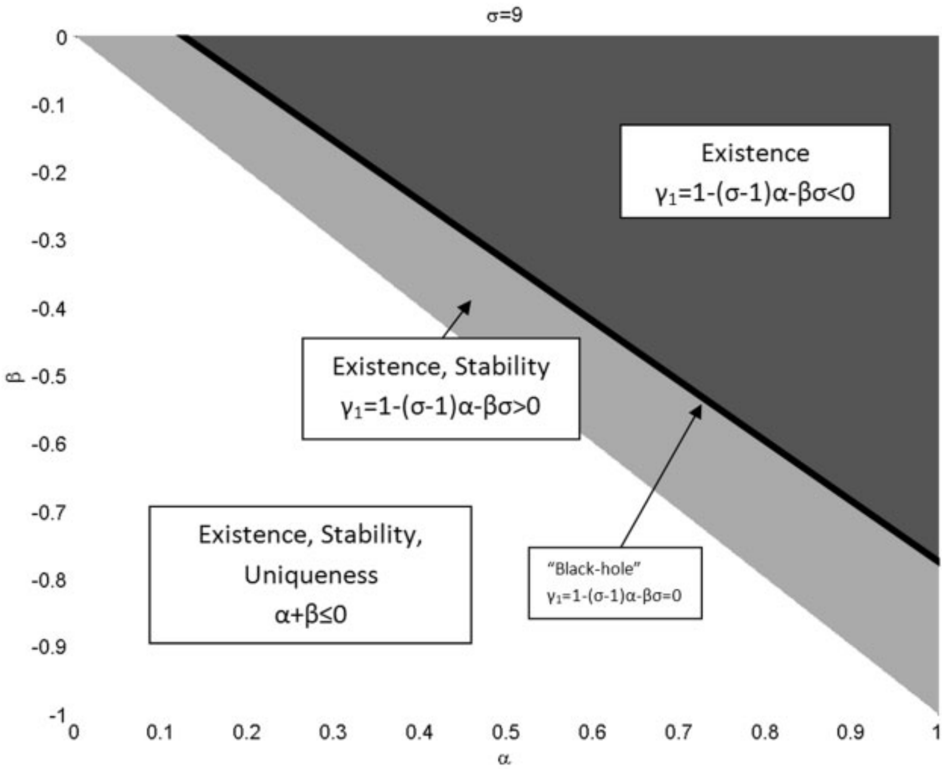


FIGURE I

Equilibria with Amenity and Productivity Spillovers

This figure shows the regions of values for the productivity spillover α and the amenity spillover β for which there exists an equilibrium, for which there exists a point-wise locally stable equilibrium, and whether that equilibrium is unique. The elasticity of substitution σ is chosen to equal 9.

5. Equilibrium Economic Activity and the Underlying Geography. When trade costs are symmetric, equations (5) and (12) (along with equations (1) and (2) governing the strength of spillovers) imply that in a regular equilibrium both wages $w(i)$ and population $L(i)$ can be written as log linear functions of the exogenous local characteristics and the price index:

$$\gamma_1 \ln w(i) = C_w - \beta(\sigma-1)\ln \bar{A}(i) - (1-\alpha(\sigma-1))\ln \bar{u}(i) + (1+(\sigma-1)(\beta-\alpha))\ln P(i), \tag{15}$$

$$\gamma_1 \ln L(i) = C_L + (\sigma-1)\ln \bar{A}(i) + \sigma \ln \bar{u}(i) + (1-2\sigma)\ln P(i), \tag{16}$$

where the scalars C_w and C_L are determined by the wage normalization and the labor market clearing, respectively.

Equations (15) and (16) provide three important implications regarding the relationship between the equilibrium distribution of economic activity and the geography of the space. First, because bilateral trade costs only appear in the price index, the price index is a sufficient statistic for geographic location. Second, as long as $\gamma_1 > 0$, the population will be higher in locations with high exogenous productivities and amenities and lower in locations with higher price indices. In contrast, the equilibrium wages may increase or decrease depending on the sign of β as the underlying productivity increases and may increase or decrease as the exogenous amenity of a location or price index increases, depending on the signs of $1-\alpha(\sigma-1)$ and $1+(\sigma-1)(\beta-\alpha)$, respectively. Third, conditional on the price index (which, you will recall, is an endogenous variable itself), productivity and amenity spillovers only change the elasticity of the equilibrium distribution of economic activity to the underlying geography.[12] If $\gamma_1 > 0$, stronger spillovers (i.e., larger α or β) result in the equilibrium distribution of population becoming more sensitive to underlying geographic differences.

II.B. Geographic Component

In this subsection, we present a micro-foundation for the bilateral trade cost function by assuming that bilateral trade costs are the total trade costs incurred traveling from an origin to a destination along the least-cost route.

Suppose now that S is a compact manifold in $\mathbb{R}^N$.[13] In what follows, we focus on the one-dimensional cases where S is a finite line or a finite circle and the two-dimensional case where S is a finite plane, although the following results hold for any finite-dimensional manifold.

Let $\tau : S \to \mathbb{R}_+$ be a continuous function where $\tau(i)$ gives the "instantaneous" trade cost incurred by crossing point $i \in S$. Define

12. Unless trade costs are zero, the strength of productivity and amenity spillovers will also affect the equilibrium distribution of population through general equilibrium effects on the price index.

13. A manifold is a topological space that is locally Euclidean, or intuitively, a space that can be "charted" in Euclidean space.

$t(i,j)$ to be the solution to the following least-cost path minimization problem:

$$t(i,j) = \inf_{g \in \Gamma(i,j)} \int_0^1 \tau(g(t)) \left\| \frac{dg(t)}{dt} \right\| dt, \tag{17}$$

where $g : [0, 1] \to S$ is a path and $\Gamma(i,j) \equiv \{g \in C^1 | g(0) = i, g(1) = j\}$ is the set of all possible continuous and once differentiable paths that lead from location i to location j. The notation $\| \cdot \|$ stands for the Euclidean norm. If the bilateral trade cost function T is such that for all $i, j \in S$, $T(i,j) = f(t(i,j))$, for some monotonically increasing function $f : \mathbb{R}_+ \to [1, \infty)$ with $f(0) = 1$, we say that the bilateral trade costs are *geographic*. Note that when bilateral trade costs are geographic, there exists a unique mapping from the instantaneous trade cost function τ (which has a domain of S) to the bilateral trade costs T (which has a domain of $S \times S$), so that assuming trade costs are geographic reduces the dimensionality of the problem by its square root.

Geographic trade costs provide a flexible means of approximating the true costs associated with moving goods across space. Transportation networks such as roads and railroads can be incorporated by assuming that the instantaneous trade costs are lower where roads or railroads exist. Borders can be incorporated by constructing (positive measure) "walls" between regions where the instantaneous trade costs are large; such "walls" can also be placed alongside roads or railroads to so that they are accessible at only a finite number of entrance ramps or stations. The instantaneous trade costs can also reflect differences in natural geography, such as ruggedness and water. Two properties of geographic trade costs deserve special mention. First, because traveling over a particular point $i \in S$ incurs the same cost regardless of the direction of travel, geographic trade costs are symmetric, that is, for all $i, j \in S$, $T(i,j) = T(j,i)$. Second, because the topography of the surface is smooth, nearby locations will face similar trade costs to all other destinations. Formally, for all $s, i, j \in S$, we have $\lim_{s \to i} T(s,j) = T(i,j)$. Although we believe these are attractive properties for trade costs arising from transportation costs, they abstract from alternative sources of trade costs, e.g. origin-specific tariffs or information frictions (see, e.g., Allen forthcoming). We will allow for such nongeographic trade costs when we estimate the total bilateral trade costs for the United States in Section III.

Equation (17) is a well-studied problem that arises in a number of fields. For any origin $i \in S$ and destination $j \in S$, its solution is characterized by the following eikonal partial differential equation (see, e.g., Mantegazza and Mennucci 2003):

$$\|\nabla t(i,j)\| = \tau(j), \tag{18}$$

where the gradient is taken with respect to the destination j.

Because we care only about the total bilateral trade costs (rather than the actual least-cost route), for our purposes it suffices to focus on the set of iso-cost contours, that is, the set of curves defined by the set of destinations $\{j|t(i,j) = C\}$ for all C. Equation (18) implies that as C increases, the iso-cost contour expands outward at a rate inversely proportional to the instantaneous trade cost in a direction that is orthogonal to the contour curve. Hence, the evolution of the contour of the bilateral trade costs is equivalent to the propagation of a wave front outward from the origin along the surface at a speed inversely proportional to the instantaneous trade cost. Intuitively, when instantaneous trade costs are large, the iso-cost contour expands more slowly, reflecting the fact that a given increase in distance results in a larger increase in the total geographic trade costs.

For any initial point $i \in S$, it is possible to determine the bilateral trade costs to all other destinations $j \in S$ using a simple iterative procedure based on the eikonal equation (18). Given any contour set $\{j|t(i,j) = C\}$, we can construct for each $j \in \{j|t(i,j) = C\}$ a vector from j of length $\frac{\varepsilon}{\tau(j)}$ and orthogonal to the iso-cost contour. By connecting the ends of these vectors, we arrive at a new contour set $\{j'|t(i,j') = C + \varepsilon\}$. Figure II illustrates the propagation process. By starting from an arbitrarily small contour around i, we can apply this process iteratively to determine the complete set of iso-cost contours and hence calculate the bilateral trade cost from i to all destinations $j \in S$. This algorithm is known as the fast marching method (FMM) (see Sethian 1996, 1999).

The FMM relies on the fact that because the instantaneous trade costs are positive everywhere, bilateral trade costs will always increase as one "marches" outward from any iso-cost contour.[14] As a result, subsequent contours can be constructed using

14. This does not imply that the bilateral trade costs always increase with distance from the origin. For example, nearby destinations that are surrounded by "mountains" of high instantaneous transportation costs can have higher bilateral

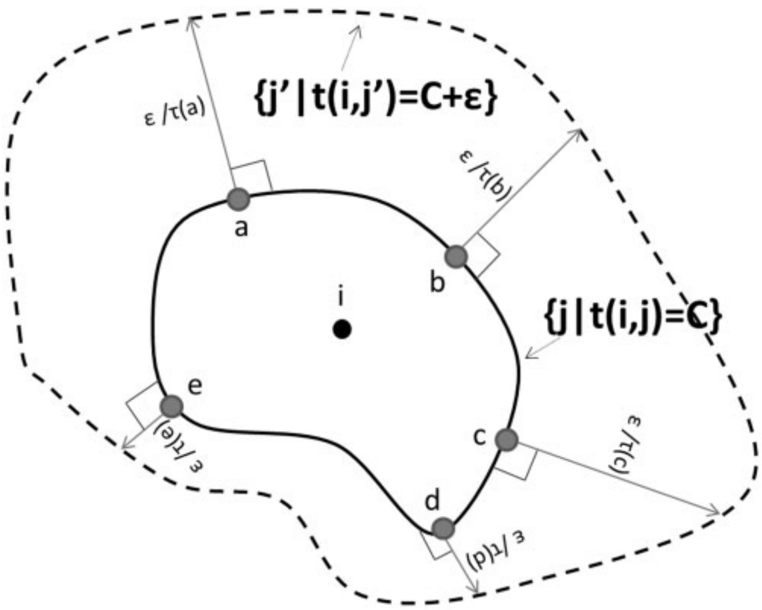


FIGURE II

Propagation of Geographic Trade Costs

This figure shows how the geographic trade costs evolve across a surface. Given a contour of points on a surface such that the geographic trade cost to location i is equal to a constant C (the solid line), for an arbitrarily small $\varepsilon > 0$, we can construct the contour line for bilateral trade costs $C + \varepsilon$ (the dashed line) by propagating the initial contour outward at a rate inversely proportional to the instantaneous trade cost.

only the immediately previous contour. This has a number of implications. First, the FMM is extremely efficient, with a run time of $O(n \log n)$, where n is the number of pixels approximating the instantaneous trade cost function τ. Practically speaking, even with high resolution images of τ, the FMM takes less than a second to determine the distance from any i to all $j \in S$ (however, because FMM has to be run separately for every origin, determining trade costs from all locations to all other locations can take a couple hours at high resolutions).

Second, the FMM bears a close resemblance to the Dijkstra algorithm used to calculate shortest paths over graphs, which

trade costs than further away destinations that can be reached without having to cross such mountains. This occurs because the direction of the propagation is determined locally by the shape of the contour set rather than by the location of the origin.

also relies on an outward expansion from the origin. Indeed, the FMM can be interpreted as a generalization of Dijkstra to continuous spaces: bilateral costs can be determined by approximating a surface with a grid (i.e., a network) and taking the appropriate weighted average over different paths along the grid (see Tsitsiklis 1995). However, it is important to note that applying the Dijkstra algorithm directly using a grid to approximate the space will not result in accurate bilateral distances because of the so-called digitization bias. Digitization bias arises because any chosen grid necessarily restricts the possible directions of travel, biasing estimated distances upward, where the bias is systematically correlated with how different the optimal path is from the allowed directions of travel over the grid (see, e.g., Mitchell and Keirsey 1984).

Third, the FMM can be easily generalized to allow for the direction of travel to affect trade costs, allowing it to incorporate such physical realities as elevation changes or one-way roads. This is because only two pieces of information are required to determine the vector at a point $j \in \{j|t(i,j) = C\}$ used to construct the subsequent iso-cost contour: (i) the slope of the current iso-cost contour (which determines the direction of the vector), and (ii) the instantaneous trade cost (which determines the length of the vector). Because the direction of the vector does not depend on the instantaneous trade cost, we can simply allow the instantaneous trade cost to depend on the direction of travel $\vec{d}$, that is, $\tau(i, \vec{d})$. We provide a simple example of the direction of travel mattering in Section II.C. Note, however, that if instantaneous trade costs are affected by the direction of travel, total bilateral trade costs will no longer be symmetric as Theorem 2 requires.

For the rest of the analysis we use a specific formulation for the geographic costs: $T(i,j) = e^{t(i,j)}$. This exponential form has the interpretation that the instantaneous trade costs are of iceberg form, as it is the limit of the product of many incremental iceberg costs as the distance between the increments tends to zero.[15] That is, the exponential form provides a micro-foundation for why the total bilateral trade costs are of an iceberg form. However, it can be shown (see the the Online Appendix) that any log subadditive monotonically increasing function f such that $f(0) = 1$ will generate bilateral iceberg trade costs that are weakly greater than 1

15. In other words, $e^{\int_a^b \tau(x)dx} = \prod_a^b (1 + \tau(x)dx)$, where $\prod_a^b$ denotes a type II product integral.

and satisfy the triangular inequality, that is, $T(i,j) \le T(i,k)T(k,j)$ for all i, k, j.

