Trade and the Topography of the Spatial Economy*

Treb Allen — Northwestern

Costas Arkolakis — Yale and NBER

First Version: November 2012
This Version: February 2014

Abstract

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 which 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.

*We thank George Alessandria, Lorenzo Caliendo, Arnaud Costinot, Dave Donaldson, Gilles Duranton, 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. Xiangliang Li provided excellent research assistance. This paper 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 Introduction

There exists an enormous disparity in economic activity across space. For example, in the year 2000, the population density in McLeod County, MN was 26 persons/km^2 and the payroll per capita was \$13,543, while 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, etc).[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 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 paper, 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 *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 ("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

[1] The literature examining the factors contributing to the productivity of a location is immense, see e.g. Sachs (2001), Acemoglu, Johnson, and Robinson (2002).

[2] The idea of analyzing economic activity on a surface has a long tradition, see e.g. Beckmann (1952); Beckmann and Puu (1985, 1990).

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 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 several stylized examples of the mechanisms of the model 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 its 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 twenty percent of the spatial variation in income across the United States in the year 2000 can be explained by geographic location alone. Second, we examine the effect of removing the Interstate Highway System. We estimate that without the Interstate Highway

[3]There is a (small) literature which provides analytical characterizations of the equilibrium distribution of economic activity across space, e.g. Matsuyama (1999) and Fabinger (2011).

System, welfare would decline by 1.1 to 1.4 percent, 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) and Krugman (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. While 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.

The primary goal of the paper is to provide an empirically implementable framework to study the role of economic geography. While there has been much empirical work examining the implication of space for the allocation of agents (Davis and 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),

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

and Ahlfeldt, Redding, Sturm, and Wolf (2012) use a quantitative framework to analyze the spatial distribution of economic activity. This paper follows in their tradition, and develops a number tools to facilitate future quantitative analysis of economic geography.

Finally, our empirical work is related to the recent literature estimating the impact of the transportation network on economic output. Donaldson (2012) and Cervantes (2012) consider the impact of railroads in India and the US, respectively, when labor is immobile, while Donaldson and Hornbeck (2012) consider the impact of the railroad network in the US when labor is mobile. While we show how such transportation networks can be incorporated in our framework, we can also incorporate geographical characteristics (e.g. mountains) that do not have obvious network representations.

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

2 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 on a surface for several simple examples of the model.

2.1 Economic component

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

2.1.1 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

[5]The continuum of locations is not important for much of what follows. In particular, as we discuss below, Theorems 1 and 2 generalize for the case of discrete number of locations; however, Proposition 1 is only true

$i \in S$ produces a unique differentiated variety of a good. Trade is costly: trade costs are of the iceberg form and are described by the function $T : S \times S \to [1, \infty)$, 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 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.

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 $L : S \to \mathbb{R}_{+}$ and $w : S \to \mathbb{R}_{++}$ to be the density of workers and their wage, respectively.

In order to allow for the possibility of productivity or congestion externalities, both productivity and amenities may depend on the density of workers. In particular, we assume that overall (or composite) productivity in location i can be written as:

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

where $\bar{A}(i)$ 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:

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

where $\bar{u}(i)$ is the exogenous utility derived from living in location i inherent to the location

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.

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. While 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 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)-Redding (2012) framework with $1-\delta$ being the budget share spent on an immobile factor, e.g. 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. In this case, the amenity spillover can be interpreted as also capturing the disutility of higher housing prices. The model is also isomorphic to a model where land is a factor of production if $\alpha = 1-\delta$, where δ is the share of labor in the Cobb-Douglass production function, in which case the productivity spillover can be interpreted as also capturing the diminishing returns to labor in the production function. Finally, if $\beta = -\frac{1}{\theta}$, our model is isomorphic to one where 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 discuss above regards trade flows, wages, population and welfare, but not necessarily other aspects of these models. Independently of their interpretation, the degree of this agglomeration and dispersion externalities are crucial to guarantee uniqueness and existence of a spatial equilibrium, as we will discuss in detail below.

We define the *geography* of S to be the set of functions $\bar{A}$, $\bar{u}$, and T, where $\bar{A}$ and $\bar{u}$ comprise the *local characteristics* and T comprises the *geographic location.* S is said to have a *regular geography* if $\bar{A}$, $\bar{u}$, 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 $\int_S w(s)\, ds = 1$.

2.1.2 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

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

2.1.3 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, i.e. 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 features. A spatial equilibrium is said to be *regular* if w and L are continuous and every location is inhabited, i.e. 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.[6]

2.1.4 Existence, uniqueness, and stability

We now discuss sufficient conditions for the existence and uniqueness of regular spatial equilibria. Using equations (3) to substitute out for trade flows 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 $A(i) = \bar{A}(i)$ and $u(i) = \bar{u}(i)$) 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 $w(i)^{1-\sigma}$, respectively. Note that the kernels of the two equations are transposes of each other. As a result, we have 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 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

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

solution. Part (ii) of the theorem 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:

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

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

Equations (10) and (11) are a system of two non-linear integral equations, which 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, i.e. $T(i,s) = T(s,i)$ for all $i, s \in S$, it turns out that the system can be written as a single non-linear integral equation, which will allow us to provide a simple characterization of the equilibrium system.[7] To see this, suppose that the left hand sides of equations (8) and (9) are equal up to scale:

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

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):

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

[7]It is interesting to note that this method of reducing a system of non-linear equations into a single non-linear 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 and Arkolakis (2013).

where

$$\gamma_1 \equiv 1 - \alpha(\sigma - 1) - \beta\sigma,$$
$$\gamma_2 \equiv 1 + \alpha\sigma + (\sigma - 1)\beta,$$

and $\tilde{\sigma} \equiv (\sigma - 1) / (2\sigma - 1)$.

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 non-linear integral equation known as a homogeneous Hammerstein equation of the second kind (see, e.g. p.807 of Polyanin and Manzhirov, 2008). 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 for $\gamma_1 \neq 0$.

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 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 assume that iceberg trade costs are symmetric and parameters are 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 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:

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

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 $\gamma_1 > 0$.[8]

Figure 1 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 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 $\gamma_2/\gamma_1 > 1$). We provide specific examples of the possible multiple equilibria below. 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, i.e. 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).[9]

The existence and uniqueness results of the Theorem 1 and 2 generalize for a discrete number of locations, as we discuss in the Appendix, in which case the set S is finite or countable.[10] 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) since a change in the population in one location will affect the price index.

[8] 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.

[9] 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 Appendix A.

[10] 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 "$\sum_S$".

2.1.5 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 both wages $w(i)$ and population $L(i)$ can be written as log linear functions of the 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 will decrease 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, productivity and amenity spillovers only change the elasticity of the equilibrium distribution of economic activity to the underlying geography. 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.[11]

2.2 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 $\mathbb{R}^N$.[12] 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

[11] 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.

[12] A manifold is a topological space that is locally Euclidean, or intuitively, a space that can be 'charted' in the Euclidean space.

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

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

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 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 railroad 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, water, etc. 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, i.e. 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 $\lim_{s \to i} T(s,j) = T(i,j)$. While 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 (2012)). We will allow for such non-geographic trade costs when we estimate the total bilateral trade costs for the United States in Section 3.

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, i.e. 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 normal 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 2 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 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, the bilateral trade costs will only increase as you move away from the origin. As a result, subsequent contours can be constructed using 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 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 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 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 so-called "digitization bias." Digitization bias arises because any chosen grid necessarily restricts the possible directions of travel, biasing estimated distances upwards, 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 incorporate such physical realities like elevation changes or one-way roads, for example. 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: 1) the slope of the current iso-cost contour (which determines the direction of the vector); and 2) 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 to depend on the direction of travel $\vec{d}$, i.e. $\tau\left(i, \vec{d}\right)$. We provide a simple example of the direction of travel mattering in Section 2.3 below. 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 will 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.[13] 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 online appendix) that any log sub-additive monotonically increasing function f such that $f(0) = 1$ will generate bilateral iceberg trade costs that are weakly greater than one and satisfy the triangular inequality, i.e. $T(i,j) \leq T(i,k)\, T(k,j)$ for all i, k, j.

[13]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.

2.3 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 to illustrate the different types of equilibria that may arise and discuss their stability properties.

The line

Let S be the $[-\pi,\pi]$ interval and suppose that $\alpha=\beta=0$ and $\bar{A}(i)=\bar{u}(i)=1$, i.e. there are no spillovers and all locations have homogeneous exogenous productivities and amenities. Suppose that instantaneous trade costs are constant, i.e. $\tau(i)=\tau$ 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 $T(i,s)=e^{b+\tau|i-s|}$.[14] 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 $\frac{\partial \ln P(0)}{\partial i}=\tau\left(1-e^{(1-\sigma)b}\right)/\left(1+e^{(1-\sigma)b}\right)$ 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 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 $k_1\equiv(1-\sigma)^2\tau^2+2(1-\sigma)\tau W^{1-\sigma}$. Given the boundary conditions above, the equilibrium distribution of labor for each line segment $[-\pi,0]$ and $[0,\pi]$ is characterized by the weighted sum of the cosine and sine functions (see example 8.8.16 in Polyanin and Manzhirov

[14]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).

(2008)):

$$L\left(i\right)=\left(k_2\cos\left(i\sqrt{k_1}\right)+k_3\sin\left(i\sqrt{k_1}\right)\right)^{\frac{1}{\tilde{\sigma}}}.$$

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.[15] 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.[16]

Figure 3 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 4 shows how a border affects the equilibrium population distribution with a positive instantaneous trade cost. As is evident, the larger the border, 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 the 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 $\bar{A}\left(i\right) = e^{\frac{A}{\sigma-1}i}$. Then the differential equation becomes:

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

so that the equilibrium distribution of population is shifted rightward when $A > 0$. Figure 5 depicts this reallocation of labor toward locations with higher productivity. In this case, it can be shown that an analytical solution of $L\left(i\right)$ exists in terms of Bessel functions of the first and the second kind.

[15] More general formulations of the exogenous productivity or amenity functions result to more general specifications of the second order differential equation illustrated above (see Polyanin and Zaitsev (2002) section 8.1 for a number of tractable examples).

[16] 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.

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 6 depicts the population for higher values of α, and the resulting increase in the concentration. Notice that further 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 while the cost of traveling to the left on the line is τ_l, where $\tau_r \geq \tau_l$. Figure 7 illustrates that is it becomes increasingly costly to travel to the right relative to travel to the left, the equilibrium distribution of the population shifts leftward, where the price index is lowest.

The circle

The example of the circle illustrates the possibility of multiplicity of spatial equilibria. Figure 8 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. This remains an equilibrium when $\alpha+\beta > 0$, but 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 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.[17]

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 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 and its details are provided in the online appendix.

[17] If we further increase $\alpha + \beta$ to the point that the sign of γ_1 turns negative we can only find numerically 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.

In the line and circle examples above, trade costs act as an agglomeration force. However, in economic geography models such as those of Krugman (1991); Fujita, Krugman, and Venables (1999), trade costs also generate a dispersion effect by creating a home market for the manufacturing sector. This effect arises from the presence of an additional 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.

3 The topography of the real world spatial economy

In this section, we use the model developed in Section 2 to analyze the actual topography of economic activity in the continental United States. The section is composed of three parts. In the first part, we estimate the underlying geography of the United States. In the second part, we determine the fraction of the observed spatial variation in income due to geographic location. In the third part, 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 eight.[18]

3.1 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 $\bar{A}$ and amenities $\bar{u}$. To do so, we proceed in two steps. We first estimate trade costs using the observed transportation networks in order 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 $\bar{A}$ and amenities $\bar{u}$.

To estimate the underlying geography of the U.S., we rely on different types of data which we summarize here; see Appendix B for details. The first type of data is the complete high-

[18]While 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.

way, rail, and navigable water networks in the United States, which we collect from several sources (NDC, 1999; CTA, 2003; NHPN, 2005). Figure 9 depicts the networks; the networks are quite detailed and include the entire U.S. highway system (400,000 miles of interstates, other highways and arterial roads), all railroads in the U.S. (approximately 140,000 miles), and all navigable waterways (approximately 190,000 miles). Using GIS software, we project the transportation networks onto a 1032×760 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.[19] 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 10 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 12 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 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

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

and no spatial productivity and amenity spillovers become.[20] We feel that treating each county as a distinct location provides a reasonable balance of the two trade-offs.

3.1.1 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 Fast Marching Method (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 non-geographic 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 above, we create an instantaneous cost function $\tau_m : S \to \mathbb{R}_{++}$, 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 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 below.

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 1 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, while, increases in water distance are actually associated with greater shipments via road (column 3), suggesting that

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

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.[21]

We next determine the relative cost of trade across different modes of transport using a discrete choice framework.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 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.

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 $m \in \{1, ..., M\}$ modes of transport and the iceberg cost of trader t shipping goods from i to j using mode m is $\exp\left(\tau_m d_m\left(i,j\right) + f_m + \nu_{tm}\right)$, 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.[22] Finally, suppose that ν_{tm} is distributed i.i.d. across traders and modes of transportation with a Gumbel distribution with shape parameter θ, i.e. $\Pr\{\nu \leq x\} = e^{-e^{-\theta x}}$.[23] (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 $\pi_m\left(i,j\right)$ 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:

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

[21]The results are similar if we constrain our analysis to only trade between metropolitan statistical areas rather than all CFS areas; results available upon request.

[22]While the introduction of a fixed cost violates the continuity assumption of Section 2, this is not a practical concern here because we consider only a discrete number of locations.

[23]Our discrete choice framework bears a resemblance to the one presented in Lux (2011); in that framework, there were a continuum of goods, where each goods 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 share 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 non-linear 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:

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

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

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

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

Table 2 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 non-geographic observable on trade costs. Because the estimation procedure is a multiple-stage process, we calculate bootstrapped standard errors derived from re-doing 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). While 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 10 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 percent for domestic distribution costs in a representative rich country). The estimated non-geographic 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 11 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 11 shows that there is not any systematic relationship between the predicted and actual trade flows and distance, suggesting that the assumed exponential relationship between trade costs and distance is a reasonable approximation.

3.1.2 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, there exists a unique topography of overall productivities and amenities that generate that equilibrium.

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, $\bar{A} = AL^{-\alpha}$ and $\bar{u} = uL^{-\beta}$.*

Proof. See Appendix A.1.5. ■

It is important to note that Theorem 3 does not rely upon 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, $\bar{A}$ and $\bar{u}$ 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 $\bar{A}$ and $\bar{u}$ can be chosen to generate the composite productivities necessary to generate the observed equilibrium distribution of economic activity.[24]

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, i.e. $T(i,s) = T(j,s)$ 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.[25] Equations (1) and (2) simply extend this intuition to allow for differences in trade costs across locations.