II.C. Examples

In this subsection, we present solutions for two simple manifolds when trade costs are geographic: the line and the circle. These two cases help us illustrate the different types of equilibria that may arise and discuss their stability properties.

1. The Line. Let S be the $[-\pi, \pi]$ interval and suppose that $\alpha = \beta = 0$ and $\overline{A}(i) = \overline{u}(i) = 1$, that is, there are no spillovers and all locations have homogeneous exogenous productivities and amenities. Suppose that instantaneous trade costs are constant, that is, $\tau(i) = \tau$ for all $i \in S$ apart from a border b in the middle of the line; that is, trade costs between locations on the same side of the line are $T(i,s) = e^{\tau|i-s|}$ and those on different sides are $T(i,s) = e^{b+\tau|i-s|}$.[16] While the T function in this case is discontinuous, so that the sufficient conditions of Theorems 1 and 2 are not satisfied, we can still obtain a unique explicit solution.

Taking logs of equation (16) and differentiating yields the following differential equation:

$$\frac{\partial \ln L(i)}{\partial i} = (1-2\sigma)\frac{\partial \ln P(i)}{\partial i}. \tag{19}$$

It is easy to show that $\frac{\partial \ln P(-\pi)}{\partial i} = -\tau$ and $\frac{\partial \ln P(\pi)}{\partial i} = \tau$ in the two edges of the line and $\frac{\partial \ln P(0)}{\partial i} = \tau \frac{\left(1-e^{(1-\sigma)b}\right)}{\left(1+e^{(1-\sigma)b}\right)}$ in the location of the border, which gives us boundary conditions for the value of the differential equation at locations $i = -\pi, 0, \pi$. Intuitively, moving rightward while on the far left of the line reduces the distance to all other locations by τ, thereby reducing the (log) price index by τ. To obtain a closed-form solution to equation (19), we differentiate

16. This border cost is reminiscent of the one considered in Rossi-Hansberg (2005). As in that model, our model predicts that increases in the border cost will increase trade between locations that are not separated by border and decrease trade between locations separated by the border. Unlike Rossi-Hansberg (2005), however, in our model the border does not affect what good is produced (since each location produces a distinct differentiated variety) nor is there an amplification effect through spillovers (since spillovers are assumed to be local).

equation (13) twice to show that the equilibrium satisfies the following second-order differential equation:

$$\frac{\partial^2}{\partial i^2} L(i)^{\tilde{\sigma}} = k_1 L(i)^{\tilde{\sigma}} \text{ for } i \in (-\pi, 0) \cup (0, \pi), \tag{20}$$

where $k_1 \equiv (1-\sigma)^2 \tau^2 + 2(1-\sigma)\tau W^{1-\sigma}$ can be shown to be negative. Given the boundary conditions, the equilibrium distribution of labor in both intervals is characterized by the weighted sum of the cosine and sine functions (see example 2.1.2.1 in Polyanin and Zaitsev 2002):

$$L(i)^{\tilde{\sigma}} = k_2 \cos\left(i\sqrt{-k_1}\right) + k_3 \left|\sin\left(i\sqrt{-k_1}\right)\right|.$$

The values of k_1 and the ratio of k_2 to k_3 can be determined using the boundary conditions. Given this ratio, the aggregate labor clearing condition determines their levels.[17] Notice that in the case of no border or an infinite border, the solution is the simple cosine function or two cosine functions one in each side of the border, respectively, and $k_3 = 0$, so that the aggregate labor clearing condition directly solves for k_2.[18]

Figure III depicts the equilibrium labor allocation in this simple case for different values of the instantaneous trade cost but no border. As the instantaneous trade cost increases, the population concentrates in the middle of the interval where the locations are less economically remote. The lower the trade costs, the less concentrated the population; in the extreme where $\tau = 0$, labor is equally allocated across space. With symmetric exogenous productivities and amenities, wages are lower in the middle of the line to compensate for the lower price index. Figure IV shows how a border affects the equilibrium population distribution with a positive instantaneous trade cost. As is evident, the larger the border,

17. More general formulations of the exogenous productivity or amenity functions result to more general specifications of the second-order differential equation illustrated before (see Polyanin and Zaitsev 2002, section 2.1.2 for a number of tractable examples).

18. Mossay and Picard (2011) obtain a characterization of the population based on the cosine function in a model where there is no trade but agglomeration of population arises due to social interactions that decline linearly with distance. In their case, population density may be zero in some locations while in our case the CES Armington assumption generates a strong dispersion force that guarantees that the equilibrium is regular when agglomeration forces are not too strong, as discussed in Theorem 2.

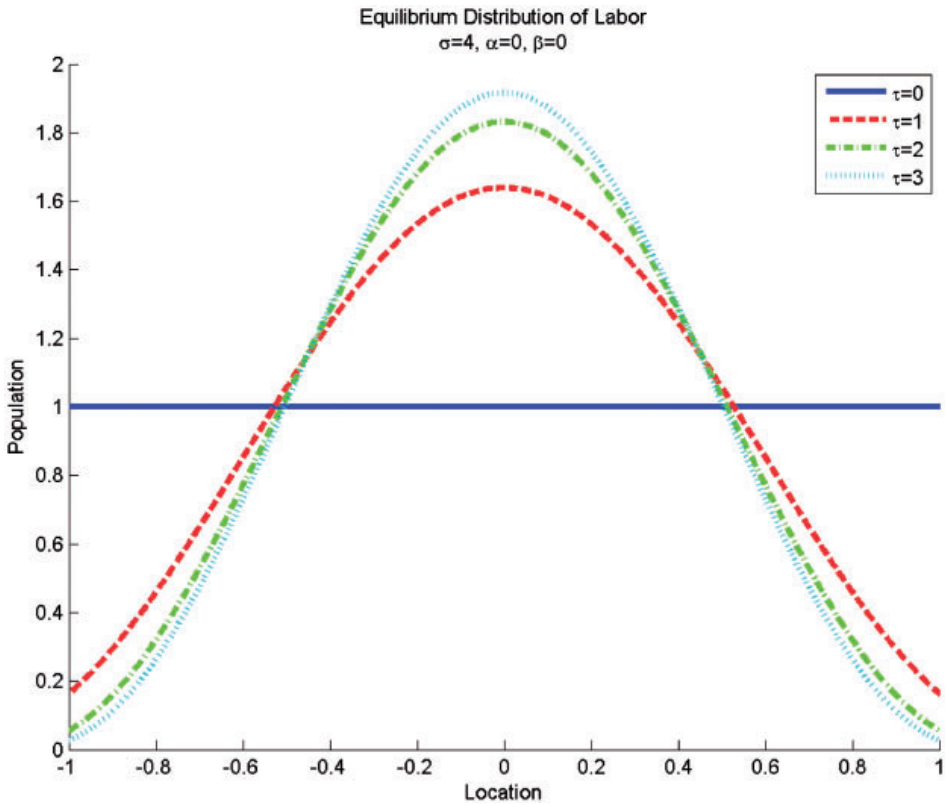


FIGURE III

Economic Activity on a Line: Trade Costs

This figure shows how the equilibrium distribution of population along a line is affected by changes in the trade cost. When trade is costless, the population is equal along the entire line. As trade becomes more costly, the population becomes increasingly concentrated in the center of the line where the consumption bundle is cheapest.

the more economic activity moves toward the middle of each side in the line; in the limit where crossing the border is infinitely costly, it is as if the two line segments existed in isolation.

Differences in exogenous productivities, amenities, and spillovers also play a key role in determining the equilibrium allocation of labor and wages. We use numerical methods to compute these more general cases. Assume, for example, that there are no spillovers, but $\overline{A}(i) = e^{\frac{A}{\sigma-1}i}$. Then the differential equation becomes:

$$\frac{\partial \ln L(i)}{\partial i} = Ai + (1 - 2\sigma)\frac{\partial \ln P(i)}{\partial i},$$

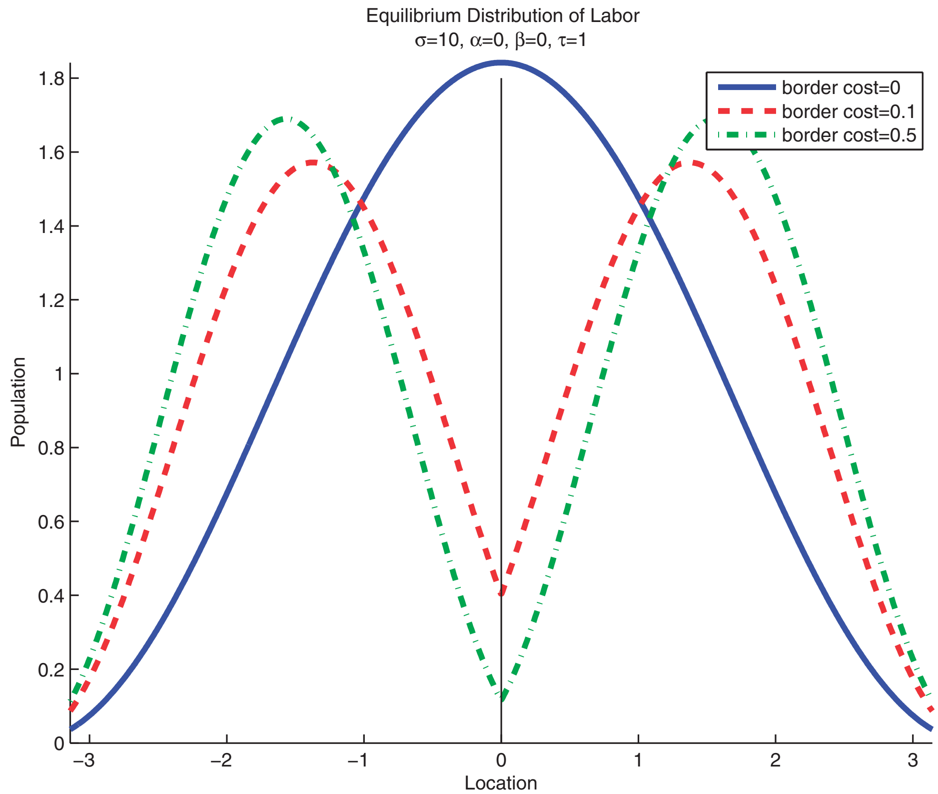


FIGURE IV

Economic Activity on a Line: Border Costs

This figure shows how the equilibrium distribution of population along a line is affected by the presence of a border in the center of the line. As crossing the border becomes increasingly costly, the equilibrium distribution of population moves toward the center of each half of the line.

so that the equilibrium distribution of population is shifted rightward when $A > 0$. Figure V depicts this reallocation of labor toward locations with higher productivity. In this case, it can be shown that an analytical solution of $L(i)$ exists in terms of Bessel functions of the first and the second kind.

A different result is obtained if we increase the parameter α that regulates productivity spillovers, but leave exogenous productivities homogeneous. As mentioned in the previous subsection, as long as $\gamma_1 > 0$, this change increases the elasticity of the labor supply to changes in the geography, which increases the concentration of population in the already highly populated locations. Figure VI depicts the population for higher values of α, and the resulting increase in the concentration. Notice that further

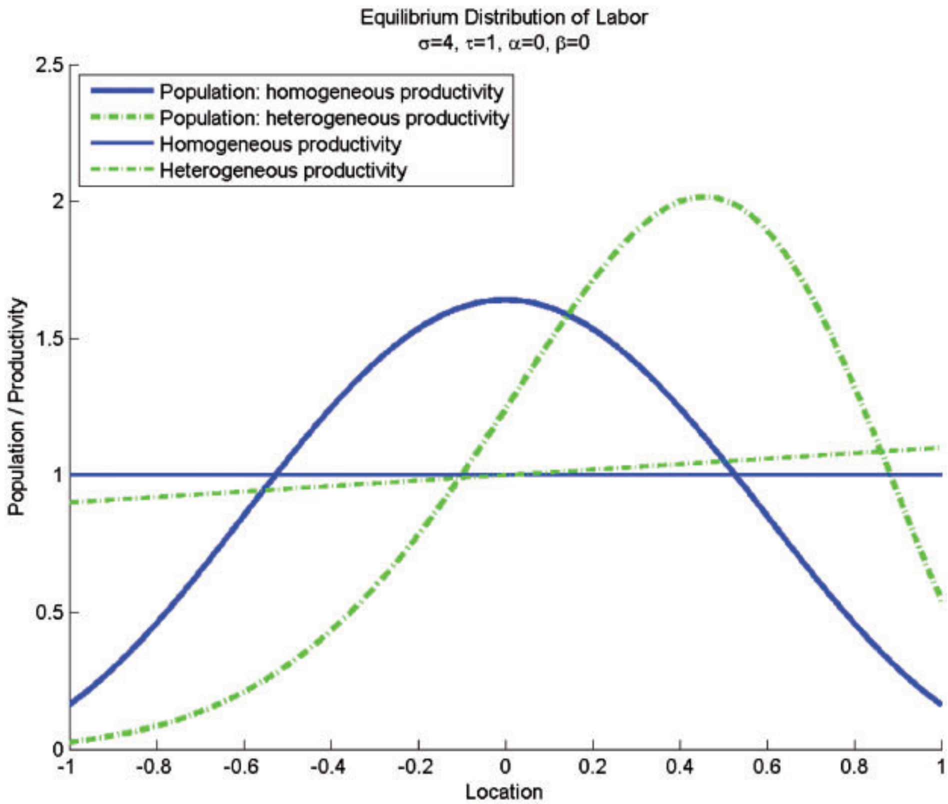


FIGURE V

Economic Activity on a Line: Exogenous Productivity Differences

This figure depicts how the equilibrium distribution of population along a line is affected by exogenous differences in productivity across space. With homogeneous productivities and positive trade costs, the population is concentrated at the center of the line. When productivity is higher toward the right, the population concentrates in regions to the right of the center of the line.

increases in α, to the point that $\gamma_1 < 0$, results in a completely different regular spatial equilibrium where most of the population is concentrated at the two edges of the line. This equilibrium, however, is not locally point-wise stable, as a small number of workers could move from the edges to the center and become better off.

Finally, we can consider what would happen if the instantaneous trade costs depended on the direction of travel. Suppose that the cost of traveling to the right on the line is τ_r and the cost of traveling to the left on the line is τ_l, where $\tau_r \geq \tau_l$. Figure VII illustrates that is it becomes increasingly costly to travel to the

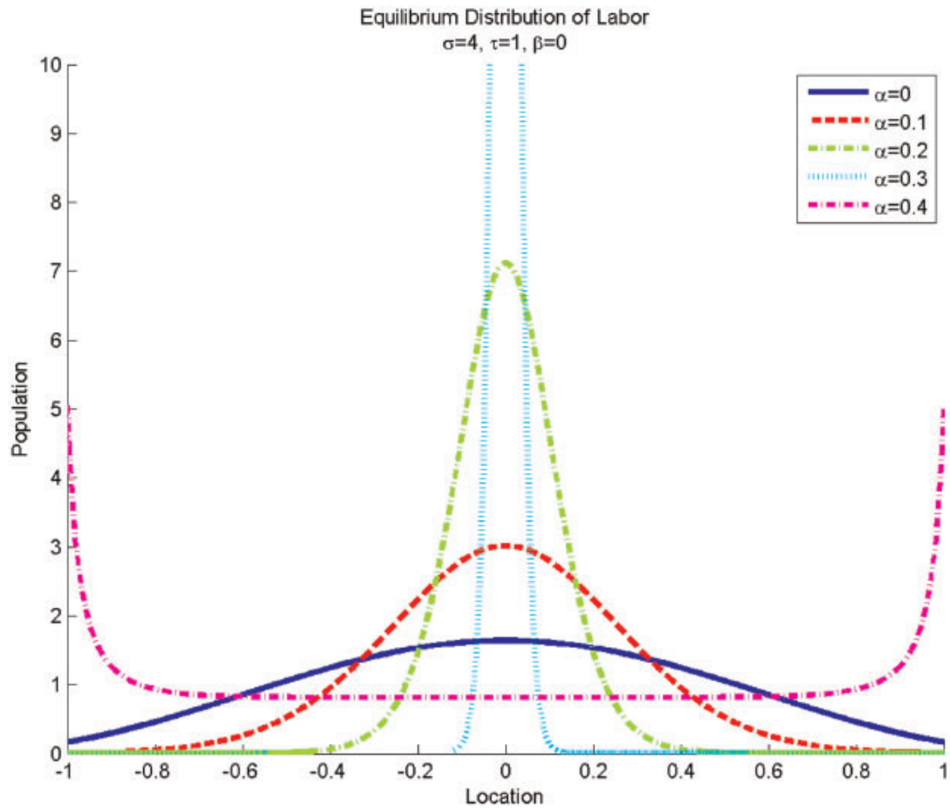


FIGURE VI

Economic Activity on a Line: Productivity Spillovers

This figure shows how the equilibrium distribution of population along a line is affected by varying degrees of productivity spillovers. As the productivity spillovers increase, the population becomes increasingly concentrated in the center of the line. A nondegenerate equilibrium can be maintained as long as $\gamma_1 = 1 - \alpha(\sigma - 1) - \sigma\beta > 0$.

right relative to travel to the left, the equilibrium distribution of the population shifts leftward, where the price index is lowest.

2. The Circle. The example of the circle illustrates the possibility of multiplicity of spatial equilibria. Figure VIII shows the cases $\alpha + \beta = 0$ (left panel) and $\alpha + \beta > 0$ (right panel). When $\alpha + \beta = 0$ there is a unique equilibrium with symmetric population across all locations. Although this remains an equilibrium when $\alpha + \beta > 0$, there are also (a continuum of) additional equilibria, where any location on the circle could be the one where economic activity is more concentrated. Thus, $\gamma_1 = 1$, which corresponds to $\alpha + \beta = 0$, is a bifurcation point that moves us from a

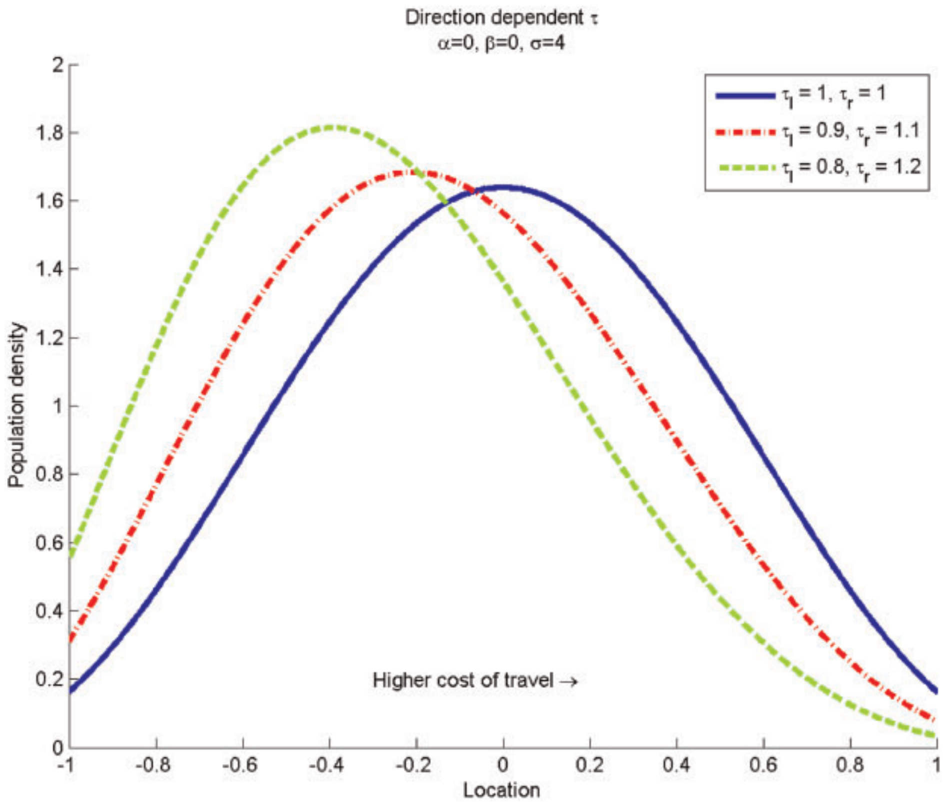


FIGURE VII

Economic Activity on a Line: Direction of Travel

This figure shows how the equilibrium distribution of population along a line is affected by instantaneous trade costs that depend on the direction of travel. As the cost of traveling to the right becomes increasingly more expensive than traveling to the left, the equilibrium distribution of population shifts toward the left.

parameter space with a unique spatial equilibrium to one with a continuum of equilibria. When $\alpha + \beta > 0$ higher trade costs may act as an additional agglomeration force, favoring differentially regions with already concentrated economic activity.[19]

It is possible to obtain a characterization of the equilibrium in a circle when two borders are located into symmetrically opposite points on the circle. Using the methodology of Fabinger (2011) we can obtain an approximation of the solution for the

19. If we further increase $\alpha + \beta$ to the point that the sign of γ_1 turns negative, we know of only one regular spatial equilibrium, which is again the symmetric one. This equilibrium is not point-wise locally stable, as increasing the population of any point in the circle increases the welfare workers living there.

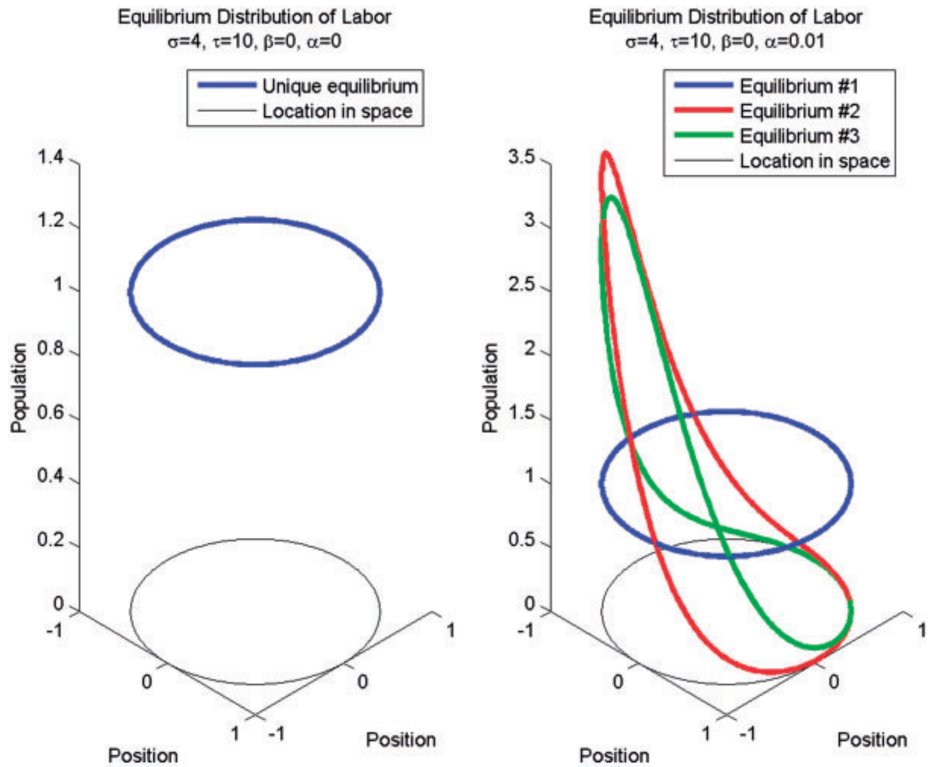


FIGURE VIII

Economic Activity on a Circle: Multiple Equilibria

This figure provides an example of multiple equilibria when the surface is a one-dimensional circle. The left panel shows the unique homogeneous distribution of population along the circle when $\alpha + \beta = 0$. When $\alpha + \beta > 0$ (here $\alpha = 0.01$ and $\beta = 0$), uniqueness is no longer guaranteed. In the case of the circle, there are uncountably many equilibria, each of which has an increased concentration of population around a different point of the circle. The right panel depicts two such equilibria.

population function using Fourier series for small values of the border. As expected, this approximation implies that as the cost of the border increases, population moves away from the border; its details are provided in Online Appendix.

In the line and circle examples, lower values of trade costs lead to larger dispersion of economic activity, as in Helpman (1998). However, in economic geography models such as those of Krugman (1991) and Fujita, Krugman, and Venables (1999), lower values of trade costs lead to a core-periphery structure, effectively increasing the agglomeration of economic activity. Intuitively, in these models with two sectors where one sector features increasing returns to scale and trade costs, a core-periphery structure arises as the result of a home market effect. Mechanically, however, this effect is simply the result of the presence of a second sector with zero trade costs. In the Online Appendix, we incorporate a second sector in our model and show that in the case of a line, increasing trade costs in one sector will reduce the agglomeration of economic activity only if the trade costs in the other sector are sufficiently small.

III. The Topography of the U.S. Economy

In this section, we use the model developed in Section II to analyze the actual topography of economic activity in the continental United States. The section is composed of three parts. First we estimate the underlying geography of the United States. Second we determine the fraction of the observed spatial variation in income due to geographic location. Third we examine the welfare impact and the resulting redistribution of economic activity arising from the construction of the Interstate Highway System. In what follows, we assume the elasticity of substitution $\sigma = 9$, which, consistent with Eaton and Kortum (2002), yields a trade elasticity of 8.[20]

III.A. Determining the Real-World Geography

The goal of this subsection is to recover the underlying geography of the continental United States, namely, the bilateral trade cost function T and the topography of exogenous productivities $\overline{A}$ and amenities $\overline{u}$. To do so, we proceed in two steps. We first estimate trade costs using the observed transportation networks to best match the observed bilateral trade flows between locations. We then find the unique overall productivities A and amenities u that generate the observed distribution of wages and population given the trade costs. Given particular values of α and β, we can then back out the underlying exogenous productivities $\overline{A}$ and amenities $\overline{u}$.

To estimate the underlying geography of the United States, we rely on different types of data which we summarize here; see Online Appendix B for details. The first type of data is the

20. Although this is toward the high end of the accepted range of international trade elasticities, because the elasticity of substitution is between products produced in different locations within a country, it seems reasonable to assume it is higher than the elasticity of substitution of products across countries.

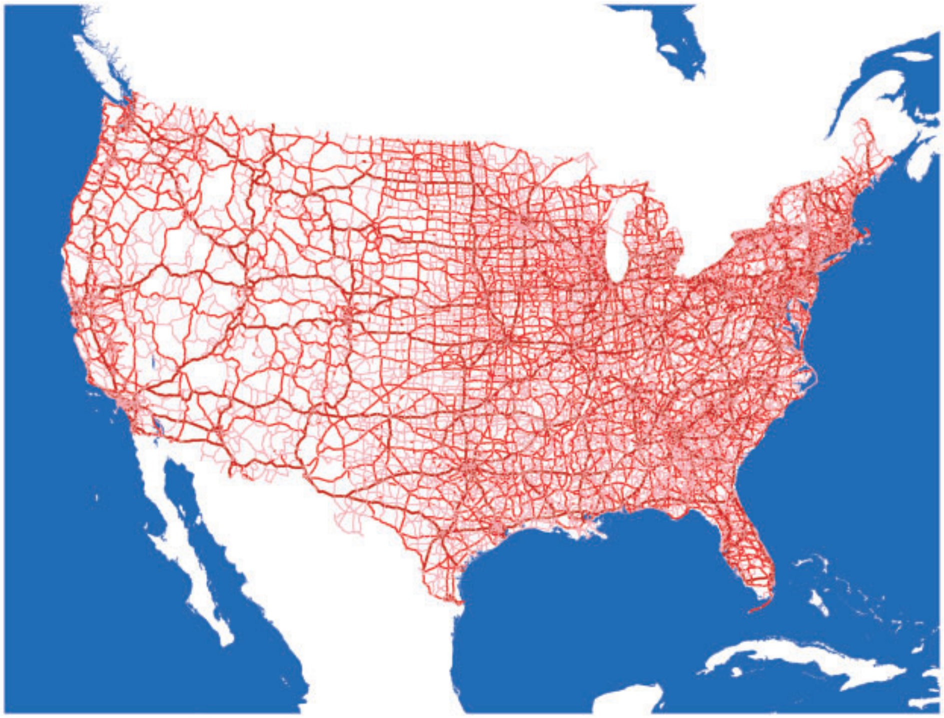
complete highway, rail, and navigable water networks in the United States, which we collect from several sources (NDC 1999; CTA 2003; NHPN 2005). Figure IX depicts the networks; the networks are quite detailed and include the entire U.S. highway system (650,000 km of interstates, other highways, and arterial roads), all railroads in the United States (approximately 225,000 km), and all navigable waterways (approximately 300,000 km). Using GIS software, we project the transportation networks onto a $1{,}032 \times 760$–pixel image of the United States, which we use to construct the mode-specific instantaneous trade cost function.