Using equations (1) and (2) and the bilateral trade costs estimated in the previous section, we identify the unique composite amenities and productivities of each U.S. county in the

[24] 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.

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

year 2000 that are consistent with the observed distribution of labor and wages from the 2000 U.S. Census. Figure 13 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, while composite productivities are much higher. Figure 14 depicts the resulting exogenous amenities and productivities when $\alpha = 0.1$ and $\beta = -0.3$ (values which are roughly consistent with the estimates of productivity spillovers from Rosenthal and Strange (2004) and the share of income spent on housing BLS (2000)).[26] The topography of exogenous productivities and amenities seem reasonable; amenities in southern Florida, southern California, and Arizona are high, while 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.

3.2 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:

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

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 15 depicts the geographic variation in the estimated price index. There exists substantial heterogeneity in the price index; both within and across states, counties with better geographic location (i.e. lower price indices) are wealthier.

[26] 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 "housefurnishings 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 "housefurnishings and equipment" is included, 24.7% when all three categories are included). Given our isomorphism that implies $\beta = \delta / (1-\delta)$, 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.

To determine the relative contribution of the effect of local characteristics (i.e. $\bar{A}(i)$ and $\bar{u}(i)$) 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 mis-specified (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 downwards the estimated contribution of the price index. As a result, the contribution of the price index should be considered a lower bound for the importance of the geography of trade costs in explaining the differences in income across space.

Figure 16 reports the fraction of the spatial variation in income in the United States in the year 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 upwards 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.

3.3 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 re-calculate the bilateral trade cost function T using the estimates from Section 3.1.1 supposing that there were no interstate highways, but keeping all other modes of transportation (including other national highways and arterial roads) unchanged. For a given strength of spillovers α and β, we hold fixed the exogenous productivities $\bar{A}$ and amenities $\bar{u}$ at the values estimated in Section 3.1.2 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 combinations of $\alpha \in [0,1]$ and $\beta \in [-1,0]$ such that $\alpha + \beta \leq 0$, a restriction which from Theorem 2 guarantees the uniqueness of equilibrium.

To illustrate the effect of the removal of the IHS on trade costs, Figure 17 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 9); additionally, locations 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 4 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 4), there is substantially less local variation of population changes than when $\alpha = 0.1$ and $\beta = -0.3$.

Finally, Figure 19 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 $1.1 - 1.4\%$.[27] Given this estimate of the welfare losses of removing the IHS, 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

[27] 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.

(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,[28] 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.

4 Conclusion

We view this paper as taking a number of steps toward the rigorous quantification of spatial theory. First, we develop a 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 to, in the words of Duranton (2008), "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?

[28]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).

References

Acemoglu, D., S. Johnson, and J. Robinson (2002): "Reversal of fortune: Geography and institutions in the making of the modern world income distribution," *The Quarterly Journal of Economics*, 117(4), 1231–1294.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2012): "The Economics of Density: Evidence from the Berlin Wall," *CEPR Working paper*, 1154.

Allen, T. (2012): "Information Frictions in Trade," *mimeo, Yale University*.

Allen, T., and C. Arkolakis (2013): "A note on the existence and uniqueness of trade equilibria," *mimeo, Northwestern and Yale Universities*.

Amiti, M., and L. Cameron (2007): "Economic Geography and Wages," *Review of Economics and Statistics*, 89(1), 1231–1294.

Anderson, J. E. (1979): "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69(1), 106–116.

Anderson, J. E., and E. Van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1), 170–192.

——— (2004): "Trade Costs," *Journal of Economic Literature*, 42(3), 691–751.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

Arkolakis, C., S. Demidova, P. J. Klenow, and A. Rodríguez-Clare (2008): "Endogenous Variety and the Gains from Trade," *American Economic Review, Papers and Proceedings*, 98(4), 444–450.

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

Beckmann, M. (1952): "A continuous model of transportation," *Econometrica*, pp. 643–660.

Beckmann, M. J., and T. Puu (1985): *Spatial economics: density, potential, and flow*. North-Holland Amsterdam.

——— (1990): *Spatial structures*. Springer-Verlag Berlin.

Birkhoff, G. (1957): "Extensions of Jentzsch's theorem," *Transactions of the American Mathematical Society*, 85(1), 219–227.

BLS (2000): "2000 Consumer Expenditure Survey," http://www.bls.gov/cex/csxann00.pdf.

Breinlich, H. (2006): "The Spatial Economic Structure in the European Union: What Role for Economic Geography?," *Journal of Economic Geography*, 12(6), 593–617.

CBO (1982): "The interstate highway system: Issues and options," *Congressional Budget Office of the United States*.

Cervantes, F. P. (2012): "Railroads and Economic Growth: A Trade Policy Approach," *mimeo*.

CFS (2007): "2007 Commodity Flow Survey," *Bureau of Transportation Statistics*.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707–1721.

CTA (2003): "Railroad Network," *Center for Tranportation Analysis*.

Davis, D., and D. Weinstein (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

——— (2008): "A Search for Multiple Equilibria in Urban Industrial Structure," *Journal of Regional Science*, 48(1), 29–65.

Diamond, R. (2012): "The determinants and welfare implications of U.S. workers' diverging location choices by skill: 1980-2000," *mimeo*.

Donaldson, D. (2012): "Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure," Manuscript, MIT.

Donaldson, D., and R. Hornbeck (2012): "Railroads and American Economic Growth: New Data and Theory," Discussion paper.

Duranton, G. (2008): "Spatial Economics," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. Blume, vol. 2nd Edition. Palgrave Macmillan.

DURANTON, G., AND D. PUGA (2004): "Microfoundations of urban agglomeration economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. Thisse, vol. 4. North-Hol.

DURANTON, G., AND M. A. TURNER (2012): "Urban growth and transportation," *The Review of Economic Studies*, 79(4), 1407–1440.

EATON, J., AND S. KORTUM (2002): "Technology, Geography and Trade," *Econometrica*, 70(5), 1741–1779.

ELLISON, G., AND E. GLAESER (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

FABINGER, M. (2011): "Trade and Interdependence in a Spatially Complex World," *mimeo, Penn State*.

FHA (2008): "2008 status of the nation's highways, bridges, and transit: Conditions and perfomance," *Federal Highway Administration and Federal Transit Administration*.

FUJIMOTO, T., AND U. KRAUSE (1985): "Strong ergodicity for strictly increasing nonlinear operators," *Linear Algebra and its Applications*, 71, 101–112.

FUJITA, M., P. KRUGMAN, AND A. J. VENABLES (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Boston, Massachussetts.

HANSON, G. H. (2005): "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economics*, 67(1), 1–24.

HEAD, K., AND T. MAYER (2006): "Regional Wage and Employment Responses to Market Potential in the EU," *Regional Science and Urban Economics*, 36(5), 573–595.

HELPMAN, E. (1998): "The Size of Regions," *Topics in Public Economics. Theoretical and Applied Analysis*, pp. 33–54.

HILLBERRY, R., AND D. HUMMELS (2008): "Trade responses to geographic frictions: A decomposition using micro-data," *European Economic Review*, 52(3), 527–550.

KARLIN, S., AND L. NIRENBERG (1967): "On a theorem of P. Nowosad," *Journal of Mathematical Analysis and Applications*, 17(1), 61–67.

Kline, P., and E. Moretti (2014): "Local economic development, agglomeration economies and the big push: 100 years of evidence from the tennessee valley authority," *Quarterly Journal of Economics*, 129, 275–331.

Krasnosel'Skii, M., and L. Boron (1964): *Positive Solutions of Operator Equations*. P. Noordhoff Groningen.

Krasnosel'skii, M. A., A. H. Armstrong, and J. Burlak (1964): *Topological methods in the theory of nonlinear integral equations*. Pergamon Press New York.

Krugman, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–959.

——— (1991): "Increasing Returns and Economic Geography," *The Journal of Political Economy*, 99(3), 483–499.

Lucas, R. E., and E. Rossi-Hansberg (2003): "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445–1476.

Lux, M. (2011): "Defying gravity: The substitutability of transportation in international trade," *mimeo*.

Mantegazza, C., and A. Mennucci (2003): "Hamilton-Jacobi Equations and Distance Functions on Riemannian Manifolds," *Applied Mathematics and Optimization*, 47(1), 1–26.

Matsuyama, K. (1999): "Geography of the World Economy," *CMS-EMS Discussion Paper 1239*, Presented as the Fukuzawa Lecture, Far Eastern Econometric Society Meeting at Singapore.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Mitchell, J. S., and D. M. Keirsey (1984): "Planning strategic paths through variable terrain data," in *1984 Technical Symposium East*, pp. 172–179. International Society for Optics and Photonics.

Mossay, P., and P. M. Picard (2011): "On spatial equilibria in a social interaction model," *Journal of Economic Theory*, 146(6), 2455–2477.

MPC (2011a): "National Historical Geographic Information System: Version 2.0," *Minnesota Population Center*.

——— (2011b): "National Historical Information System: Version 2.0," *Minnesota Population Center*, http://www.nhgis.org.

NDC (1999): "U.S. Waterway Data," *Navigation Data Center*.

NHPN (2005): "National Highway Planning Network," *Federal Highway Administration*.

Nowosad, P. (1966): "On the integral equation kf equals 1/f arising in a problem in communication," *Journal of Mathematical Analysis and Applications*, 14, 484–492.

NSTIFC (2009): "Paying our way: A new framework for transportation finance," *National Surface Transportation Infrastructure Financing Commission*.

Polyanin, A., and A. Manzhirov (2008): *Handbook of Integral Equations*. Chapman & Hall/CRC.

Polyanin, A., and V. Zaitsev (2002): *Handbook of exact solutions for ordinary differential equations*. Chapman & Hall/CRC.

Redding, S., and D. Sturm (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766–1797.

Redding, S. J. (2012): "Goods Trade, Factor Mobility and Welfare," *mimeo*.

Roback, J. (1982): "Wages, rents, and the quality of life," *The Journal of Political Economy*, pp. 1257–1278.

Rosenthal, S. S., and W. C. Strange (2004): "Evidence on the nature and sources of agglomeration economies," *Handbook of regional and urban economics*, 4, 2119–2171.

Rossi-Hansberg, E. (2005): "A Spatial Theory of Trade," *American Economic Review*, 95(5), 1464–1491.

Rossi-Hansberg, E., and M. L. J. Wright (2007): "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, 97(5), 1639–1666.

Sachs, J. (2001): "Tropical underdevelopment," Discussion paper, National Bureau of Economic Research.

Sethian, J. (1996): "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences*, 93(4), 1591–1595.

——— (1999): *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, vol. 3. Cambridge University Press.

Shorrocks, A. (2013): "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value," *Journal of Economic Inequality*, 11(1), 99–126.

Starrett, D. (1978): "Market allocations of location choice in a model with free mobility," *Journal of Economic Theory*, 17(1), 21–37.

Tsitsiklis, J. (1995): "Efficient Algorithms for Globally Optimal Trajectories," *Automatic Control, IEEE Transactions on*, 40(9), 1528–1538.

Yang, Z., and D. O'Regan (2005): "Positive solvability of systems of nonlinear Hammerstein integral equations," *Journal of mathematical analysis and applications*, 311(2), 600–614.

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

Tables and figures

Table 1: COMMODITY FLOWS AND MODE-SPECIFIC SHIPPING DISTANCES

Dependent variable:	Road Shipments			Rail Shipments			Water Shipments		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
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	9177	9177	9177	1434	1434	1434	58	58	58

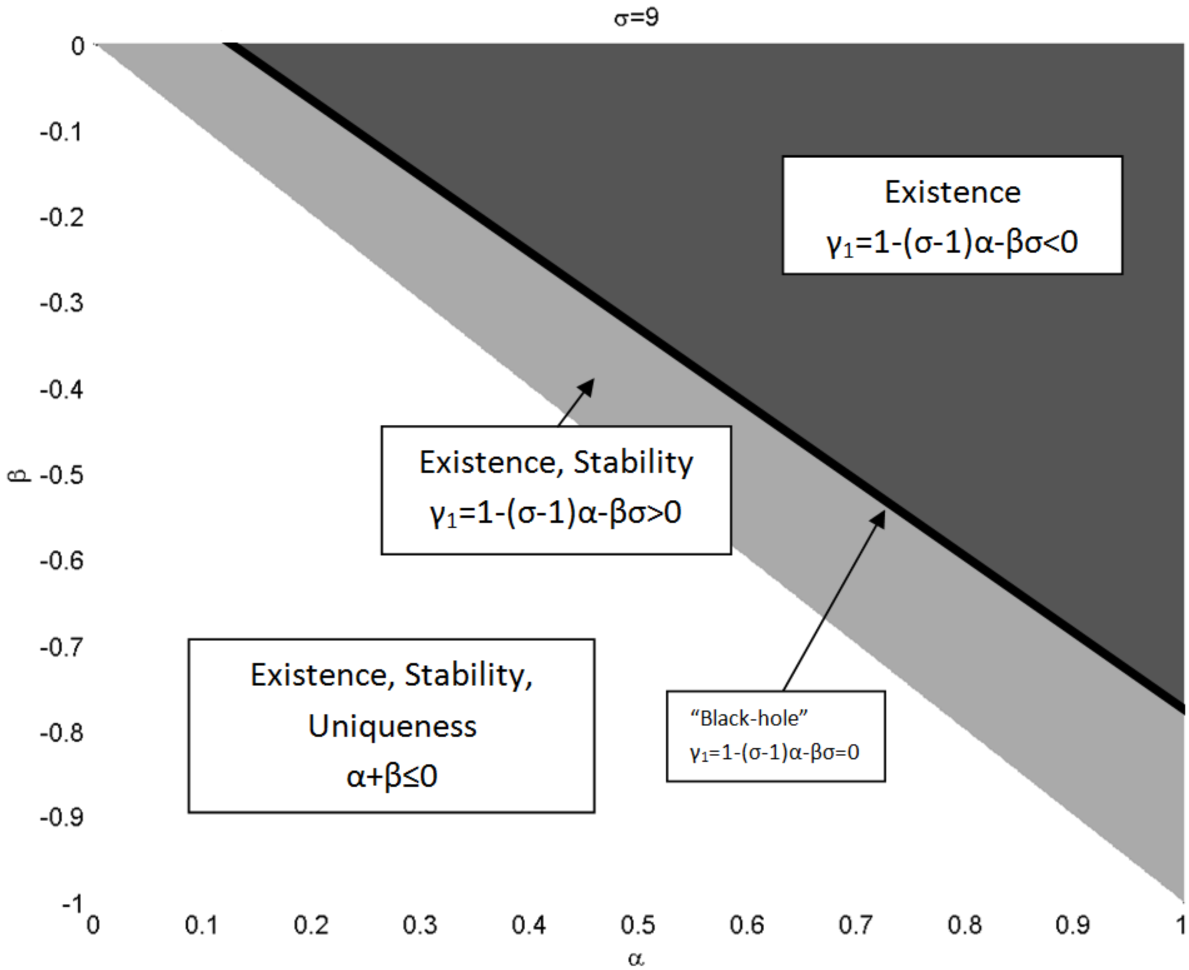
Notes: Ordinary least squares. Each observation is the observed (log) value traded from a CFS region to another CFS region in 2007 by a particular mode of transport. Standard errors are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Table 2: Estimated mode-specific relative cost of travel