The second type of data is bilateral trade flow data, which we take from 2007 Commodity Flow Survey (CFS 2007). The CFS is the primary source of within-U.S. domestic freight shipments and the only public source of commodity flow data by U.S. highways. It is collected every five years as a part of the Economic Census and reports the value of trade flows between each CFS area and every other CFS area by each mode of travel.[21] We treat each CFS area as a single location, and assign its location on the image of the United States using the latitude and longitude of its centroid. In what follows, we focus on four modes of travel: road, rail, water, and air. The left panel of Figure X depicts how the share of each mode of travel varies with straight-line distance in the data. The vast majority of trade (in value terms) in the United States is shipped via road; however, this fraction declines as distance increases.

The third type of data is county-level income and demographic characteristics, which we take from the 2000 U.S. Census (MPC 2011b). Figure XII depicts the observed spatial distribution of relative labor and wages. We treat each of the 3,109 counties in the contiguous United States as a distinct location and assign each a location on the image of the United States using the latitude and longitude of their centroid.

A few words are necessary regarding the assumption that each CFS area (in the estimation of trade costs) and each county (in the estimation of overall productivities and amenities) are distinct locations. To calculate an equilibrium, it is necessary to approximate the continuous space with a discrete number of locations. However, there is a trade-off in determining the

21. The CFS micro-data, which is not publicly available, reports establishment level shipment data at the ZIP code level; see Hillberry and Hummels (2008).



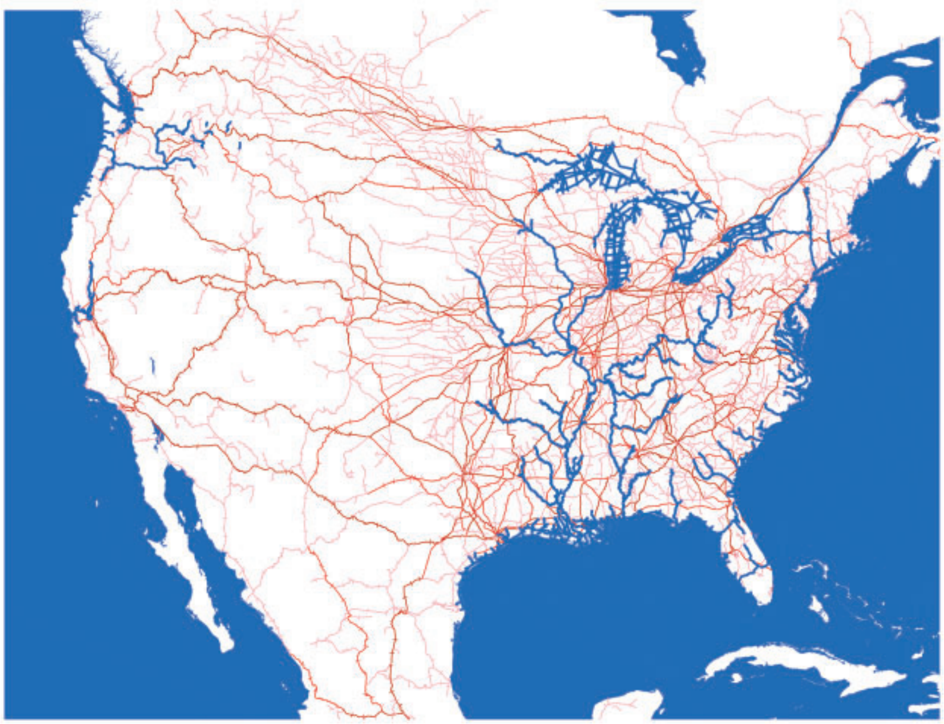


FIGURE IX

U.S. Transportation Networks

(continued)

optimal size of each discrete location. The major advantage of a finer discretization (i.e., more locations) is that the approximation of the continuous space solution improves. There are two disadvantages of a finer discretization. The first is practical: the greater the number of locations, the more computationally intense the problem; the second is conceptual: the smaller each discrete location, the more egregious the assumptions regarding no commuting and no spatial productivity and amenity spillovers become.[22] We feel that treating each county as a distinct location provides a reasonable balance of the two trade-offs.

1. Step 1: Estimating Trade Costs. We first estimate the bilateral trade cost function T. The basic procedure is as follows: for any origin-destination pair, we apply the FMM algorithm to the observed transportation network to get a (normalized) distance between the two locations for each mode of travel (road, rail, water, and air). We then compare these mode-specific distances to the observed mode-specific bilateral trade *shares* using a discrete choice framework to infer the relative geographic trade cost of each mode of travel. Given the structure of the discrete choice framework, we can combine these estimates to determine the total geographic trade cost up to scale. Finally, we estimate the scale using the observed bilateral trade levels and the gravity equation implied by the model. The last step has the advantage of allowing us to incorporate proxies for nongeographic trade costs.

We begin by determining the normalized mode-specific distance between all locations in the United States. Using the detailed transportation networks data detailed already, we create

FIGURE IX Continued

The top panel of the figure shows the U.S. road network; interstate highways are black (dark red in the online version), other U.S. highways are dark gray (red online), and arterial roads are light gray (light red online). The bottom panel of the figure shows the U.S. rail and water network. Railroads are indicated by hatched lines where the color indicates the importance of the line: class A railroads are black (dark red online), class B railroads dark gray (red online), and other railroads are light gray (light red online). Navigable inland waterways are indicated by black lines (blue online).

22. In the Online Appendix, we extend the model to allow for commuting and find similar estimated amenities and productivities.

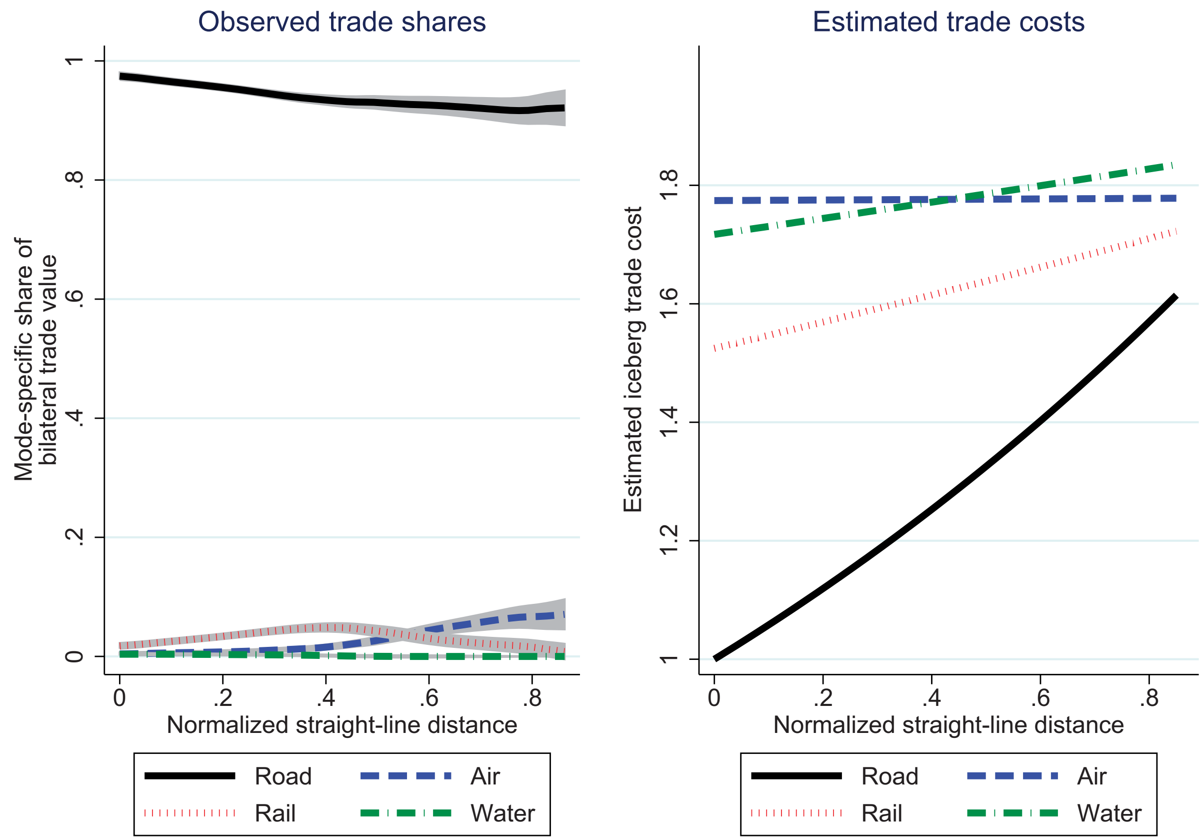


FIGURE X

Mode-Specific Bilateral Trade Shares by Distance

This figure shows the relationship between mode specific trade flows and distance. The left panel shows how the share of bilateral trade (measured in value) by each mode of transport varies with the straight-line distance between the origin and destination. Each line is a nonparametric local mean smoothed regression using an Epanechnikov kernel with a bandwidth of 0.1. 99% confidence intervals are reported in gray. The right panel shows how the estimated trade costs for each mode of transportation vary with distance. In both panels, distance is normalized so that the width of the United States has distance of 1.

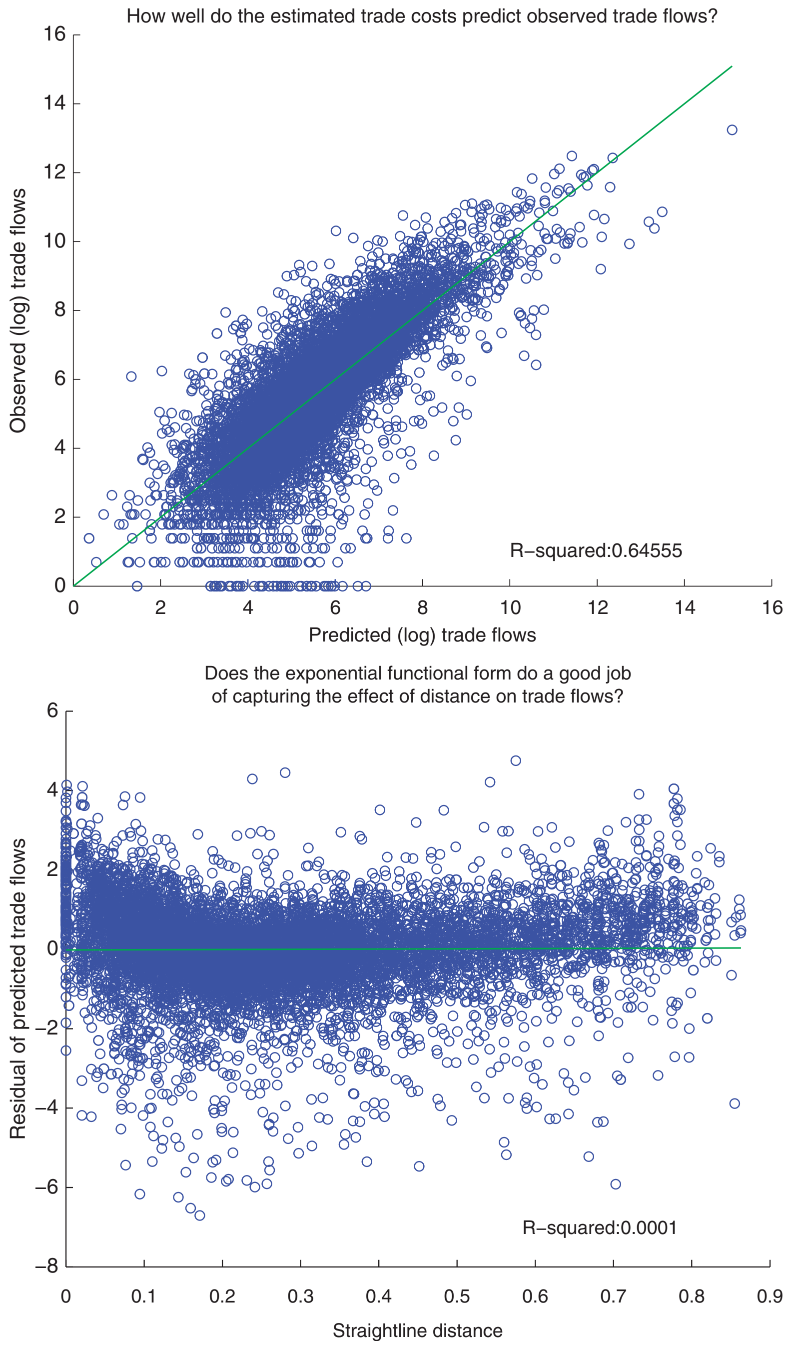


FIGURE XI

Assessing the Predicted Trade Costs

This figure assesses the quality of the estimated trade costs. The top panel compares the bilateral trade flows implied by the estimated trade costs with the bilateral trade flows observed in the 2007 Commodity Flow Survey. The bottom panel shows that the difference between the trade flows implied by the estimated trade costs and the observed bilateral trade flows (i.e., the residuals) are uncorrelated with straight-line distance.

an instantaneous cost function $\tau_m : S \rightarrow \mathbb{R}_{++}$, where locations i on the network are assigned a low value of τ_m and locations off the network are assigned a high value τ_m (see Online Appendix B.3 for details). For any origin $i \in S$ and destination $j \in S$ and mode $m \in M$, we can apply the FMM algorithm using τ_m to determine the normalized mode-specific distance $d_m(i,j)$. We normalize the scale of distance so that the cost of traveling the width of the United States would be one if there existed a straight-line route via a particular network. We estimate the relative costs of trade across different modes of transport later.

Before proceeding, it is informative to note that simple reduced-form regressions show that the normalized mode-specific distances $d_m(i,j)$ do indeed appear to be capturing the cost of traveling via different modes of travel. Table I reports the results of regressions of the mode-specific value of bilateral trade flows on the normalized mode-specific distances, conditional on origin and destination fixed effects. The log value of road shipments is strongly negatively correlated with the log road distance (column (1)), and remains so even conditional on straight-line distance (column (2)). Conditional on road distance, there is no statistically significant relationship between road shipments and rail distance, whereas increases in water distance are actually associated with greater shipments via road (column (3)), suggesting that traders substitute across modes of transport. Similar patterns are present for shipments via rail (columns (4)–(6)) and water (columns (7)–(9)), although the results are not as statistically significant, possibly because there are fewer observations and the different measures of distance are highly correlated.[23]

We next determine the relative cost of trade across different modes of transport using a discrete choice framework. We should emphasize that the discrete choice framework is entirely distinct from the economic geography model developed in Section II and is used only as a tool to estimate trade costs based on mode-specific trade shares. While it would be possible to estimate travel cost parameters using variation in bilateral trade levels across origins and destinations without using a discrete choice framework, such a procedure would be subject to concerns about the endogeneity of the location of transportation networks (e.g., there exists a

23. The results are similar if we constrain our analysis to only trade between metropolitan statistical areas rather than all CFS areas.

TABLE I

Commodity Flows and Mode-Specific Shipping Distances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable:		Road shipments			Rail shipments			Water shipments	
Log road distance	−1.369***	−0.945***	−1.362***			−0.061			0.371
	(0.015)	(0.160)	(0.141)			(0.506)			(3.028)
Log rail distance			−0.083	−0.457***	−0.382	−0.296			−0.622
			(0.147)	(0.056)	(0.421)	(0.515)			(3.067)
Log water distance			0.083***			−0.109	−0.730*	−0.444	−0.498
			(0.032)			(0.100)	(0.349)	(0.754)	(0.911)
Log straightline distance		−0.407***			−0.069			−0.313	
		(0.154)			(0.388)			(0.718)	
Origin fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.488	0.489	0.489	0.051	0.051	0.052	0.384	0.403	0.397
Observations	9,177	9,177	9,177	1,434	1,434	1,434	58	58	58

Notes. Ordinary least squares. Each observation is the observed (log) value traded from a CFS region to another CFS region in 2007 by a particular mode of transport. Standard errors are reported in parentheses. Stars indicate statistical significance: $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

highway between Chicago and New York because the two cities trade a large amount with each other). In contrast, the discrete choice framework provides a method of estimating travel cost parameters using mode-specific trade *shares* between a given origin and destination (e.g., what fraction of trade between Chicago and New York occurs via rail rather than road). This procedure effectively controls for the overall level of bilateral trade flows, mitigating endogeneity concerns in much the same way that an origin-destination fixed effect does in a linear regression. However, we should emphasize that there still exist endogeneity concerns based on the relative mode-specific trade shares (e.g., there exists a railroad between Chicago and New York because the two cities tend to trade more goods that are best shipped via rail).

Suppose for every origin $i \in S$ and destination $j \in S$ there exists a mass of identical traders who choose a particular mode of transport in order to minimize the trade costs incurred from shipping a unit amount from i to j. Suppose there are $m \in \{1, \ldots, M\}$ modes of transport and the iceberg cost of trader t shipping goods from i to j using mode m is $\exp(\tau_m d_m(i,j) + f_m + \nu_{tm})$, where τ_m is the mode-specific variable cost, f_m is the mode-specific that is fixed with respect to distance, and ν_{tm} is a trader-mode specific idiosyncratic cost.[24] Finally, suppose that ν_{tm} is distributed i.i.d. across traders and modes of transportation with a Gumbel distribution with shape parameter θ, that is, $\Pr\{\nu \leq x\} = e^{-e^{-\theta x}}$.[25] (Note that this implies $\Pr\{e^{\nu} \leq x\} = e^{-x^{-\theta}}$, i.e., e^{ν} is distributed according to a Fréchet distribution with shape parameter θ.)

Let $\pi_m(i,j)$ denote the fraction of trade shipped from i to j using mode of transportation m. Given the distribution assumption of ν_{tm}, it is straightforward to show that:

$$\pi_m(i,j) = \frac{\exp(-a_m d_m(i,j) - b_m)}{\sum_k (\exp(-a_k d_k(i,j) - b_k))}, \tag{21}$$

24. While the introduction of a fixed cost violates the continuity assumption of Section II, this is not a practical concern here because we consider only a discrete number of locations.

25. Our discrete choice framework bears a resemblance to the one presented in Lux (2011); in that framework, there is a continuum of goods, where each good had an idiosyncratic mode-specific transportation costs; here, there is a single good but a continuum of traders and each trader is assumed to have an idiosyncratic mode-specific transportation cost.

where $a_m \equiv \theta\tau_m$ and $b_m \equiv \theta f_m$. Given mode-specific distances $\{d_m\} : M \times S \times S \to \mathbb{R}_+$, we can estimate $\{a_m\}$ and $\{b_m\}$ using equation (21) by choosing $\{a_m\}$ and $\{b_m\}$ such that the predicted mode-specific shares of bilateral trade most closely match the observed mode-specific trade shares. As is standard in discrete choice estimation, mode-specific trade shares are invariant to a multiplicative shifter on the trade costs. To pin down the relative scale, we assume that traders do not incur a fixed cost of traveling via road. We then estimate $\{a_m\}$ and $\{b_m\}$ from equation (21) using a nonlinear least squares routine.

Given our estimates of $\{a_m\}$ and $\{b_m\}$, we can estimate total bilateral trade costs using the observed level of bilateral trade flows. From the discrete choice framework, the average geographic trade cost incurred in trading from i to j, $T_g(i,j)$, is:

$$T_g(i,j) = \frac{1}{\theta}\Gamma\left(\frac{1}{\theta}\right)\left(\sum_m (\exp(-a_m d_m(i,j) - b_m))\right)^{-\frac{1}{\theta}}. \tag{22}$$

Suppose that total trade costs T are a composite function of geographic trade costs T_g and nongeographic trade costs T_{ng}, where the latter can be approximated by a vector of nongeographic bilateral observables $\mathbf{C}(i,j)$, for example, similarity in language and ethnicities.[26] Taking logs of equation (3) and substituting in the functional forms of T_g and T_{ng} yields the following gravity equation:

$$\begin{aligned} \ln X_{ij} = {} & \frac{\sigma-1}{\theta}\ln\sum_m \left(\exp\left(-\hat{a}_m d_{mij} - \hat{b}_m\right)\right) \\ & + (1-\sigma)\beta' \ln \mathbf{C}_{ij} + \delta_i + \delta_j + \varepsilon_{ij}, \end{aligned} \tag{23}$$

where the notation indicates that that we observe a finite number of bilateral trade flows. Hence, given an elasticity of trade σ, we can estimate θ (which thereby determines $\{\tau_m\}$ and $\{f_m\}$) and β (which determines the nongeographic trade costs). Note that varying the elasticity of trade will simply scale the estimates accordingly; this is the well-known result (see, e.g., Anderson and Van Wincoop 2003)

26. While we follow the large gravity literature in interpreting the coefficient on these bilateral nongeographic trade costs as causal, we acknowledge that these variables may be affected by bilateral trade flows.

that observed trade flows are matched equally well with a high trade elasticity and a low level of trade costs or vice versa.

Table II reports estimated values of the mode-specific variable and fixed trade costs $\{\tau_m\}$ and $\{f_m\}$, the estimated shape parameter θ, and the effect of each nongeographic observable on trade costs. Because the estimation procedure is a multiple-stage process, we calculate bootstrapped standard errors derived from redoing the entire estimation procedure 1,000 times. We do the estimation for trade flows between all CFS areas as well as trade flows only between metropolitan statistical areas (MSAs) (or subsets thereof). Although the former sample has more observations, the latter sample corresponds more closely with our theoretical conception of a location; reassuringly, the results in the two samples are very similar.

The right panel of Figure X depicts how the estimated mode-specific costs vary with distance. Given that the vast majority of trade occurs over roads, it is not surprising that travel via roads is always estimated to have the lowest cost, regardless of distance. As distance between origin and destination increases, however, the cost of travel via air, water, and rail decline relative to travel via road, which is consistent with the declining share of trade occurring via road with distance. Overall, the magnitude of the trade costs is roughly consistent with estimates of domestic trade costs in the literature (e.g., Anderson and Van Wincoop 2004 estimate an iceberg trade cost of 55% for domestic distribution costs in a representative rich country). The estimated nongeographic trade costs also appear reasonable. Trade costs are estimated to be approximately 30 percentage points lower when the origin and destination are in the same state, and a 10% increase in the ethnic similarity between an origin and destination is associated with a 9% decline in trade costs. Somewhat surprisingly, trade costs are estimated to increase with the similarity in languages between the origin and destination, although this effect is not statistically significant when the sample only includes MSAs.

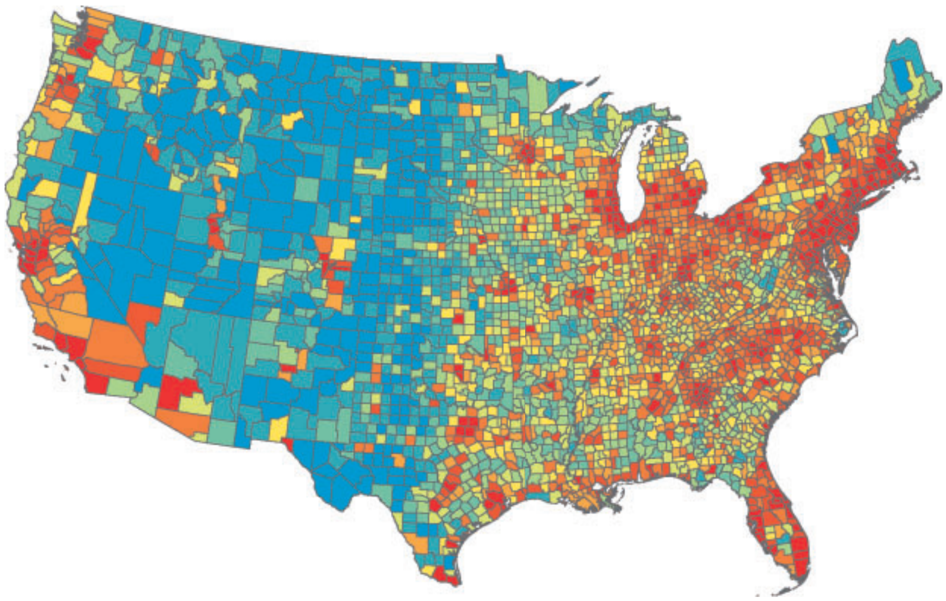
How well do the estimated trade costs predict trade flows? The top panel of Figure XI compares the bilateral trade flows predicted by the estimated trade costs to those observed in the CFS. Overall, the predicted trade flows can explain 65% of the observed variation in trade flows, and there does not appear to be any systematic bias in the estimates with the observed volume of trade shares. The bottom panel of Figure XI shows that there is

TABLE II

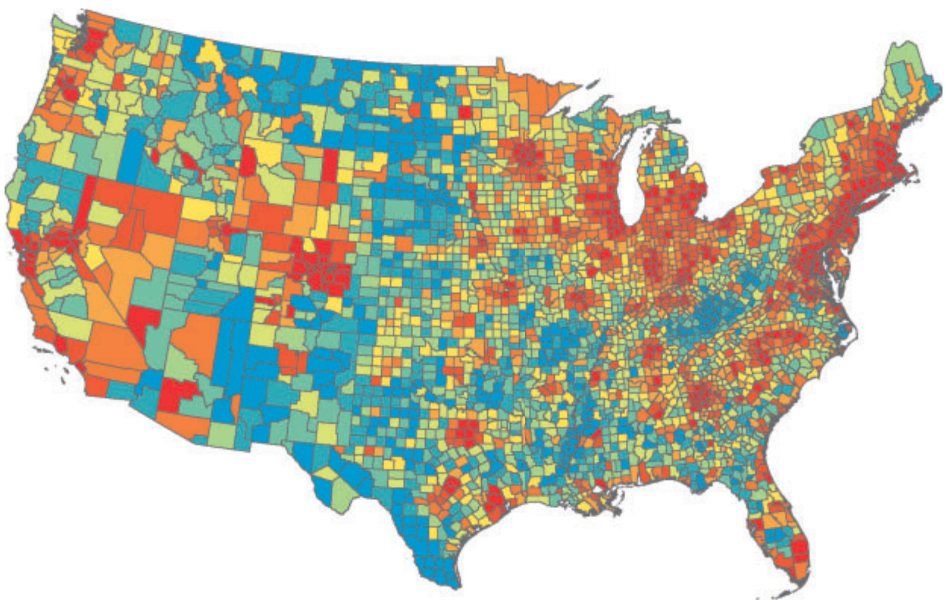
Estimated Mode-Specific Relative Cost of Travel

	All CFS areas				Only MSAs			
	Road	Rail	Water	Air	Road	Rail	Water	Air
Geographic trade costs								
Variable cost	0.5636***	0.1434***	0.0779***	0.0026	0.4542***	0.1156***	0.0628***	0.0021
	(0.0120)	(0.0063)	(0.0199)	(0.0085)	(0.0233)	(0.0210)	(0.0265)	(0.0176)
Fixed cost	0	0.4219***	0.5407***	0.5734***	0	0.34***	0.4358***	0.4621***
	N/A	(0.0097)	(0.0236)	(0.0129)	N/A	(0.0235)	(0.0375)	(0.0264)
Estimated shape			14.225***				17.6509***	
parameter (θ)			(0.3375)				(1.4194)	
Nongeographic trade costs								
Similar ethnicity			−0.0888***				−0.0803***	
			(0.0153)				(0.0275)	
Similar language			0.063***				0.0286	
			(0.0223)				(0.0359)	
Similar migrants			−0.0191				−0.0135	
			(0.0119)				(0.0203)	
Same state			−0.2984***				−0.3104***	
			(0.0101)				(0.0176)	
R-squared (within)			0.4487				0.4113	
R-squared (overall)			0.6456				0.5995	
Observations with positive bilateral flows	9,601	9,601	9,601	9,601	3,266	3,266	3,266	3,266
Observations with positive mode-specific bilateral flows	9,311	1,499	78	1,016	3,144	340	26	471

Notes. This table shows the estimated cost of traveling via different modes of travel. Relative costs are estimated using the mode-specific shares of total bilateral trade values from the 2007 Commodity Flow Survey. The mode specific bilateral cost is exp(*variablecost* * *distance* + *fixedcost*), where the distance of the width of the country is normalized to 1. The shape parameter (θ) is estimated from a gravity regression and pins down the scale of the variable and fixed costs (see the text for details). Similarity in ethnicity, language, and migrants is measured by the correlation across respective census categories between the counties nearest to the origin and destination CFS area. Boostrapped standard errors (from 1,000 repetitions) reported in parentheses.