	All CFS Areas				Only MSAs			
Geographic trade costs	Road	Rail	Water	Air	Road	Rail	Water	Air
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 parameter (θ)	14.225***				17.6509***			
	(0.3375)				(1.4194)			
Non-geographic 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	9601	9601	9601	9601	3266	3266	3266	3266
Observations with positive mode-specific bilateral flows	9311	1499	78	1016	3144	340	26	471

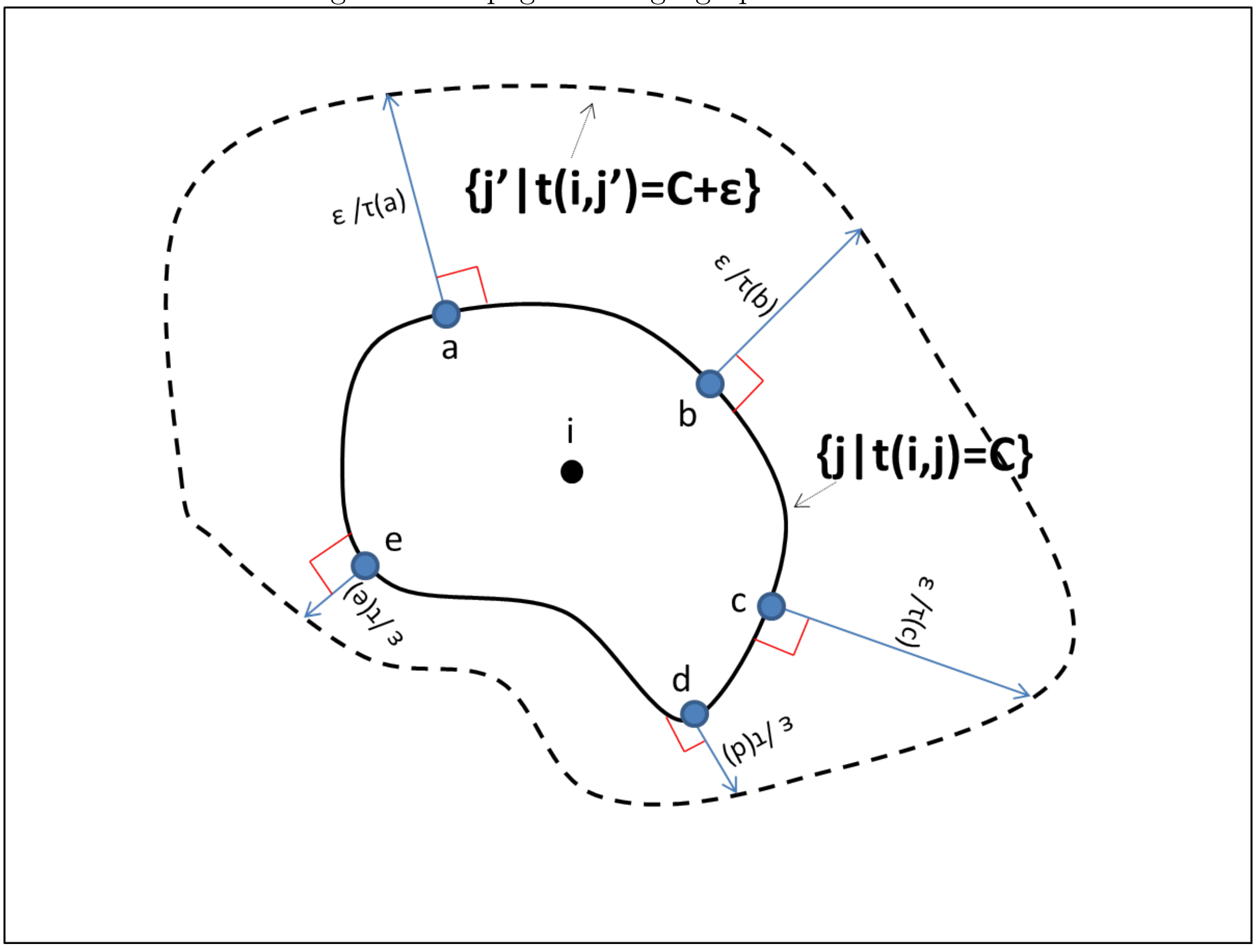
Notes: This table shows the estimated cost of traveling via different modes of travel. Relative costs are estimated using the mode-specific shares of total bilateral trade values from the 2007 Commodity Flow Survey. The mode specific bilateral cost is $\exp(variablecost * distance + fixedcost)$, where the distance of the width of the country is normalized to one. 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 1000 repetitions) reported in parentheses.

Figure 1: Equilibria with amenity and productivity spillovers



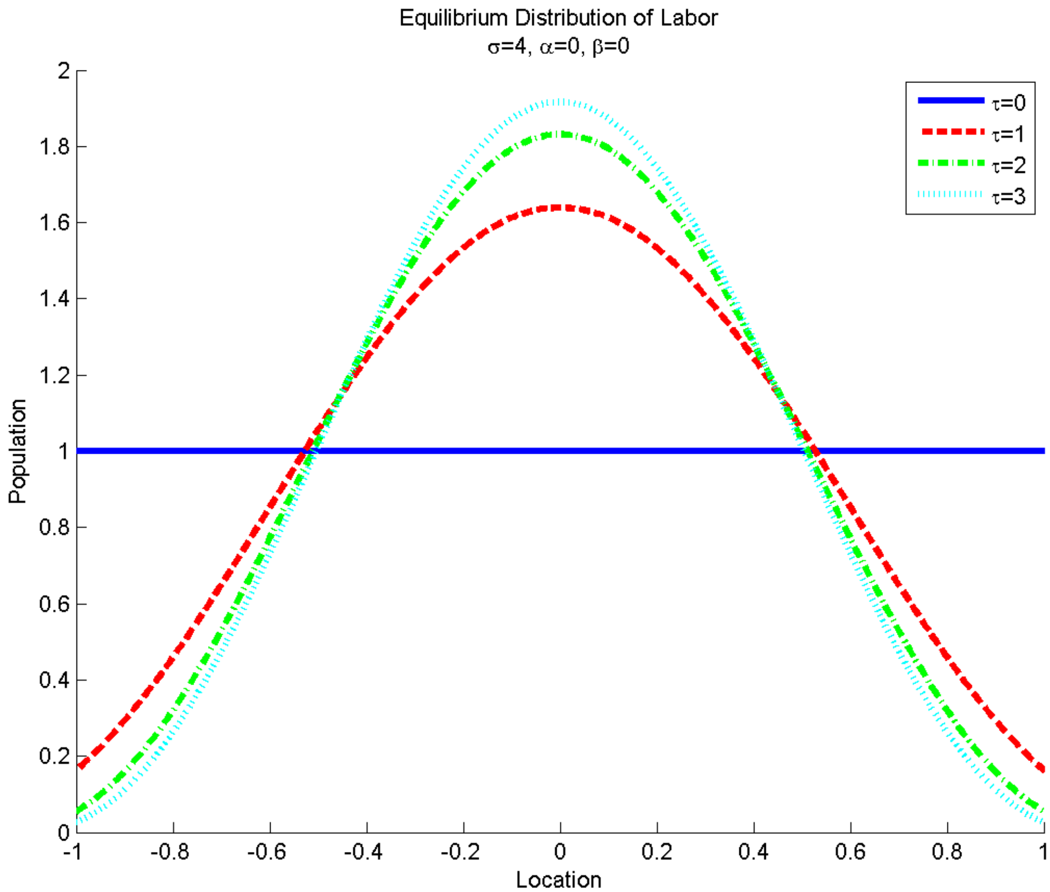
Notes: This figure shows the regions of values for the productivity spillover α and the amenity spillover β for which there exists an equilibrium, for which there exists a point-wise locally stable equilibrium, and whether that equilibrium is unique. The elasticity of substitutions σ is chosen to equal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

Figure 2: Propagation of geographic trade costs



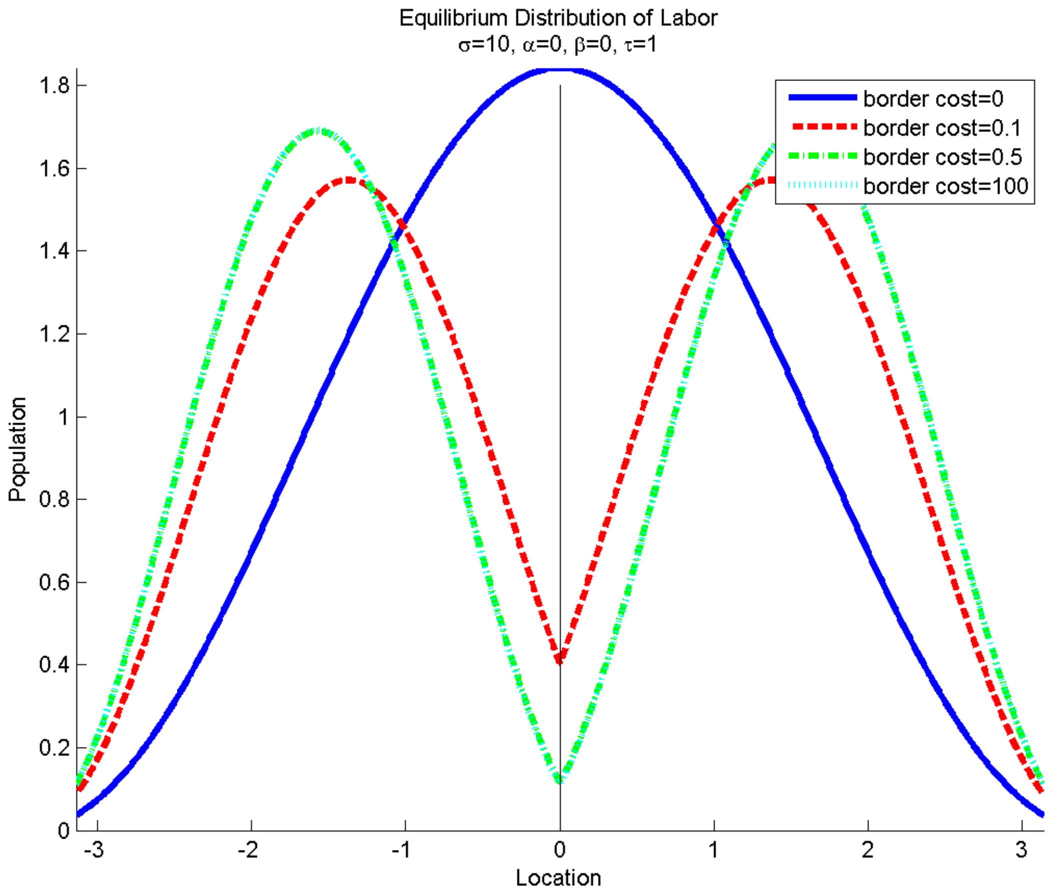
Notes: This figure shows how the geographic trade costs evolve across a surface. Given a contour of points on a surface such that the geographic trade cost to location i is equal to a constant C (the solid line), for an arbitrarily small $\varepsilon > 0$, we can construct the contour line for bilateral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

Figure 3: Economic activity on a line: Trade costs



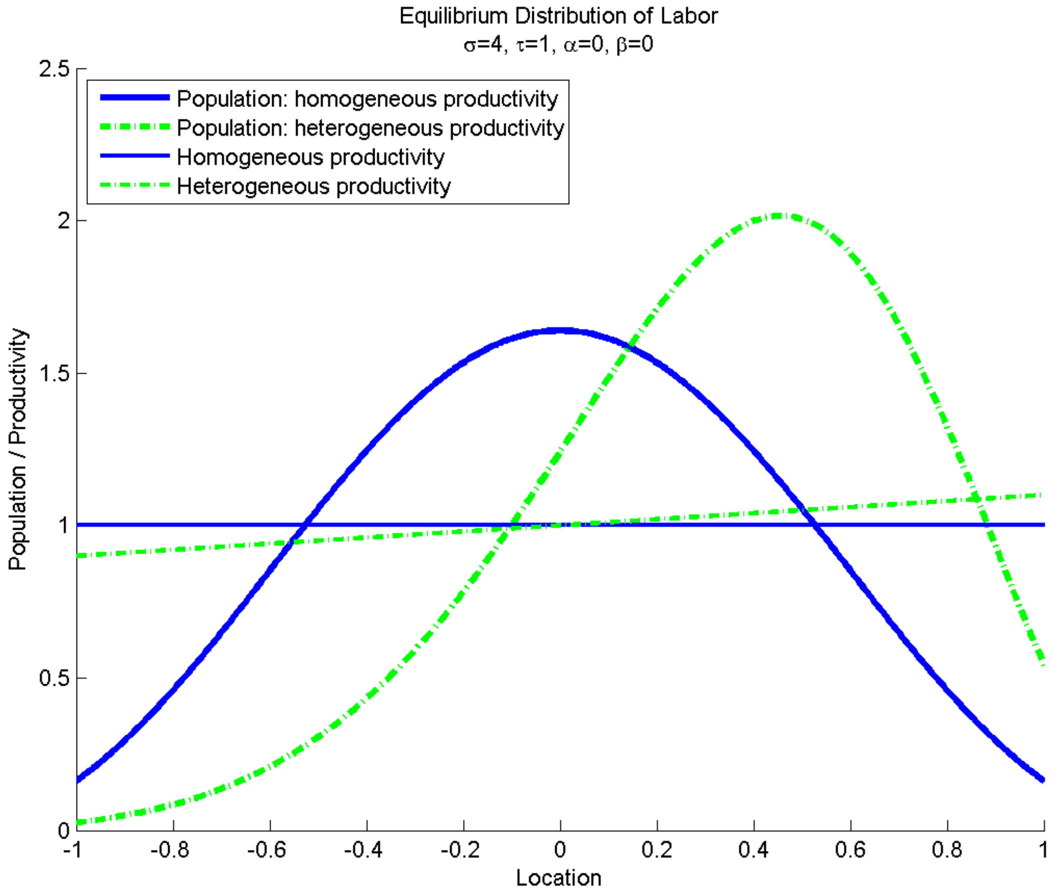
Notes: This figure shows how the equilibrium distribution of population along a line is affected by changes in the trade cost. When trade is costless, the population is equal along the entire line. As trade becomes more costly, the population becomes increasingly concentrated in the center of the line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

Figure 4: Economic activity on a line: Border costs



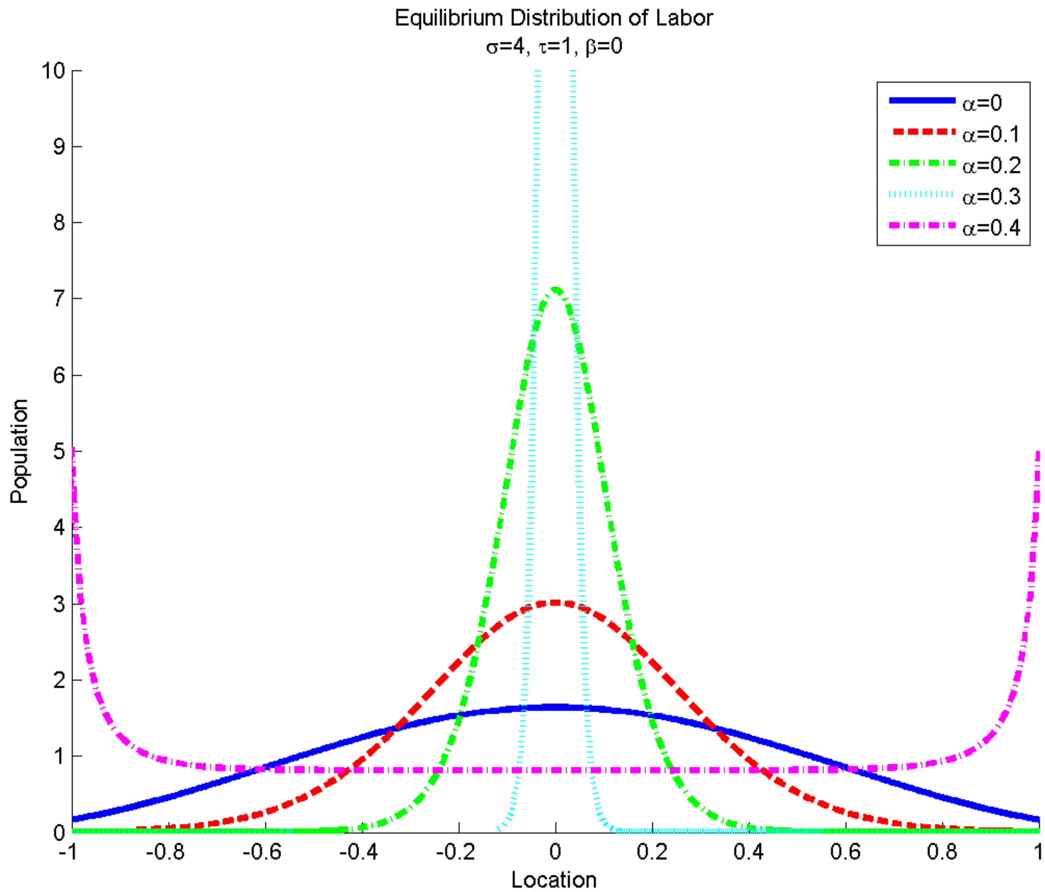
Notes: This figure shows how the equilibrium distribution of population along a line is affected by the presence of a border in the center of the line. As crossing the border becomes increasingly costly, the equilibrium distribution of population moves toward the center of each half of the line.

Figure 5: Economic activity on a line: Exogenous productivity differences



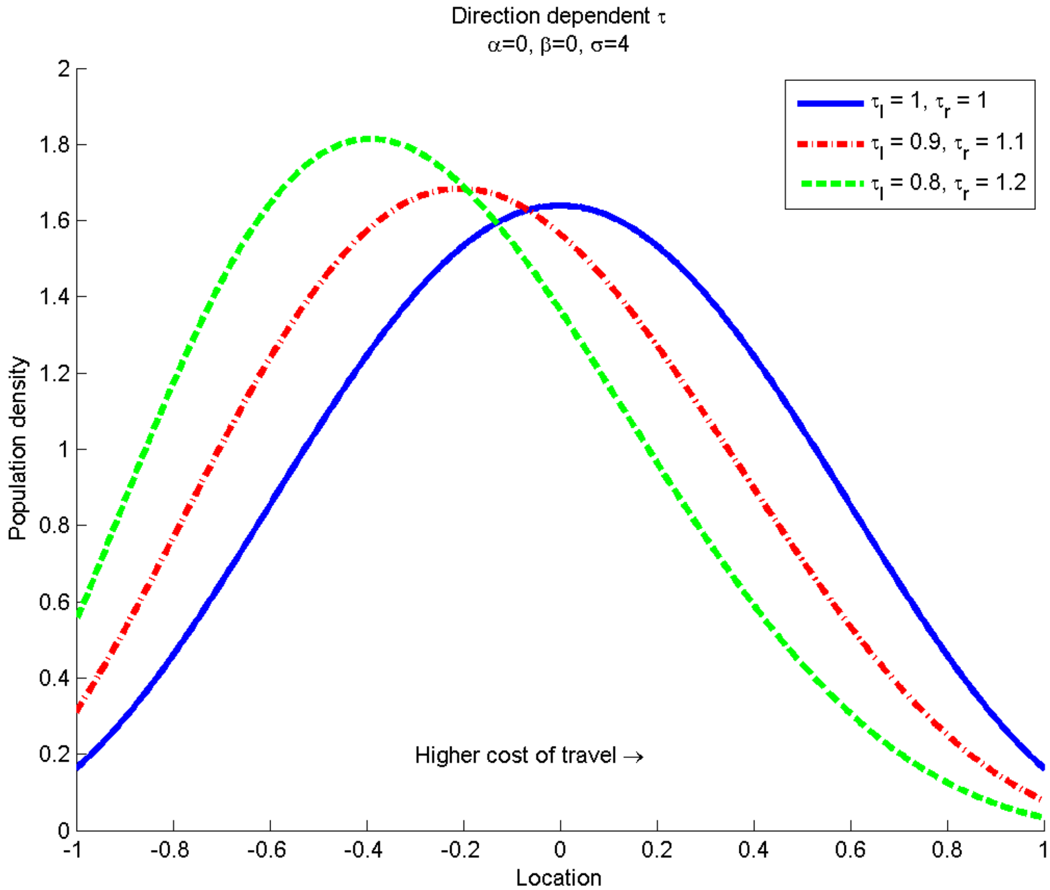
Notes: This figure depicts how the equilibrium distribution of population along a line is affected by exogenous differences in productivity across space. With homogeneous productivities, and positive trade costs, the population is concentrated at the center of the line. When productivity is higher toward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

Figure 6: Economic activity on a line: Productivity spillovers



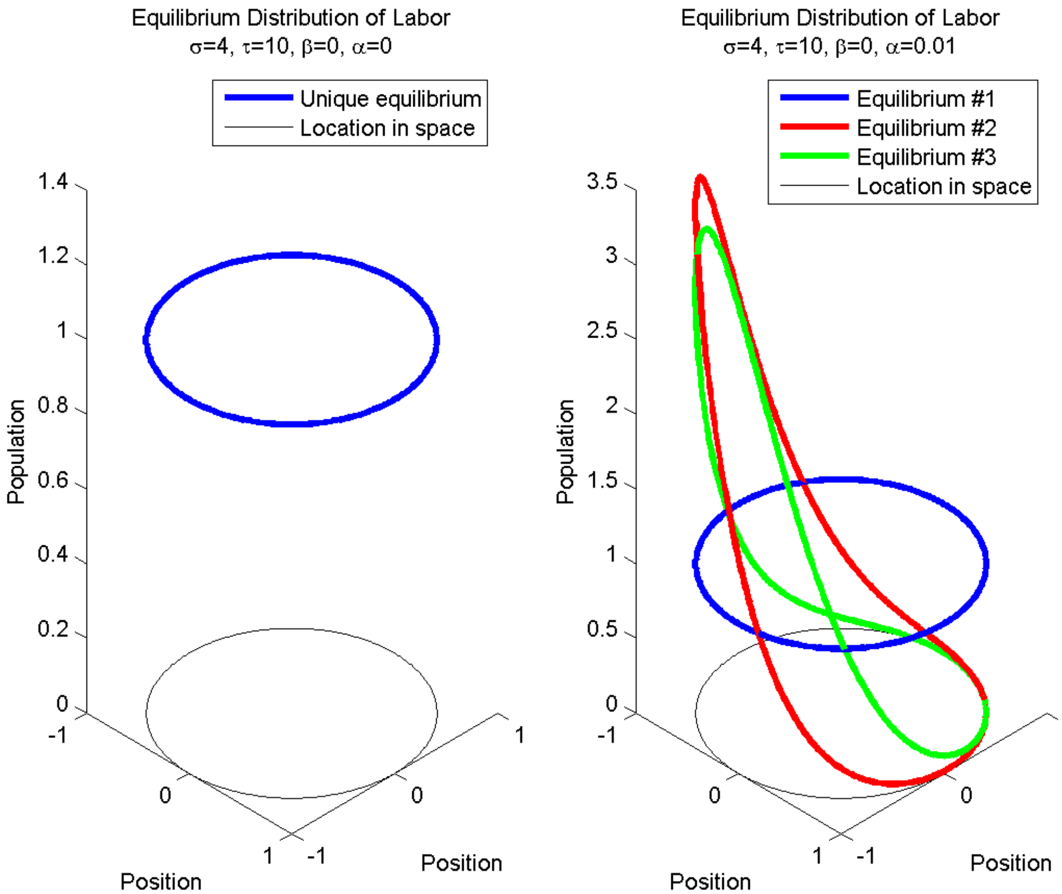
Notes: This figure shows how the equilibrium distribution of population along a line is affected by varying degrees of productivity spillovers. As the productivity spillovers increase, the population becomes increasingly concentrated in the center of the line. A non-degenerate equilibrium can be maintained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

Figure 7: Economic activity on a line: Direction of travel



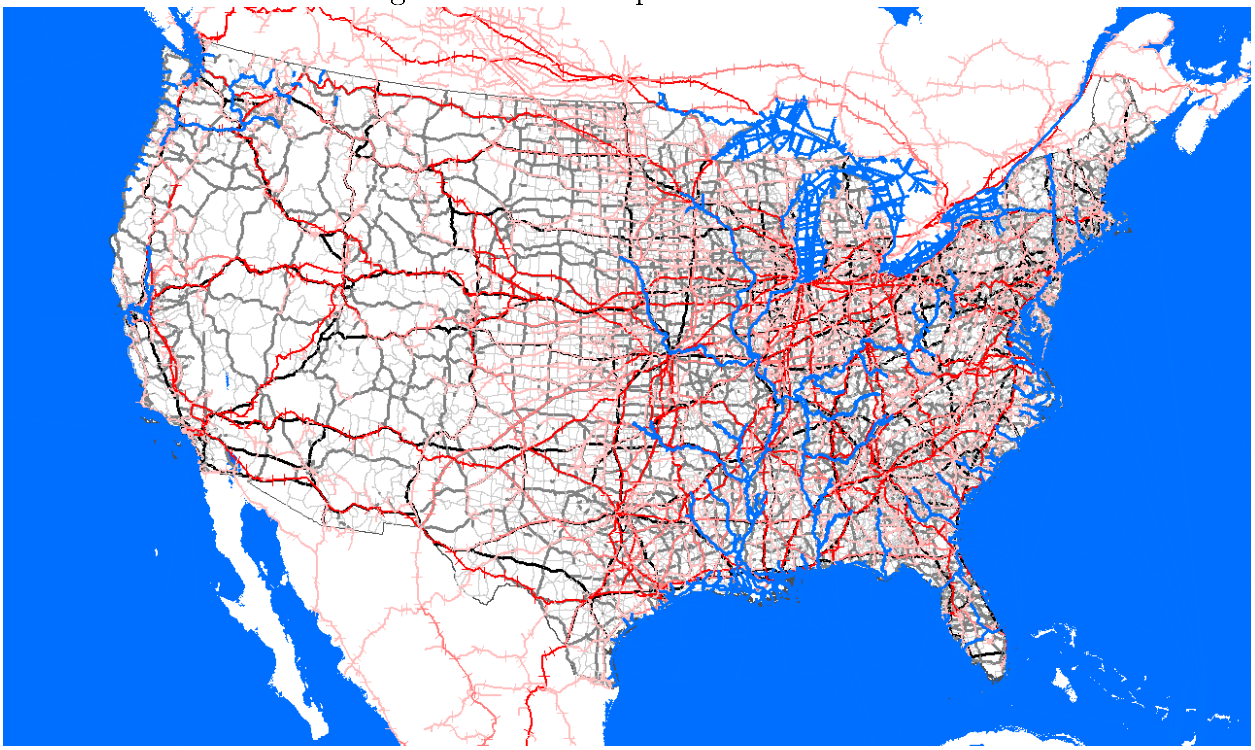
Notes: This figure shows how the equilibrium distribution of population along a line is affected by instantaneous trade costs that depend on the direction of travel. As the cost of traveling to the right becomes increasingly more expensive than traveling to the left, the equilibrium distribution of population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