Population density



Wages

FIGURE XII

U.S. Population Density and Wages in 2000

This figure shows the relative population density (top) and wages (bottom) within the United States in 2000 by decile. The data are reported at the county level; darker shading indicates higher deciles.

Source: MPC (2011a).

not any systematic relationship between the difference from the predicted to the actual trade flows and distance, suggesting that the assumed exponential relationship between trade costs and distance is a reasonable approximation.

2. Step 2: Identifying Productivities and Amenities. Suppose we observe trade costs and the equilibrium distribution of economic activity. Can we identify the underlying topography of overall productivities and amenities? The following theorem guarantees that for any observed distribution of economic activity and given any trade costs, there exists a unique topography of overall productivities and amenities that generate that equilibrium. Because the structure of the model allows the overall productivities and amenities to be recovered for any set of trade costs, the theorem justifies the sequential estimation strategy we employ (where we first estimate trade costs and then estimate productivities and amenities).

THEOREM 3. *For any continuous functions w and L and continuous symmetric function T, all bounded above and below by strictly positive numbers, there exists unique (to-scale) positive and continuous functions A and u such that w and L comprise the regular spatial equilibrium for the geography defined by T, $\overline{A} = AL^{-\alpha}$ and $\overline{u} = uL^{-\beta}$.*

Proof. See Online Appendix A.1.5. ■

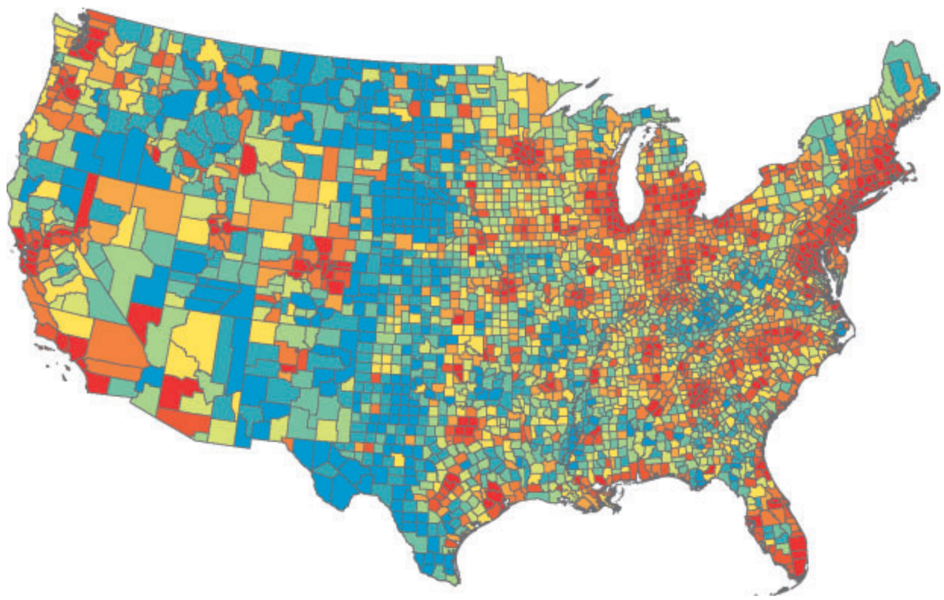
It is important to note that Theorem 3 does not rely on the assumed relationships governing spillovers in equations (1) and (2); hence, the theorem applies for any strength or source of spillovers, including, for example, spillovers that occur across space. In general, if the relationship between the strength of spillovers and the population distribution is known, then because the distribution of labor is observed, the underlying productivities and amenities can be determined by inverting the relationship given A and u. Given our assumed functional form of spillovers, $\overline{A}$ and $\overline{u}$ can be identified given L, α and β using equations (1) and (2). The converse of this is that the strength of spillovers (in our case, α and β) cannot be identified from the observed cross-sectional distribution of wages and population: for any α and β, unique functions $\overline{A}$ and $\overline{u}$ can be chosen to generate the composite productivities necessary to generate the observed equilibrium distribution of economic activity.[27]

Intuitively, the identification of composite productivities and amenities from the observed distribution of population and welfare works as follows. Consider two points i and $j \in S$ with the same geographic locations, that is $T(i,s) = T(j,s)$ for all $s \in S$. Because the two points have the same geographic location, they share the same price index, which implies the (observed) ratio of their nominal wages is equal to the ratio of their real wages. Because welfare is the same in both locations, it must be the case that any difference in relative real wages must be fully compensated by differences in amenities; hence the relative amenities are simply the inverse of the relative wages. Similarly, because the two locations have the same geographic location, differences in demand for their produce arises only because of differences in their marginal costs of production, which depends only on wages and productivity. From market clearing, income is equal to the total quantity sold, so the relative productivity of the two locations can be inferred by comparing the total income and wages in each location.[28] Using the structure of the model allows us to extend this intuition to allow for differences in trade costs across locations.

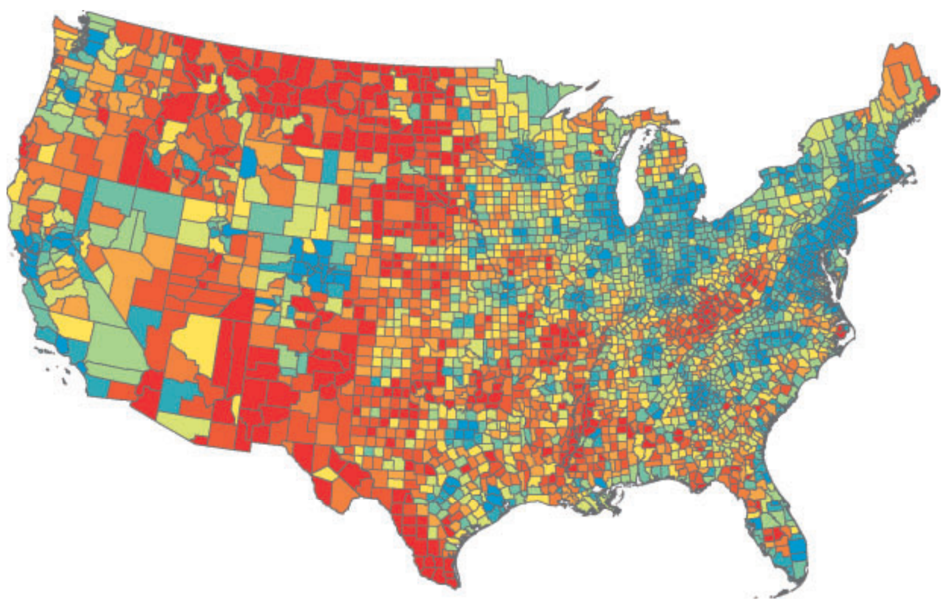
Using the structure of the model and the bilateral trade costs estimated in the previous section, we identify the unique composite amenities and productivities of each U.S. county in 2000 that are consistent with the observed distribution of labor and wages from the 2000 U.S. Census. We should emphasize that because the trade costs are estimated in the first stage, the amenities and productivities should be interpreted as estimates themselves. Figure XIII depicts the unique distribution of *composite* amenities and productivities that are consistent with the estimated trade costs and the observed distribution of labor and population. Composite amenities are much lower in more populated counties,

27. Ellison and Glaeser (1997) make a similar point about the inability to disentangle the natural advantage of a location from spillovers using cross-sectional data alone.

28. In particular, it is straightforward to show that market clearing implies $\frac{A(i)}{A(j)} = \left(\frac{(L(i)w(i)^{\sigma})}{(L(j)w(j)^{\sigma})}\right)^{\frac{1}{\sigma-1}}$.



Composite productivity



Composite amenity

FIGURE XIII

Estimated Composite Productivity and Amenity

This figure shows the estimated composite productivity (top) and amenity (bottom) by decile. The data are reported at the county level; darker shading indicates higher deciles.

while composite productivities are much higher. Figure XIV depicts the resulting exogenous amenities and productivities when $\alpha = 0.1$ and $\beta = -0.3$ (values that are roughly consistent with the estimates of productivity spillovers from Rosenthal and Strange 2004 and the share of income spent on housing, BLS 2000).[29] The topography of exogenous productivities and amenities seem reasonable; amenities in southern Florida, southern California, and Arizona are high, whereas amenities in the central of the United States are low; productivities are highest along the Eastern Seaboard and in the upper Midwest and low in places like Montana, Nebraska, and west Texas. Note that there is only a weak positive correlation (0.12) between exogenous amenities and productivities.

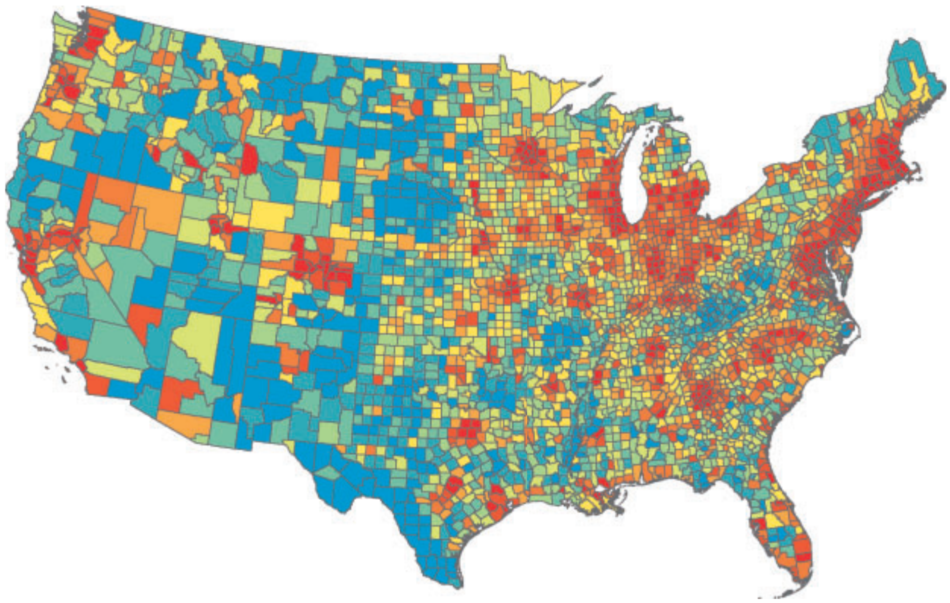
III.B. Importance of Geographic Location

Given the estimated geography of the United States, we can determine the fraction of the observed variation in incomes $Y(i) \equiv w(i)L(i)$ that is due to the geographic location of $i \in S$. To do so, note that combining equations (15) and (16) yields the following expression:

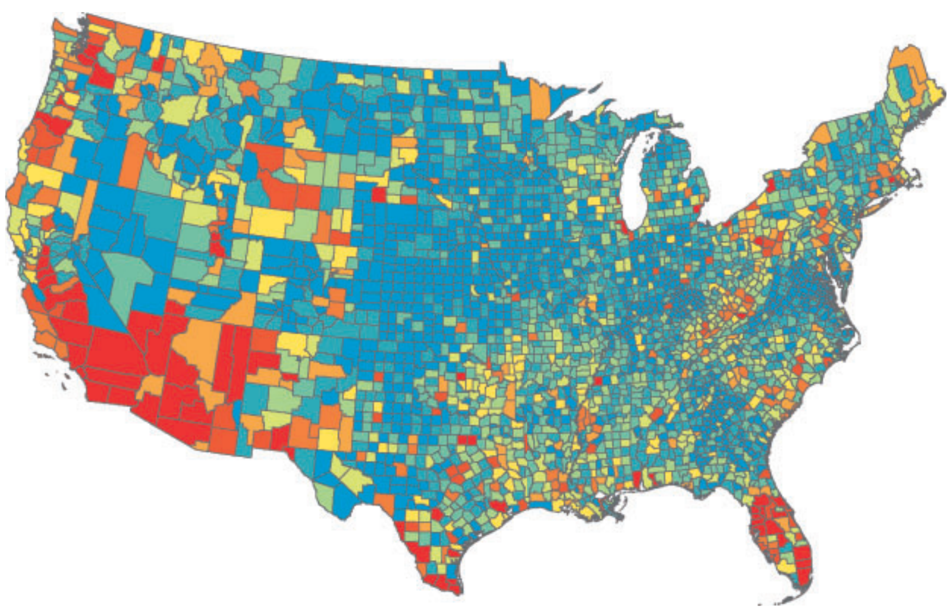
$$\frac{\gamma_1}{\sigma - 1} \ln Y(i) = \frac{C_w + C_L}{\sigma - 1} + (1 - \beta)\ln \overline{A}(i) + (1 + \alpha)\ln \overline{u}(i) - (2 + \alpha - \beta)\ln P(i). \tag{24}$$

Equation (24) provides a log linear relationship between the observed income in location i, the exogenous productivities and amenities, and the price index. As reduced-form evidence that geographic location matters for the distribution of income in the United States, Figure XV depicts the geographic variation in the

29. Rosenthal and Strange (2004) summarize estimates for the increase of productivity when population doubles of around 3–8%. We chose a roughly higher spillover term of $\alpha = 0.1$ since our model is a perfect competition model that ignores the effects of entry on overall output, but as already discussed, these additional spillovers map directly to a higher parameter α. Depending on whether one includes "house furnishings and equipment" and "household operations" with "shelter," the BLS (2000) reports that 18.7% to 24.7% of household expenditure is spent on housing (18.7% when "shelter" and half of "house furnishings and equipment" is included, 24.7% when all three categories are included). Given our isomorphism that implies $\beta = \frac{\delta}{1-\delta}$, where δ is the share of expenditure spent on housing, the range of relevant parameters for β is 0.23 to 0.325; we choose a parameter of $\beta = -0.3$ for our baseline experiment.



Exogenous productivity



Exogenous amenity

FIGURE XIV

Estimated Exogenous Productivity and Amenity

This figure shows the estimated exogenous productivity $\overline{A}$ (top) and amenity $\overline{u}$ (bottom) by decile assuming $\alpha = 0.1$ and $\beta = -0.3$. The data are reported at the county level; darker shading indicates higher deciles.