Figure 8: Economic activity on a circle: Multiple equilibria



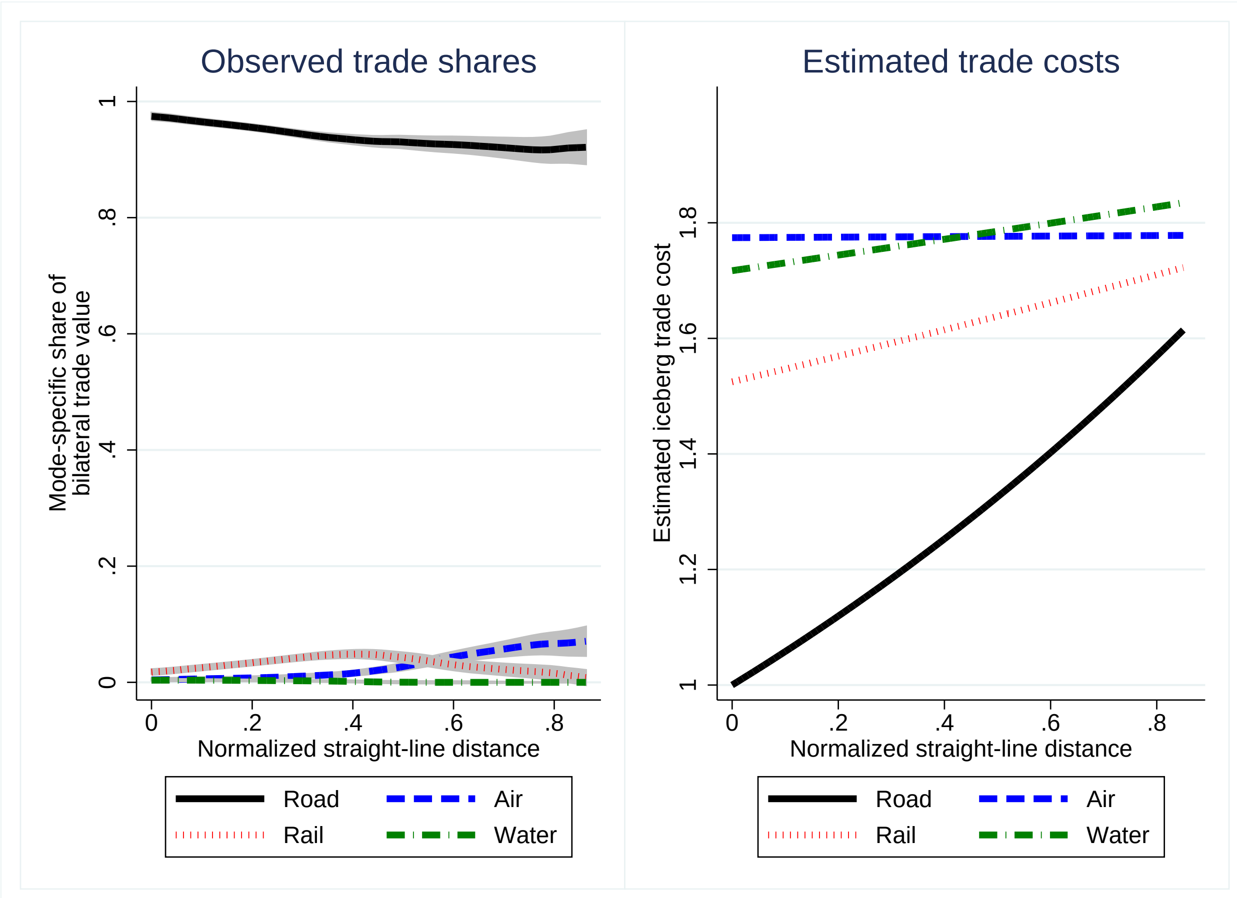
Notes: This figure provides an example of multiple equilibria when the surface is a one dimensional circle. The left panel shows the unique homogeneous distribution of population along the circle when $\alpha + \beta = 0$. When $\alpha + \beta > 0$ (here $\alpha = 0.01$ and $\beta = 0$), uniqueness is 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.

Figure 9: U.S. transportation networks



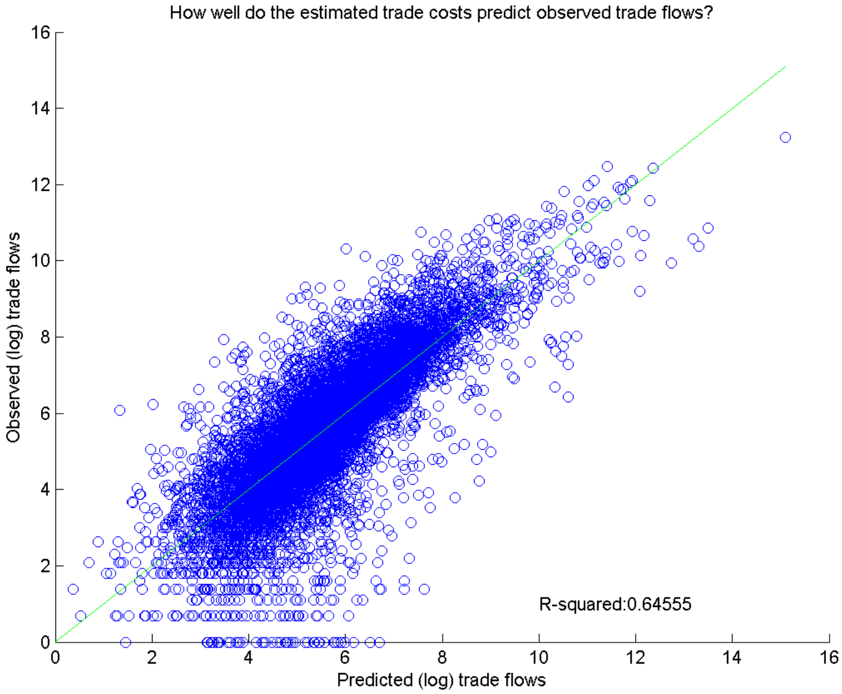
Notes: This figure shows each of the observed transportation cost networks. Interstate highways are black, other U.S. highways are dark gray, and arterial roads are light gray. Class A railroads are dark red, class B railroads are light red, and other railroads are pink. Navigable waterways are blue.

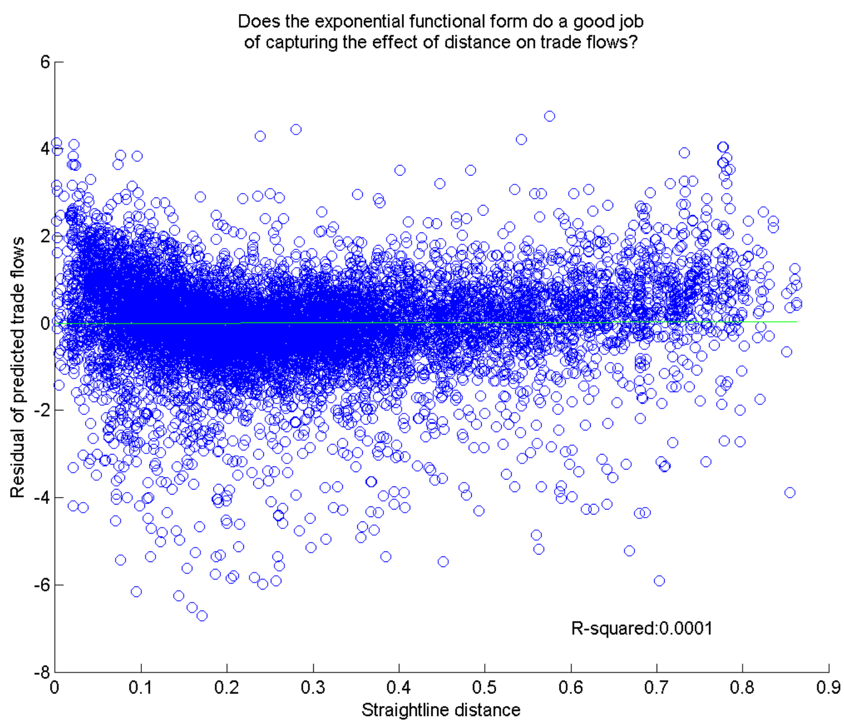
Figure 10: Mode-specific bilateral trade shares by distance



Notes: 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 non-parametric local mean smoothed regression using an Epanechnikov kernel with a bandwidth of 0.1. 99% confidence intervals are reported in grey. 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 one.

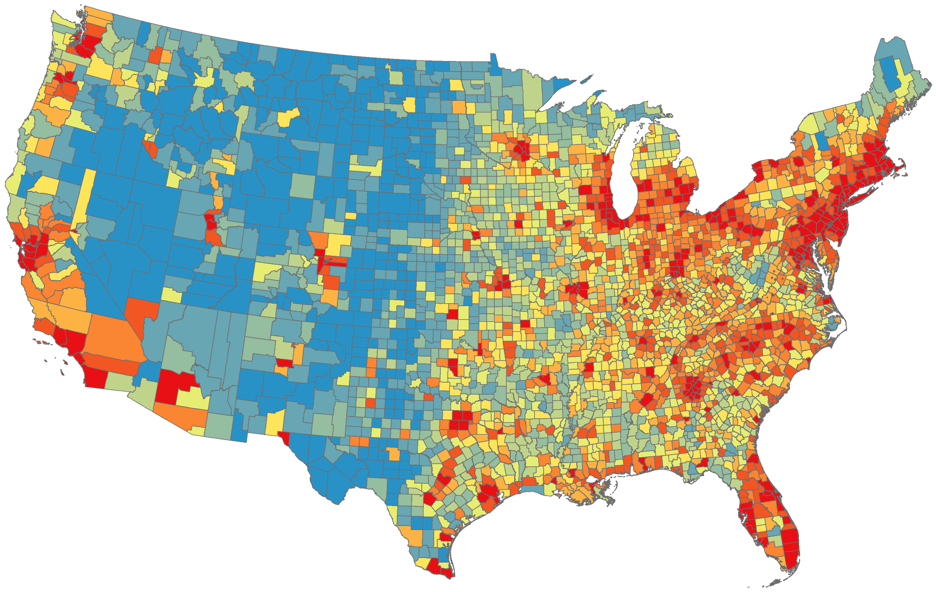
Figure 11: Assessing the predicted trade costs



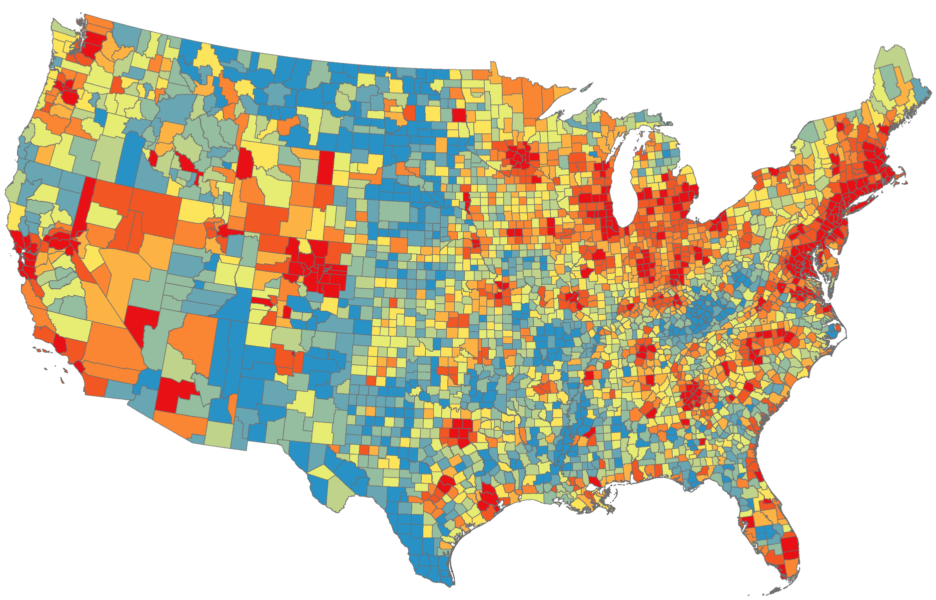


Notes: This figure assesses the quality of the estimated trade costs. The top panel compares the bilateral trade flows implied by the estimated trade costs with the bilateral trade flows observed in the 2007 Commodity Flow Survey. The bottom panel shows that the difference between the trade flows implied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

Figure 12: United States population density and wages in 2000



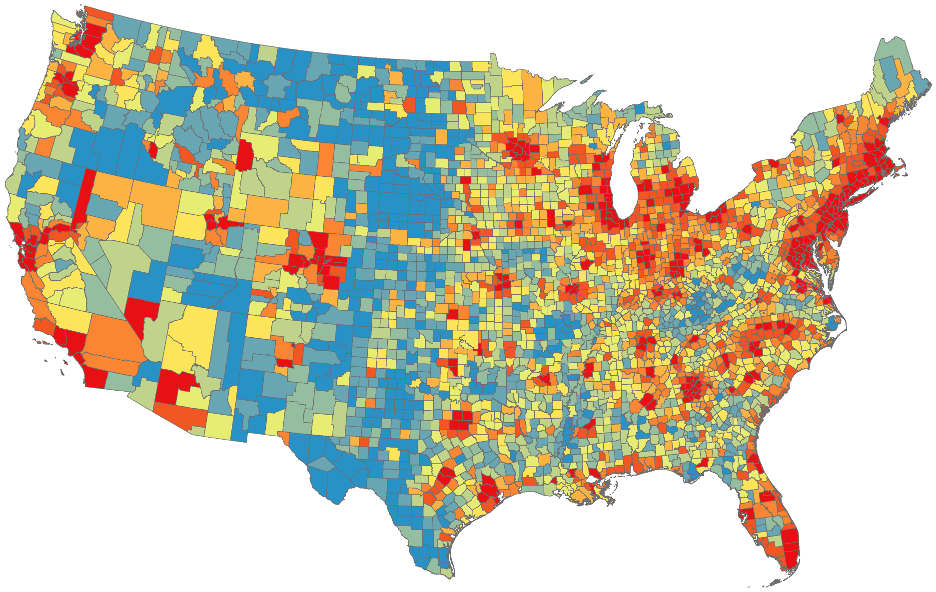
Population density



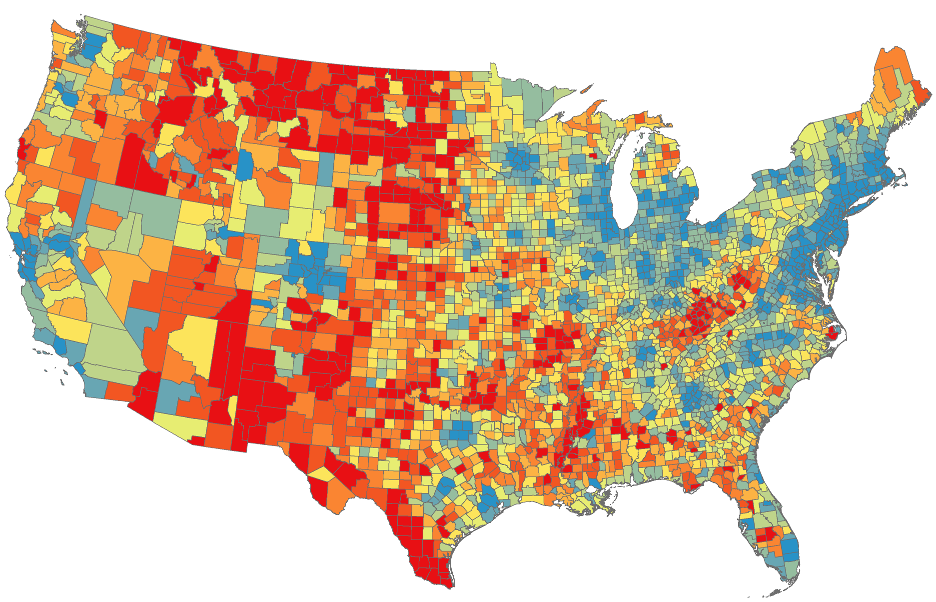
Wages

Notes: This figure shows the relative population density (top) and wages (bottom) within the United States in the year 2000 by decile. The data are reported at the county level; red (blue) indicate higher (lower) deciles. (Source: MPC (2011a)).

Figure 13: Estimated composite productivity and amenity



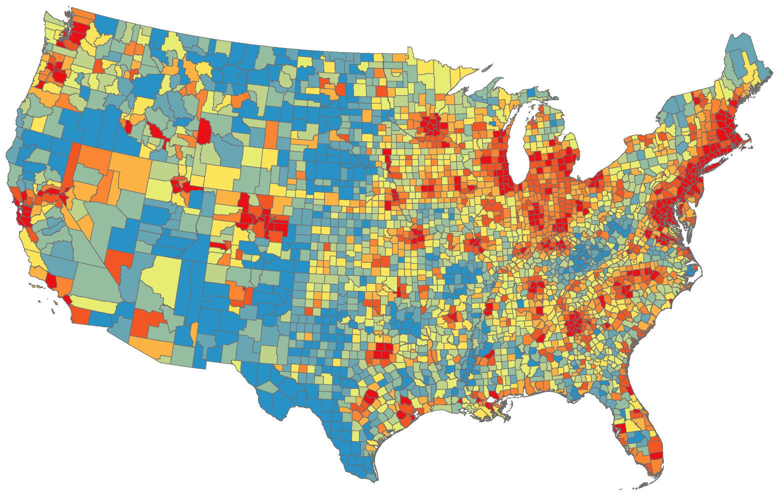
Composite productivity



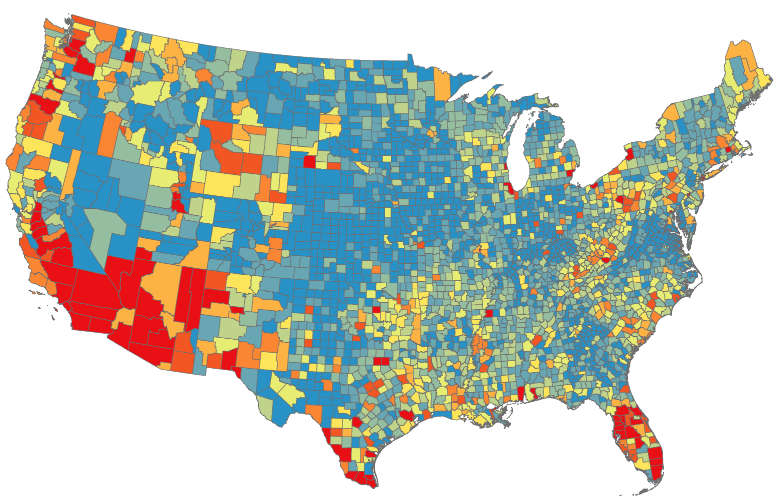
Composite amenity

Notes: This figure shows the estimated composite productivity (top) and amenity (bottom) by decile. The data are reported at the county level; red (blue) indicate higher (lower) deciles.

Figure 14: Estimated exogenous productivity and amenity



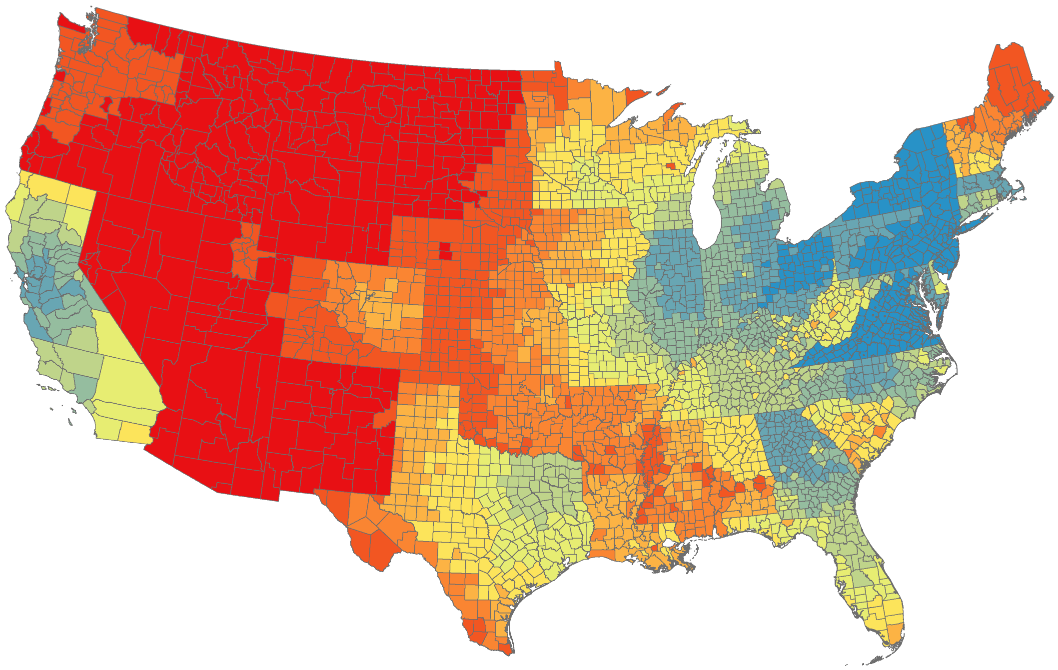
Exogenous productivity



Exogenous amenity

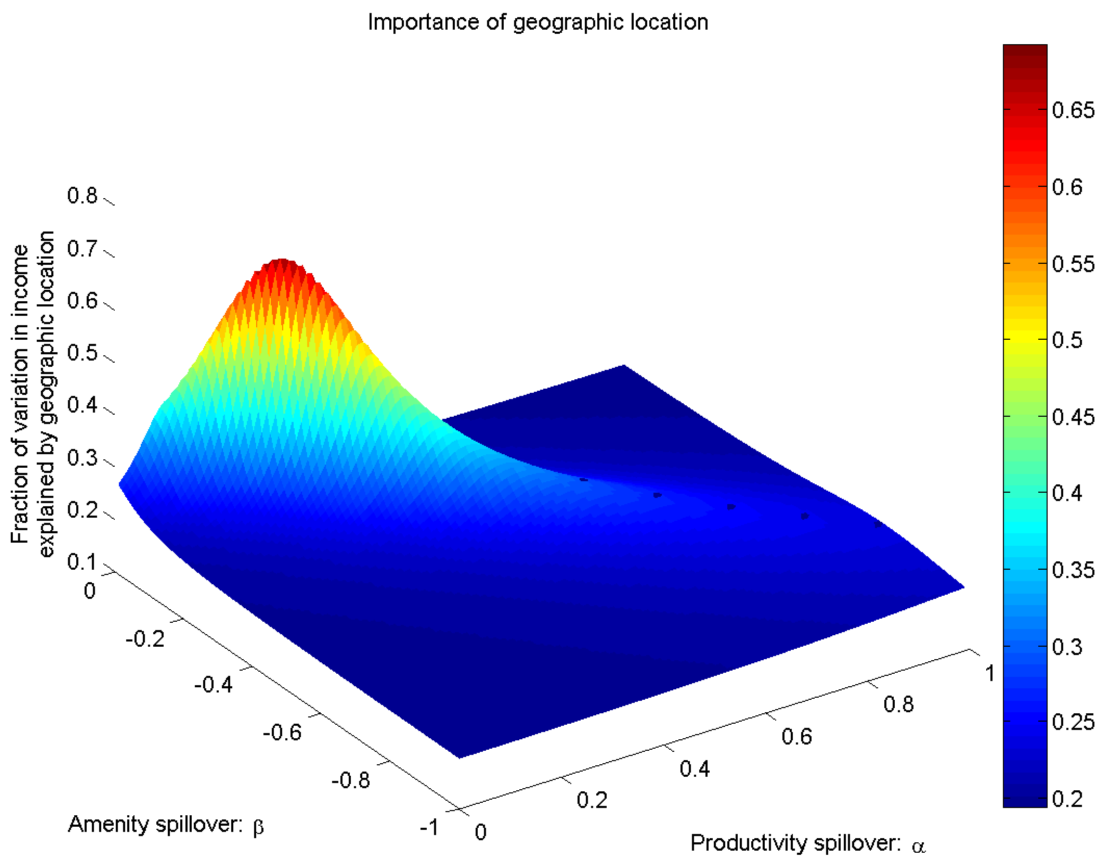
Notes: This figure shows the estimated exogenous productivity $\bar{A}$ (top) and amenity $\bar{u}$ (bottom) by decile assuming $\alpha = 0.1$ and $\beta = -0.3$. The data are reported at the county level; red (blue) indicate higher (lower) deciles.

Figure 15: Estimated price index



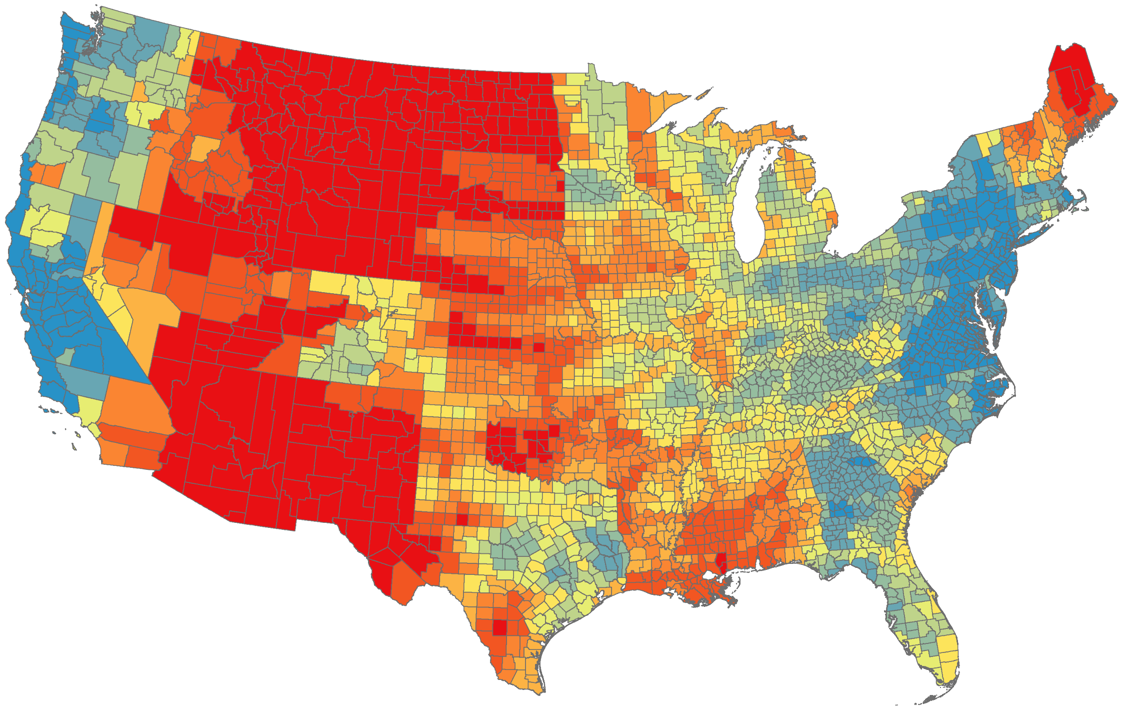
Notes: This figure shows the estimated price index by decile. The data are reported at the county level; red (blue) indicate higher (lower) deciles.

Figure 16: Fraction of spatial inequality of income due to geographic location in the United States



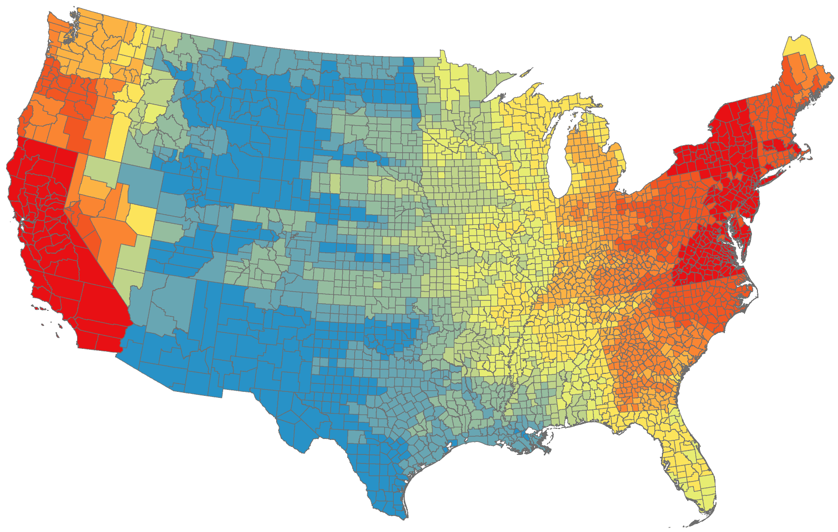
Notes: This figure shows the fraction of the observed variation in income across space in the U.S. in the year 2000 that is due to geographic location. The decomposition is calculated for all constellations of productivity spillover strength $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$.

Figure 17: Estimated increase in the price index from removing the Interstate Highway System

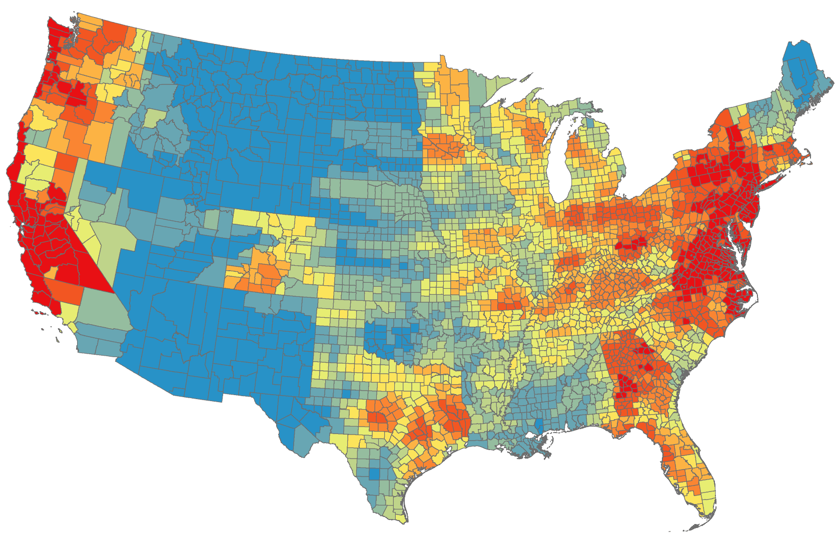


Notes: This figure depicts the estimated increase in the price index (by decile) across space from removing the Interstate Highway System (IHS), holding wages and productivities constant at the 2000 U.S. levels. Red (blue) indicate higher (lower) deciles (e.g. the removal of the IHS disproportionately increased the economic remoteness in red regions).