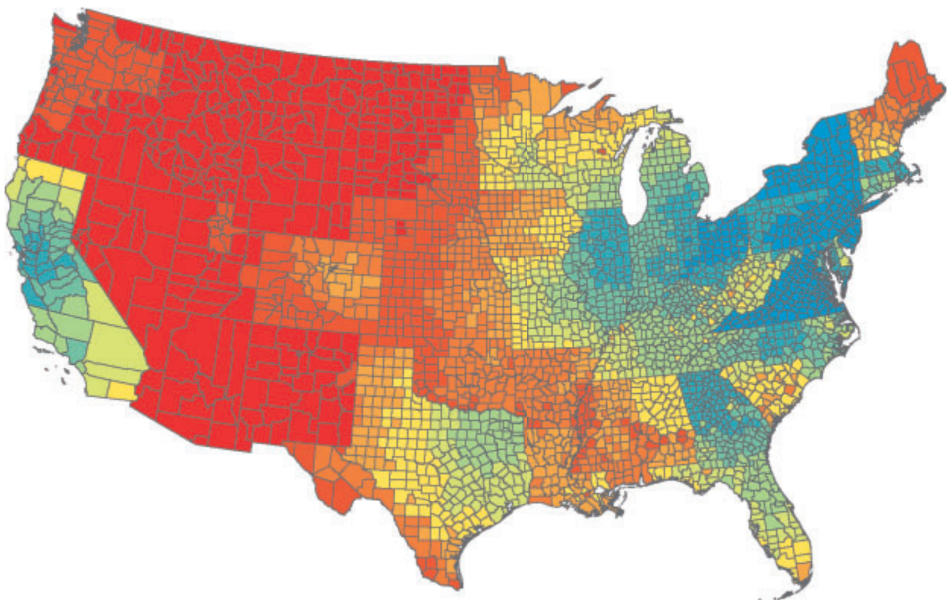


FIGURE XV

Estimated Price Index

This figure shows the estimated price index by decile. The data are reported at the county level; darker shading indicates higher deciles.

estimated price index. At first glance, a comparison of Figure XV with Figure XII suggests that counties with better geographic location (i.e., lower price indices) appear to be wealthier.

To determine the relative contribution of the effect of local characteristics (i.e., $\overline{A}(i)$ and $\overline{u}(i)$) and geographic location (i.e., $P(i)$) to the spatial dispersion of income, we apply a Shapley decomposition (see Shorrocks 2013) to equation (24). The Shapley decomposition determines the expected marginal contribution of the local characteristics and the geographic location to the total variation in observed incomes; intuitively, it provides a way of assigning what fraction of the R^2 of a regression is due to each set of explanatory variables. Because we do not observe the strength of spillovers (i.e., α and β), but they are necessary to determine amenities and productivities, we report the results of the decomposition for all combinations of $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$.

It should be noted that if the trade cost function is misspecified (and hence the price index $\ln P(i)$ is measured with error), the model would erroneously rely on amenity and productivity differences to explain observed differences in incomes, thereby biasing downward the estimated contribution of the price index. On the other hand, because the price index is calculated from the bilateral trade costs, which themselves depend on transportation networks, the geographic location may in reality depend on the economic characteristics of a location (e.g., more productive locations may be more likely to be connected via road networks).

Figure XVI reports the fraction of the spatial variation in income in the United States in 2000 that can be attributed to geographic location rather than local characteristics. While the exact value of the decomposition depends on the strength of spillovers, the decomposition suggests that at least 20% of the observed spatial variation in income is due to geographic location, and geographic location may be responsible for upward of 70% of the observed variation in income (if the spillovers are such that $\alpha = 0.23$ and $\beta = -0.14$). Hence, the results suggest that a substantial fraction of the spatial variation in incomes across the United States can be explained by variation in trade costs due to geographic location.

III.C. The Effects of the Interstate Highway System

Given the estimated geography of the United States, we can also examine how changes to the geography affect the equilibrium spatial distribution of population and wages and overall welfare. This section provides an illustrative example of such counterfactual analysis by examining what would happen if the Interstate Highway System (IHS) were removed.

The counterfactual procedure is straightforward. We first recalculate the bilateral trade cost function T using the estimates from Section III.A.1 supposing that there were no interstate highways, but keeping all other modes of transportation (including other national highways and arterial roads) unchanged. Since the counterfactual trade costs are based off of estimated parameters, the following results should also be interpreted as estimates. For a given strength of spillovers α and β, we hold fixed the exogenous productivities $\overline{A}$ and amenities $\overline{u}$ at the values estimated in Section III.A.1 and recalculate the equilibrium distribution of labor, wages, and the overall welfare level under these alternative trade costs using equations (12) and (13). Because the effect of removing the IHS will depend on the strength of spillovers, we do the counterfactual for many

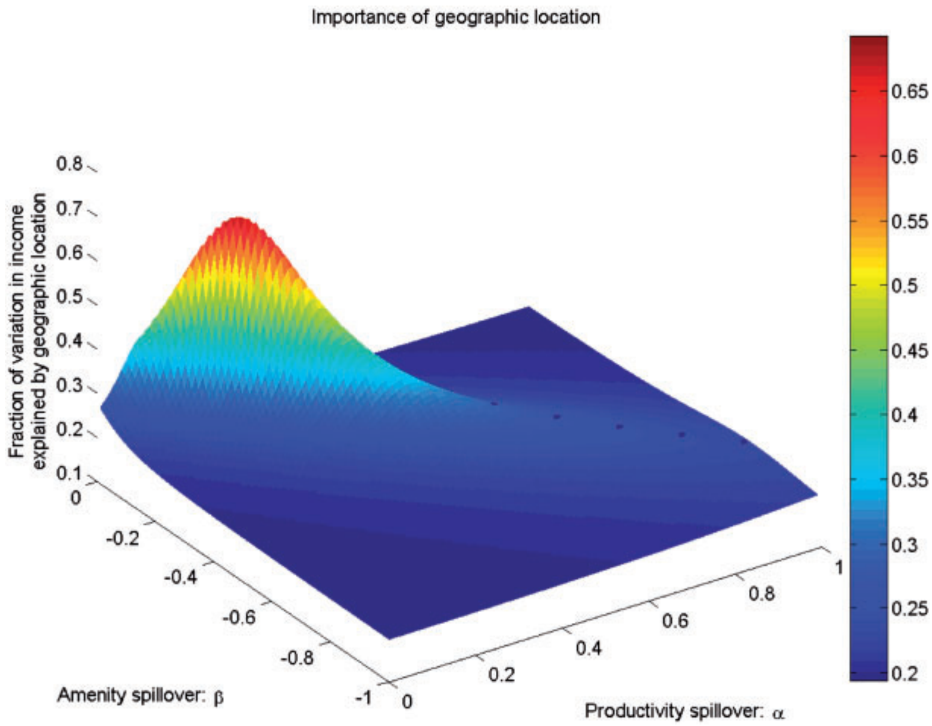


FIGURE XVI

Fraction of Spatial Inequality of Income Due to Geographic Location in the United States

This figure shows the fraction of the observed variation in income across space in the United States in 2000 that is due to geographic location. The decomposition is calculated for all constellations of productivity spillover strength $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$.

combinations of $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$ such that $\alpha + \beta \leq 0$, a restriction that from Theorem 2 guarantees the uniqueness of equilibrium.

To illustrate the effect of the removal of the IHS on trade costs, Figure XVII presents the relative change in the price index (holding wages, population, and productivities fixed at observed levels). As is evident, the price index rose the most in the Rocky Mountains, indicating that locations there saw the greatest increase in economic remoteness, whereas the price index in California and the Eastern Seaboard increased by less. There are two reasons for these differences: first, locations in California and the Eastern Seaboard had better alternative modes of transportation (see Figure IX); additionally, locations

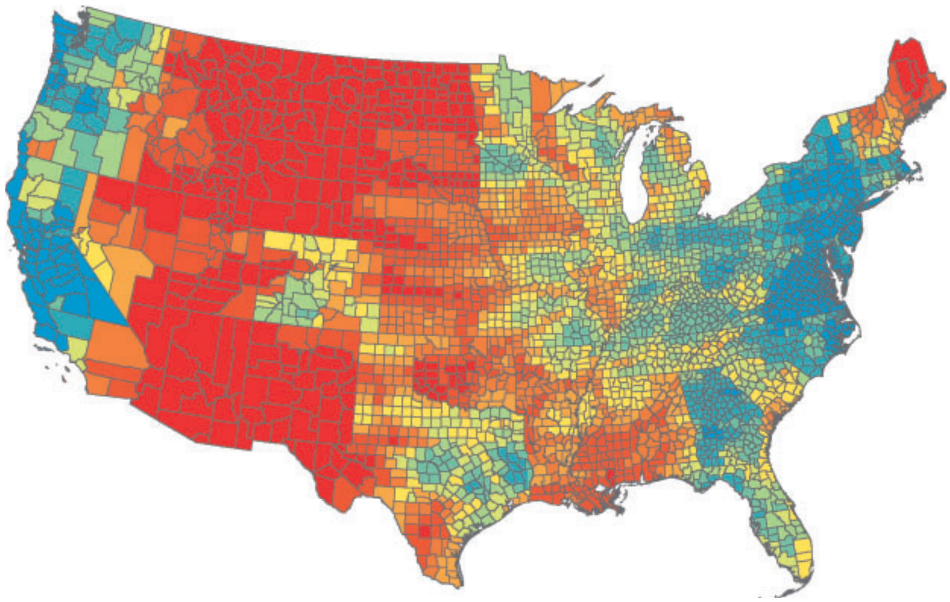
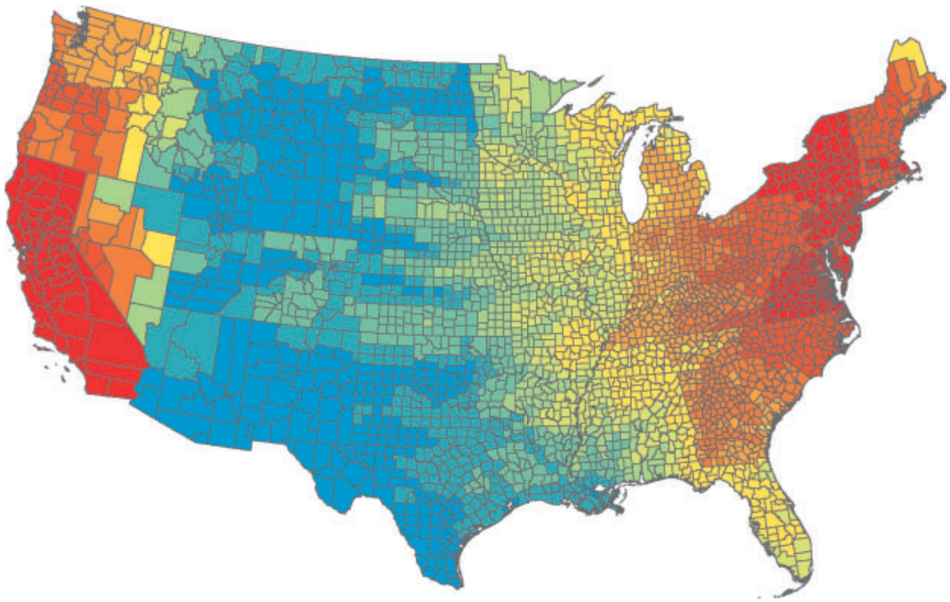


FIGURE XVII

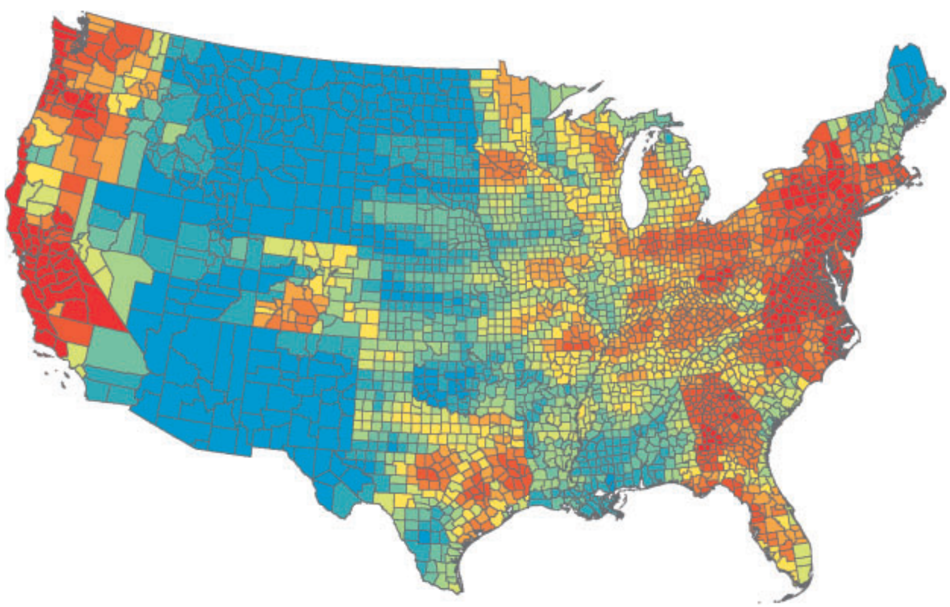
Estimated Increase in the Price Index from Removing the Interstate Highway System

This figure depicts the estimated increase in the price index (by decile) across space from removing the IHS, holding wages and productivities constant at the 2000 U.S. levels. Darker shading indicates higher deciles (e.g., the removal of the IHS disproportionately increased the economic remoteness in more darkly shaded regions).

in California and the Eastern Seaboard purchased more goods from nearby locations (since a greater amount of production was concentrated nearby), so they relied less on the IHS. The importance of the latter effect, however, depends on how the spatial distribution of population (and hence production) will endogenously change in response to changes in the trade costs. Figure XVIII shows how removing the IHS changes the spatial distribution of the population. Consistent with the fact that California and the Eastern Seaboard incur relatively small increase in economic remoteness, there is a redistribution of the population toward those locations and away from the Rocky Mountains. However, the redistribution of population across space depends importantly on the strength of spillovers: when spillovers are absent (top map of Figure XVIII), there is substantially less local variation of population changes than when $\alpha = 0.1$ and $\beta = -0.3$ (bottom map of Figure XVIII).