Figure 18: Estimated change in the population from removing the Interstate Highway System



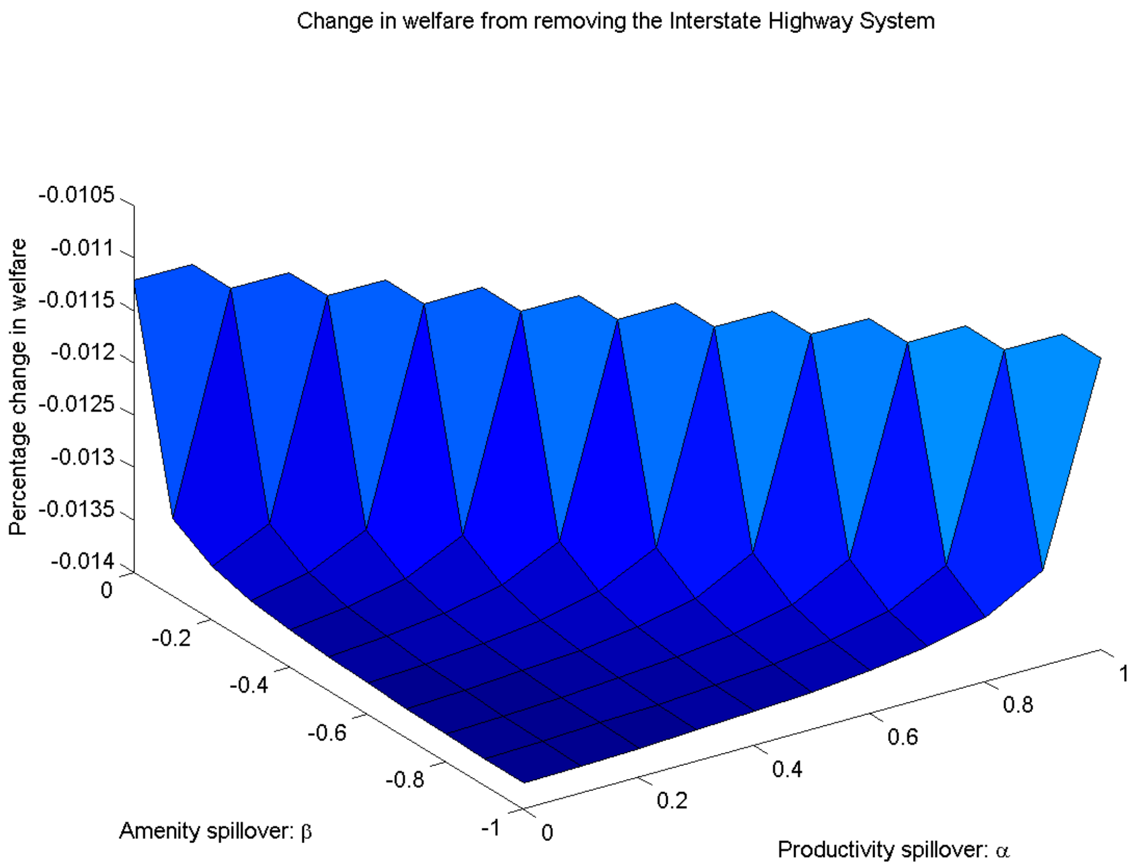
$\alpha = 0$, $\beta = 0$



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

Notes: This figure shows the estimated change in population (in deciles) from the removal of the Interstate Highway System (IHS). The top map reports the estimated population changes when there are no spillovers (i.e. $\alpha = \beta = 0$), while 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 $\beta = -0.3$). Red (blue) indicates higher (lower) deciles (e.g. the removal of the IHS increased the relative population in red areas).

Figure 19: Estimated decline in welfare from removing the Interstate Highway System



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

A Theory Appendix

This Appendix is composed of two subsections. In the first, we prove Theorems 1 and 2 and Proposition 1 regarding the existence, uniqueness, and point-wise local stability of a spatial equilibrium, as well as Theorem 3, regarding the identification of exogenous productivities and amenities. In the second, we discuss the isomorphisms existing between our framework and other spatial economic models.

A.1 Proofs of Theorems

In this section, we prove the theorems in the main text. The proofs rely heavily on results from the study of integral equations, for which Zabreyko, Koshelev, Krasnosel'skii, Mikhlin, Rakovshchik, and Stetsenko (1975) and Polyanin and Manzhirov (2008) are handy references. The proofs of the theorems apply to compact intervals $S \subset \mathbb{R}^N$ but for convenience we provide references with results for connected and compact subsets of $\mathbb{R}^N$.

A.1.1 Proof of Theorem 1

Note that when $\alpha = \beta = 0$, equation (10) can be written as:

$$g\left(i\right) = \lambda \int_S K\left(s,i\right) g\left(s\right) ds, \tag{25}$$

where $g\left(i\right) \equiv L\left(i\right) w\left(i\right)^{\sigma}$ in unknown, $K\left(s,i\right) \equiv T\left(i,s\right)^{1-\sigma} A\left(i\right)^{\sigma-1} u\left(s\right)^{\sigma-1}$ is known, and $\lambda \equiv W^{1-\sigma}$ is unknown. We can also re-write equation (11) in an identical form:

$$f\left(i\right) = \lambda \int_S K\left(i,s\right) f\left(s\right) ds, \tag{26}$$

where $f\left(i\right) \equiv w\left(i\right)^{1-\sigma}$ is unknown, $K\left(i,s\right) \equiv T\left(s,i\right)^{1-\sigma} u\left(i\right)^{\sigma-1} A\left(s\right)^{\sigma-1}$ is the transpose of $K\left(s,i\right)$, and $\lambda \equiv W^{1-\sigma}$ is unknown.

Part (i) As mentioned in the text, equations (25) and (26) are eigenfunctions. Furthermore, because in both cases all components of the Kernel are continuous and bounded above and below by a positive number, each kernel $K\left(s,i\right)$ is also continuous and bounded above and below by a positive number. As a result, by a generalization of Jentzsch's theorem (see e.g. Theorem 3 of Birkhoff (1957) where S is a Banach lattice or p.648 of Polyanin and Manzhirov (2008) where S is a connected interval of $\mathbb{R}$.[29]), there exists a unique (to-scale) strictly positive function $g(i)$ and constant λ_1 that solves equation (25) and a unique (to-scale) strictly positive function $f(i)$ and constant λ_2 that solves equation (26).

It remains to show that $\lambda_1 = \lambda_2$. From one of the Fredholm Theorems (see Theorem 1.3 on p. 31 of Zabreyko, Koshelev, Krasnosel'skii, Mikhlin, Rakovshchik, and Stetsenko (1975)[30]), because the kernel of equation (26) is the transpose of the kernel equation (25), λ_1 is a characteristic value of equation (26) and λ_2 is a characteristic value of equation (25). In addition, notice that the constants λ_1 and λ_2 correspond to the smallest characteristic values of equations (25) and (26), respectively.[31] Suppose that $\lambda_1 > \lambda_2$, we will arrive at a contradiction. In that case equation (26) has a characteristic value smaller than λ_2, which is a contradiction of Jentzsch's theorem. Similarly, we get a contradiction if we assume $\lambda_2 > \lambda_1$. Therefore, $\lambda_1 = \lambda_2$, so that there exists unique (to-scale), strictly positive functions $g(i)$ and $f(i)$ that solve equations (25) and (26). Because $g(i) \equiv L(i)\, w(i)^{\sigma}$ and $f(i) \equiv w(i)^{1-\sigma}$, wages and the labor supply can be determined (up to scale) immediately from $g(i)$ and $f(i)$.

To prove that the equilibrium is regular we need to argue that $L(i), w(i)$ are strictly positive and continuous functions for all i. The proof that all regions are populated, and thus $L(i), w(i) > 0$, is given in the proof of Theorem 2 for any $\gamma_1 > 0$. The proof of continuity is given in Part (ii) below.

Part (ii) The solution of the wages and the labor, up to scale, is the uniform limit of the successive approximation

$$f_{n+1}(i) = \frac{\int_S K(i,s)\, f_n(s)\, ds}{\int_S \int_S K(i,s)\, f_n(s)\, ds di}, \tag{27}$$

as shown by Birkhoff (1957), starting from an arbitrary guess of the function $f_0(i)$. In practice, we find that the convergence of equation (27) is rapid for both $f(i)$ and $g(i)$. Solving for equilibrium wages and the labor supply is then straightforward, as $w(i) = g(i)^{\frac{1}{1-\sigma}}$ and $L(i) = \bar{L}\frac{f(i)g(i)^{\frac{\sigma}{\sigma-1}}}{\int_S f(s)g(s)^{\frac{\sigma}{\sigma-1}} ds}$. Note that once convergence occurs, the normalization identifies

[29]Note that the compactness of S and the continuity of $K(s,i)$ are sufficient but not necessary conditions to apply Theorem 3 of Birkhoff (1957). Related to that, the boundeness of $K(s,i)$ above and below by a positive number is a stronger requirement than the linear transformation of $K(s,i)$ is uniformly bounded.

[30]Note that because λ_1 is real, the complex conjugate of λ_1, $\bar{\lambda}_1$, is equal to λ_1 (and likewise for λ_2 and $\bar{\lambda}_2$).

[31]See, for example, Krasnosel'Skii and Boron (1964) p.232. This statement derives from the results of Theorems 2.11 and 2.13, p.78 and 81, with the required conditions on the kernel stated in Theorem 2.10, page 76. The conditions require that the kernel is bounded above and below by a positive number, which we have already assumed (see Theorem 2.2).

$W^{1-\sigma}$, i.e. $W^{1-\sigma} = \bar{L}/\int_S f(s)\, g(s)^{\frac{\sigma}{\sigma-1}}\, ds$.

Notice that if we start with a continuous guess $f_0(i)$ the operator (27) is continuous and thus $\{f_n(i)\}_{n\in\mathbb{N}}$ is a sequence of continuous functions. By the uniform convergence theorem and the uniform limit result above the limit of this sequence is also continuous and thus $f(i)$ is also continuous. Since we proved that the equilibrium solutions are positive and continuous, we proved that the equilibrium is regular, completing the proof of the theorem.

A.1.2 Proof of Theorem 2

We first show that if there exists a regular spatial equilibrium, then equation (12) is the unique relationship between $w(i)$ and $L(i)$ that satisfies equations (10) and (11). Suppose there exists a regular spatial equilibrium, i.e. there exists continuous functions $w(i)$ and $L(i)$ bounded above and below by positive numbers that satisfy equations (10) and (11). Define the function $\phi : S \to \mathbb{R}_+$ as follows:

$$\phi(i) \equiv \frac{L(i)^{1-\alpha(\sigma-1)}\, w(i)^{\sigma}\, \bar{A}(i)^{1-\sigma}}{w(i)^{1-\sigma}\, \bar{u}(i)^{1-\sigma}\, L(i)^{\beta(1-\sigma)}}.$$

Note that $\phi(i)$ is positive, continuous, and bounded above and below by strictly positive numbers. Suppose too that $T(i,s) = T(s,i)$. Then from equations (10) and (11) we have:

$$\begin{aligned}
\phi(i) &= \frac{\int_S T(i,s)^{1-\sigma}\, \bar{u}(s)^{\sigma-1}\, L(s)^{1+\beta(\sigma-1)}\, w(s)^{\sigma}\, ds}{\int_S T(s,i)^{1-\sigma}\, \bar{A}(s)^{\sigma-1} w(s)^{1-\sigma}\, L(s)^{\alpha(\sigma-1)}\, ds} \iff \\
\phi(i) &= \frac{\int_S F(s,i)\phi(s)^{\beta}\, ds}{\int_S F(s,i)\phi(s)^{\beta-1}\, ds}, \qquad (28)
\end{aligned}$$

where $F(s,i) \equiv T(i,s)^{1-\sigma}\, \bar{u}(s)^{(1-\beta)(\sigma-1)}\, \bar{A}(s)^{\beta(\sigma-1)}\, L(s)^{1+\beta(\sigma-1)+\beta((\alpha-\beta)(\sigma-1)-1)}\, w(s)^{\sigma+\beta(1-2\sigma)}$ and we use the assumed symmetry of trade costs, i.e. $T(i,s) = T(s,i)$. Note that $F(s,i)$ is positive, continuous, and bounded above and below by strictly positive numbers. We can then write (28) as

$$\frac{\phi(i)^{\beta}}{\int_S F(s,i)\phi(s)^{\beta}\, ds} = \frac{\phi(i)^{\beta-1}}{\int_S F(s,i)\phi(s)^{\beta-1}\, ds}. \qquad (29)$$

Define $\lambda(i) \equiv \frac{\phi(i)^{\beta}}{\int_S F(s,i)\phi(s)^{\beta} ds}$. Note that because $\phi(i)$ is positive, continuous, and bounded above and below by strictly positive numbers, so too is $\gamma(i)$. Define functions $g_1(i) \equiv \phi(i)^{\beta}$

and $g_2(i) \equiv \phi(i)^{\beta-1}$. Then we can rewrite equation (29) as the following set of equations:

$$g_1(i) = \int_S \lambda(i) F(s,i) g_1(s) ds \tag{30}$$

$$g_2(i) = \int_S \lambda(i) F(s,i) g_2(s) ds \tag{31}$$

Because $\lambda(i) F(s,i)$ is positive, continuous, and bounded above and below by strictly positive numbers, the generalized Jentzsch theorem implies that there exists a unique (to-scale) strictly positive function that satisfies both equations (30) and (31), i.e. $g_1(i) = Cg_2(i)$, where C is a constant. As a result, $\phi(i)^{\beta} = C\phi(i)^{\beta-1}$, or equivalently, $\phi(i) = C$. Substituting in the definition of $\phi(i)$ into $\phi(i) = C$ immediately yields equation (12). Hence equation (12) is the unique relationship between $w(i)$ and $L(i)$ that satisfies equations (10) and (11) for a regular spatial equilibrium.