$\alpha = 0,\ \beta = 0$



$\alpha = 0.1,\ \beta = -0.3$

FIGURE XVIII

Estimated Change in the Population from Removing the Interstate Highway System

(continued)

Finally, Figure XIX presents the effect of the removal of the IHS on welfare for a large number of different spillover strengths. Depending on the strength of spillovers, we estimate that removing the IHS would result in a decline in welfare of between 1.1% and 1.4%.[30] Given this estimate, a simple back-of-the-envelope calculation suggests that the benefits of the IHS substantially outweigh its costs. According to the Congressional Budget Office, the total cost of constructing the IHS was $560 billion (in 2007 dollars); assuming a 5% annual cost of capital, this amounts to roughly $28 billion a year (CBO 1982). The total cost of maintaining the entire highway system is approximately $130 billion a year (FHA 2008; NSTIFC 2009). If we assume that half of that expense is spent on the IHS,[31] this suggests the total annual cost of building and maintaining the IHS is approximately $100 billion. In comparison, the U.S. GDP in 2007 was $14.25 trillion; since preferences are assumed to be homothetic, if removing the IHS would decrease (static) welfare by 1.1–1.4%, the model implies the monetary value of the IHS is between $150 and $200 billion 2007 dollars, suggesting an overall return on investment of at least 50%, or an annualized return of at least 9% ($\frac{150-100}{560}$) a year.

IV. Conclusion

We view this article as taking a number of steps toward the rigorous quantification of spatial theory. First, we develop a

Figure XVIII Continued

This figure shows the estimated change in population (in deciles) from the removal of the IHS. The top map reports the estimated population changes when there are no spillovers (i.e., $\alpha = \beta = 0$), and the bottom map reports the estimated population changes when spillovers are chosen to approximately match those from the literature (i.e., $\alpha = 0.1$ and $\beta = -0.3$). Darker shading indicates higher deciles (e.g., the removal of the IHS increased the relative population in more darkly shaded regions).

30. This estimated welfare loss arises only from the additional cost of trading goods. To the extent that the IHS had other benefits (e.g., facilitating passenger travel), the welfare loss of removing the IHS would be even greater.

31. The IHS accounts for about one quarter of all passenger miles on the system, but the maintenance costs are likely higher per passenger miles than other highways (Duranton and Turner 2012).

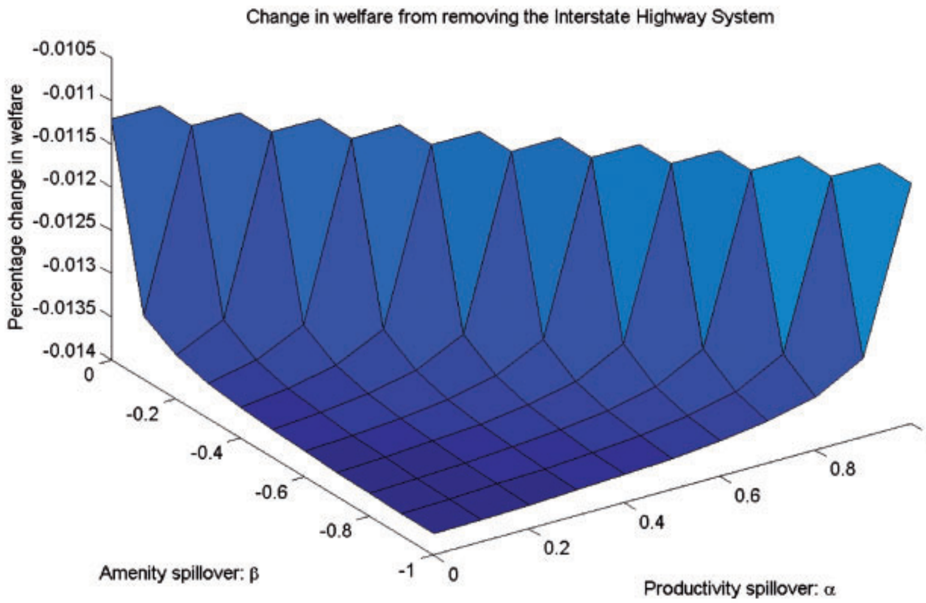


FIGURE XIX

Estimated Decline in Welfare from Removing the Interstate Highway System

This figure shows the estimated decline in welfare (in percentage terms) from the removal of the IHS for each combination of productivity spillover strength $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$ such that $\alpha + \beta \leq 0$.

unified general equilibrium framework combining labor mobility, gravity, and productivity and amenity spillovers. Within this framework, we establish conditions for the existence and uniqueness of a spatial equilibrium and derive relationships between the equilibrium distribution of economic activity and the underlying geography. Given the isomorphisms of our framework to multiple existing frameworks in the literature, we see this as helping, in the words of Duranton (2008), to "provide a unified general equilibrium approach to spatial economics and end the often uneasy coexistence between urban systems and the new economic geography." Second, we provide a micro-foundation of trade costs as the accumulation of instantaneous trade costs over the least-cost route on a surface. We then develop tools to apply our framework to the analysis of detailed real world data on spatial economic activity.

This framework could be extended to address a number of other questions, including: what is the optimal spatial taxation scheme in both the short-run and long-run? What transportation system maximizes social welfare? How would removing restrictions on cross-country migration affect the equilibrium distribution of economic activity?

NORTHWESTERN AND NBER
YALE AND NBER

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

REFERENCES

Acemoglu, D., S. Johnson, and J. Robinson, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 117 (2002), 1231–1294.
Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf, "The Economics of Density: Evidence from the Berlin Wall," CEPR Working Paper 1154, 2012.
Allen, T., "Information Frictions in Trade," Econometrica, Forthcoming.
Allen, T., C. Arkolakis, and Y. Takahashi, "Universal Gravity," Mimeo Northwestern and Yale University, 2014.
Amiti, M., and L. Cameron, "Economic Geography and Wages," *Review of Economics and Statistics*, 89 (2007), 1231–1294.
Anderson, J. E., "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69 (1979), 106–116.
Anderson, J. E., and E. Van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (2003), 170–192.
———, "Trade Costs," *Journal of Economic Literature*, 42 (2004), 691–751.
Arkolakis, C., A. Costinot, and A. Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 102 (2012), 94–130.
Arkolakis, C., S. Demidova, P. J. Klenow, and A. Rodríguez-Clare, "Endogenous Variety and the Gains from Trade," *American Economic Review, Papers and Proceedings*, 98 (2008), 444–450.
Armington, P. S., "A Theory of Demand for Products Distinguished by Place of Production," *International Monetary Fund Staff Papers*, 16 (1969), 159–178.
Beckmann, M., "A Continuous Model of Transportation," *Econometrica* (1952), 643–660.
Beckmann, M. J., and T. Puu, *Spatial Economics: Density, Potential, and Flow* (Amsterdam: North-Holland, 1985).
———, *Spatial Structures* (Berlin: Springer, 1990).
BLS, "2000 Consumer Expenditure Survey," Bureau of Labor Statistics, 2000, http://www.bls.gov/cex/csxann00.pdf.
Breinlich, H., "The Spatial Economic Structure in the European Union: What Role for Economic Geography?," *Journal of Economic Geography*, 12 (2006), 593–617.
CBO, "The Interstate Highway System: Issues and Options," Congressional Budget Office of the United States, 1982.
CFS, "2007 Commodity Flow Survey," Bureau of Transportation Statistics, 2007.
Chaney, T., "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98 (2008), 1707–1721.
CTA, "Railroad Network," Center for Tranportation Analysis, 2003.
Davis, D., and D. Weinstein, "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92 (2002), 1269–1289.
———, "A Search for Multiple Equilibria in Urban Industrial Structure," *Journal of Regional Science*, 48 (2008), 29–65.
Diamond, R., "The Determinants and Welfare Implications of U.S. Workers' Diverging Location Choices by Skill: 1980–2000," (Mimeo: Stanford University, 2012).

Donaldson, D., “Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure,” Working Paper, MIT, 2012.
Donaldson, D., and R. Hornbeck, “Railroads and American Economic Growth: New Data and Theory,” Discussion Paper, Harvard University and Massachusetts Institute of Technology, 2012.
Duranton, G., “Spatial Economics,” in *The New Palgrave Dictionary of Economics*, 2nd ed., S. N. Durlauf, and L. Blume, eds. (London: Palgrave Macmillan, 2008).
Duranton, G., and D. Puga, “Microfoundations of Urban Agglomeration Economies,” in *Handbook of Regional and Urban Economics,* vol. 4, J. V. Henderson, and J. Thisse, eds. (Amsterdam: North-Holland, 2004).
Duranton, G., and M. A. Turner, “Urban Growth and Transportation,” *Review of Economic Studies*, 79 (2012), 1407–1440.
Eaton, J., and S. Kortum, “Technology, Geography and Trade,” *Econometrica*, 70 (2002), 1741–1779.
Ellison, G., and E. Glaeser, “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy*, 105 (1997), 889–927.
Fabinger, M., “Trade and Interdependence in a Spatially Complex World,” Penn State, Mimeo, 2011.
FHA, “2008 Status of the Nation’s Highways, Bridges, and Transit: Conditions and Perfomance,” Federal Highway Administration and Federal Transit Administration, 2008.
Fujimoto, T., and U. Krause, “Strong Ergodicity for Strictly Increasing Nonlinear Operators,” *Linear Algebra and its Applications*, 71 (1985), 101–112.
Fujita, M., P. Krugman, and A. J. Venables, *The Spatial Economy: Cities, Regions, and International Trade* (Boston, MA: MIT Press, 1999).
Fujita, M., and J.-F. Thisse, *Economies of Agglomeration: Cities, Industrial Location, and Globalization* (Cambridge: Cambridge University Press, 2013).
Hanson, G. H., “Market Potential, Increasing Returns, and Geographic Concentration,” *Journal of International Economics*, 67 (2005), 1–24.
Head, K., and T. Mayer, “Regional Wage and Employment Responses to Market Potential in the EU,” *Regional Science and Urban Economics*, 36 (2006), 573–595.
Helpman, E., “The Size of Regions,” *Topics in Public Economics. Theoretical and Applied Analysis* (1998), 33–54.
Helpman, E., and P. Krugman, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy* (Cambridge, MA: MIT Press, 1985).
Hillberry, R., and D. Hummels, “Trade Responses to Geographic Frictions: A Decomposition Using Micro-data,” *European Economic Review*, 52 (2008), 527–550.
Karlin, S., and L. Nirenberg, “On a Theorem of P. Nowosad,” *Journal of Mathematical Analysis and Applications*, 17 (1967), 61–67.
Kline, P., and E. Moretti, “Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *Quarterly Journal of Economics*, 129 (2014), 275–331.
Krasnosel’skii, M. A., A. H. Armstrong, and J. Burlak, *Topological methods in the theory of nonlinear integral equations* (New York: Pergamon Press).
Krugman, P., “Increasing Returns Monopolistic Competition and International Trade,” *Journal of International Economics*, 9 (1979), 469–479.
———, “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 70 (1980), 950–959.
———, “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 99 (1991), 483–499.
Krugman, P., and A. J. Venables, “The Seamless World: A Spatial Model of International Specialization,” NBER Discussion Paper, 1995.
Lucas, R. E., and E. Rossi-Hansberg, “On the Internal Structure of Cities,” *Econometrica*, 70 (2003), 1445–1476.
Lux, M., “Defying Gravity: The Substitutability of Transportation in International Trade,” Mimeo, New York University, 2011.

Mantegazza, C., and A. Mennucci, "Hamilton-Jacobi Equations and Distance Functions on Riemannian Manifolds," *Applied Mathematics and Optimization*, 47 (2003), 1–26.
Matsuyama, K., "Geography of the World Economy," CMS-EMS Discussion Paper 1239, presented as the Fukuzawa Lecture, Far Eastern Econometric Society Meeting at Singapore, 1999.
Melitz, M. J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71 (2003), 1695–1725.
Mitchell, J. S., and D. M. Keirsey, "Planning Strategic Paths through Variable Terrain Data," in *1984 Technical Symposium East* (International Society for Optics and Photonics, 1984), 172–179.
Mossay, P., and P. M. Picard, "On Spatial Equilibria in a Social Interaction Model," *Journal of Economic Theory*, 146 (2011), 2455–2477.
MPC, "National Historical Geographic Information System: Version 2.0," (Minnesota Population Center, 2011a).
———, "National Historical Information System: Version 2.0," Minnesota Population Center, 2011b, http://www.nhgis.org.
NDC, "U.S. Waterway Data," Navigation Data Center, 1999.
NHPN, "National Highway Planning Network," Federal Highway Administration, 2005.
Nowosad, P., "On the Integral Equation kf Equals l/f Arising in a Problem in Communication," *Journal of Mathematical Analysis and Applications*, 14 (1966), 484–492.
NSTIFC, "Paying Our Way: A New Framework for Transportation Finance," National Surface Transportation Infrastructure Financing Commission, 2009.
Pérez-Cervantes, F., "Railroads and Economic Growth: A Trade Policy Approach," Mimeo: University of Chicago, 2012).
Polyanin, A., and A. Manzhirov, *Handbook of Integral Equations* (London: Chapman & Hall/CRC, 2008).
Polyanin, A., and V. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations* (London: Chapman & Hall/CRC, 2008).
Redding, S. J., "Goods Trade, Factor Mobility and Welfare," Mimeo, 2012.
Redding, S., and D. Sturm, "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98 (2008), 1766–1797.
Roback, J., "Wages, Rents, and the Quality of Life," *Journal of Political Economy* (1982), 1257–1278.
Rosenthal, S. S., and W. C. Strange, "Evidence on the Nature and Sources of Agglomeration Cconomies," *Handbook of Regional and Urban Economics*, 4 (2004), 2119–2171.
Rossi-Hansberg, E., "A Spatial Theory of Trade," *American Economic Review*, 95 (2005), 1464–1491.
Rossi-Hansberg, E., and M. L. J. Wright, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, 97 (2007), 1639–1666.
Sachs, J., "Tropical Underdevelopment," NBER Discussion Paper, 2001.
Sethian, J., "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences*, 93 (1996), 1591–1595.
———, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science,* vol. 3 (Cambridge: Cambridge University Press, 1999).
Shorrocks, A., "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," *Journal of Economic Inequality*, 11 (2013), 99–126.
Starrett, D., "Market Allocations of Location Choice in a Model with Free Mobility," *Journal of Economic Theory*, 17 (1978), 21–37.
Tsitsiklis, J., "Efficient Algorithms for Globally Optimal Trajectories," *IEEE Transactions on Automatic Control*, 40 (1995), 1528–1538.

Yang, Z., and D. O'Regan, "Positive Solvability of Systems of Nonlinear Hammerstein Integral Equations," *Journal of Mathematical Analysis and Applications*, 311 (2005), 600–614.

Zabreyko, P., A. Koshelev, M. Krasnosel'skii, S. Mikhlin, L. Rakovshchik, and V. Stetsenko, *Integral Equations: A Reference Text* (Leyden: Noordhoff International Publishing, 1975).