Since equation (12) holds for any regular equilibrium, it is sufficient to consider it along with equation (13) to determine existence and uniqueness of a regular equilibrium rather than equations (10) and (11) directly. Note that we can rewrite equation (13) as a nonlinear integral equation

$$f(i) = \lambda \int_S K(s,i) f(s)^{\frac{\gamma_2}{\gamma_1}} ds, \tag{32}$$

where $f(i) \equiv L(i)^{\gamma_1}$, $\lambda = W^{1-\sigma}$, and

$$K(s,i) \equiv \bar{u}(i)^{(1-\widetilde{\sigma})(\sigma-1)} \bar{A}(i)^{\widetilde{\sigma}(\sigma-1)} T(s,i)^{1-\sigma} \bar{A}(s)^{(1-\widetilde{\sigma})(\sigma-1)} \bar{u}(s)^{\widetilde{\sigma}(\sigma-1)}.$$

In fact, instead of characterizing (32) it suffices to find the solution for the combined variable $\widetilde{f}(i) = f(i) \lambda^{\frac{1}{\frac{\gamma_2}{\gamma_1}-1}}$. To see this, notice that

$$\widetilde{f}(i) = \int_S K(s,i) \left[\tilde{f}(s)\right]^{\frac{\gamma_2}{\gamma_1}} ds \iff \tag{33}$$
$$f(i) \lambda^{\frac{1}{\frac{\gamma_2}{\gamma_1}-1}} = \int_S K(s,i) \left[f(s) \lambda^{\frac{1}{\frac{\gamma_2}{\gamma_1}-1}}\right]^{\frac{\gamma_2}{\gamma_1}} ds,$$

which is equivalent to (32). In our case $f(i) \equiv L(i)^{\gamma_1}$ and the labor market clearing constraint implies

$$\bar{L} = \lambda^{-\frac{1}{\gamma_2-\gamma_1}} \int_S \widetilde{f}(s)^{1/\gamma_1} ds, \tag{34}$$

and thus, for each solution for $\widetilde{f}(s)$, a unique solution for λ.[32] Therefore, finding a solution for $\widetilde{f}(i)$ gives us the solution for $f(i)$ and the eigenvalue of the system λ, which in our case is inversely related to welfare. Given the above preliminaries we proceed to prove the different parts of Theorem 2.

Part (i) We first prove existence of a regular spatial equilibrium. To do so we can directly use Theorem 2 of Karlin and Nirenberg (1967) to establish existence for Equation (32). Their Theorem 2 shows the existence of a continuous solution $f(i)$ for a Hammerstein equation of the second kind

$$f(i) = \lambda \int_S K(s,i)\, \phi(s, f(s))\, ds,$$

where ϕ is a continuous function and where the bounds of integration are given by the min and the max of $K(s,i) / F(K(.,i))$ with $F(K(.,i))$ an arbitrary linear functional such that $F(f) = 1$ and $F(K(\cdot,s)) > 0$ for all $s \in S$. Furthermore, the solution $f(i)$ is bounded below by $a \equiv \min_{i,s\in S} \frac{K(i,s)}{F(K(\cdot,s))} > 0$ and bounded above by $b \equiv \max_{i,s\in S} \frac{K(i,s)}{F(K(\cdot,s))} > 0$. For our purposes, $\phi(s, f(s)) = f(s)^{\frac{\gamma_2}{\gamma_1}}$ and $F(f) \equiv \frac{1}{L}\int_S f(s)^{\frac{1}{\gamma_1}} ds$. Note both that $F(f) = 1$ from labor market clearing and $F(K(\cdot,s)) = \frac{1}{L}\int K(i,s)^{\frac{1}{\gamma_1}} ds > 0$ for all $s \in S$ since $K(s,i)$ is bounded above and below by a positive number, so the theory applies. Note too that for Karlin and Nirenberg (1967) $S = [0,1]$. However, as they point out, and as is easily verified from the steps of the proof of their Theorem 2, their result applies for any domain in $\mathbb{R}^N$ and, thus, for a compact interval, which completes the proof of existence.[33]

Part (ii) To prove that for $\gamma_1 > 0$ all equilibria are regular notice that we need to prove that all locations are inhabited in equilibrium and that the equilibrium wages and labor functions are continuous. For the first part, note that substituting the gravity equation (equation (3)) and the market clearing condition (equation (6)) into the indirect utility function (equation (5)) yields:

$$W(i) = \frac{\left(\int_S T(i,s)^{1-\sigma} P(s)^{\sigma-1} w(s) L(s)\, ds\right)^{\frac{1}{\sigma}}}{P(i)} \bar{A}(i)^{\frac{\sigma-1}{\sigma}}\, \bar{u}(i)\, L(i)^{-\frac{\gamma_1}{\sigma}}. \tag{35}$$

[32] Using the latest formulation it is easy to show that increasing $\bar{L}$ does not affect the distribution of labor across locations. In particular, an increase in $\bar{L}$ does not affect $\widetilde{f}(i)$, given equation (33), and thus translates only to a change in overall welfare, $\lambda = W^{1-\sigma}$, but not to a change in the distribution of labor across locations.

[33] The proof involves constructing a compact operator that maps the convex set of all continuous functions $f(s)$ into itself and consequently applying Schauder's fixed point theorem. These steps do not depend on the domain of the integration.

Notice that if $\gamma_1 > 0$ expression (35) guarantees that every location is populated: the utility of moving to an uninhabited location is infinite. To show that every equilibrium is continuous, we need to prove that for any $\epsilon > 0$ there exists a $\delta > 0$ such that $\|s_1 - s_2\| < \delta$ implies that that $\left|\tilde{f}(s_1) - \tilde{f}(s_2)\right| < \epsilon$, i.e. $\tilde{f}$ is continuous, and thus f is continuous. Assume that we have an equilibrium with every location population and a resulting eigenvalue λ, we will establish continuity. Note first that $K(s,i)$ is assumed to be continuous and S is compact, so that by the Heine-Cantor theorem, $K(s,i)$ is uniformly continuous on S. Then for any $\epsilon > 0$ there exists a $\delta > 0$ so that $\|s_1 - s_2\| < \delta$ implies $|K(s,s_1) - K(s,s_2)| < \epsilon$ with $s_1, s_2 \in S$. Suppose $\|s_1 - s_2\| < \delta$. Then we have:

$$\begin{aligned}
\left|\tilde{f}(s_1) - \tilde{f}(s_2)\right| &= \left|\int_S \left(K(s,s_1) - K(s,s_2)\right)\tilde{f}(s)^{\frac{\gamma_2}{\gamma_1}}\,ds\right| \\
&\leq \int_S \left|K(s,s_1) - K(s,s_2)\right|\tilde{f}(s)^{\frac{\gamma_2}{\gamma_1}}\,ds \\
&\leq \epsilon \int_S \tilde{f}(s)^{\frac{\gamma_2}{\gamma_1}}\,ds \\
&\leq \epsilon \left(\int_S \tilde{f}(s)^{\frac{1}{\gamma_1}}\,ds\right)^{\gamma_2} |S|^{1-\gamma_2} \\
&\leq \epsilon \bar{L}^{\gamma_2}\lambda^{\frac{\gamma_2}{\gamma_2-\gamma_1}}|S|^{1-\gamma_2}
\end{aligned}$$

where the second to last line used Holder's inequality, the last line used equation (34), and $|S| \equiv \int_S ds$. Hence for any $\epsilon > 0$, we can choose a $\delta > 0$ such that $\|s_1 - s_2\| < \delta$ implies that $|K(s,s_1) - K(s,s_2)| < \frac{\epsilon}{\bar{L}^{\gamma_2}\lambda^{\frac{\gamma_2}{\gamma_2-\gamma_1}}|S|^{1-\gamma_2}}$ and thus $\left|\tilde{f}(s_1) - \tilde{f}(s_2)\right| < \epsilon$, establishing continuity.

Part (iii) We now prove uniqueness of a regular equilibrium when $|\frac{\gamma_2}{\gamma_1}| \leq 1$.

We already discussed the case $\gamma_2 = \gamma_1$ (which would occur if $\alpha + \beta = 0$).

Next suppose instead that $|\frac{\gamma_2}{\gamma_1}| < 1$. In this case, we can apply Theorem 2.19 from Zabreyko, Koshelev, Krasnosel'skii, Mikhlin, Rakovshchik, and Stetsenko (1975) (p.401), which states that if i) $K(i,s)$ is positive and continuous and ii) $\widetilde{f}(s)$ is strictly positive and it is non-decreasing and $\widetilde{f}(s)/s^c$ is non-increasing for $c \in (0,1)$ or it is non-increasing and $\widetilde{f}(s)s^c$ is increasing for $c \in (0,1)$ then there there exists a unique positive solution to equation (33). Furthermore, that solution is the uniform limit of the successive approximations:

$$\widetilde{f}_{n+1}(i) = \int_S K(s,i)\,\widetilde{f}_n(s)^{\frac{\gamma_2}{\gamma_1}}\,ds. \tag{36}$$

for any arbitrary non-zero, non-negative $\widetilde{f}_0(i)$.[34] Notice that i) is satisfied given the restrictions on $K(i,s)$ and $\widetilde{f}(s)^{\frac{\gamma_2}{\gamma_1}}$ satisfies ii) and in particular with $\frac{\gamma_2}{\gamma_1} \in [0,1)$, the first condition or with $\frac{\gamma_2}{\gamma_1} \in (-1,0)$ the second condition. Thus, there exists a unique positive function $\widetilde{f}(i)$ that solves (33) and uniqueness is proved when $|\frac{\gamma_2}{\gamma_1}| < 1$

Finally, assume that $\frac{\gamma_2}{\gamma_1} = -1$. For this case Remark 1 of Karlin and Nirenberg (1967) implies that as long as $K(s,i)$ continuous, non-negative and $K(i,i) > 0$, there exists a unique continuous and positive function $f(i)$ that satisfies (32) for $\frac{\gamma_2}{\gamma_1} = -1$, in the case of $S = [0,1]$. Their argument for uniqueness trivially extends to any compact subset of S of $\mathbb{R}^N$.

It remains to characterize the parameter space when $|\frac{\gamma_2}{\gamma_1}| \leq 1$ in terms of α, β and σ. Recall that $\frac{\gamma_2}{\gamma_1} = \frac{1+\alpha\sigma+(\sigma-1)\beta}{1-\alpha(\sigma-1)-\beta\sigma}$; it can then be shown that:

$$
\begin{aligned}
|\frac{\gamma_2}{\gamma_1}| \leq 1 \iff & \left(\beta+\alpha < \frac{1+\alpha}{\sigma} \text{ and } \beta+\alpha \leq \min\{2(1+\alpha), 0\}\right) \text{ if } \gamma_1 > 0 \text{or} \\
& \left(\beta+\alpha > \frac{1+\alpha}{\sigma} \text{ and } \beta+\alpha \geq \max\{2(1+\alpha), 0\}\right) \text{ if } \gamma_1 < 0.
\end{aligned}
$$

Note that if $\gamma_1 > 0$ (which implies $\frac{1+\alpha}{\sigma} > \beta + \alpha$) and $\alpha \geq 0$ then a sufficient condition for uniqueness is that $\beta + \alpha \leq 0$, as we discuss in the main text.

A.1.3 Proofs of Theorems 1 and 2 for a discrete number of locations

Theorem 1 This theorem extends in straightforward manner to the case of a discrete number of locations. The analogous result to Jentzsch's theorem for matrices – related to the case of a discrete number of locations – is the celebrated Perron-Frobenius theorem (in fact, Jentzsch's theorem is a generalization of Perron-Frobenius theorem, which regards matrices and eigenvectors, for continuous kernels and eigenfunctions). The analogous result to the Fredholm theorem used comes from the fact that for a square matrix its eigenvalue is the same as the eigenvalue of its transpose. Finally, the matrix operator in this case is ergodic in the sense that an iterative approach as the one in (27) converges to the true solution. The algorithm in this case is the same as for the case of continuous variables.

Theorem 2, part (i) We first prove existence of a regular spatial equilibrium. To do so we can directly use Theorem 1.1 of Krasnosel'skii, Armstrong, and Burlak (1964) to establish

[34] Notice that there is a typo in the statement of the second condition of Theorem 2.19 in Zabreyko, Koshelev, Krasnosel'skii, Mikhlin, Rakovshchik, and Stetsenko (1975). A statement of the Theorem for a connected compact interval in $\mathbb{R}$ is given by Polyanin and Manzhirov (2008) p. 831.

existence of a regular equilibrium for Equation (32). The Theorem states that for Γ to be the boundary of an open set defined on a Banach space of positive functions, define a positive continuous and compact operator $\mathbf{T}$ (i.e. one that maps positive, continuous and compact functions to positive continuous and compact functions) for which

$$\inf_{x\in\Gamma} \|\mathbf{T}x\| > 0$$

Then the operator has at least one positive eigenvector $\mathbf{x_0}$ such that

$$\mathbf{T}x_0 = \lambda x_0$$

to which it corresponds a positive eigenvalue λ.

Our assumptions guarantee that the Kernel of the integral equation is bounded, continuous and positive and thus the operator is positive, continuous and compact. In addition, given that the operator is bounded below and transforms positive functions to functions that are strictly bounded above zero satisfying the requirements of the Theorem. Thus, there exists at least one positive eigenvector (eigenfunction) and a corresponding eigenvalue that solve Equation (32).

Theorem 2, parts (ii) and (iii) First notice that for $\gamma_1 > 0$ the same argument as in the continuum case establishes that no location will be inhabited in a spatial equilibrium. The uniqueness proof for Theorem 2 also applies to the case of a discrete number of locations for $\gamma_2/\gamma_1 \in (0,1]$. In particular, Fujimoto and Krause (1985) show that any operator T that is strictly increasing and satisfies $T(\lambda x) = f(\lambda) T(x)$ with $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f(\lambda)/\lambda$ is non-increasing and $f(0) = 0$, has a unique positive solution and is strongly ergodic. Our operator in the discrete case is $T(f) = \sum_s K(s,i) f(s)^{\frac{\gamma_2}{\gamma_1}}$ and $\gamma_2/\gamma_1 \in (0,1]$ all these restrictions on the operator are satisfied, proving the result. Finally for $\gamma_2/\gamma_1 \in [-1,0)$ the theorem of Karlin and Nirenberg (1967) directly applies for the discrete case as in Part iii of Theorem 2, which completes the uniqueness proof.

A.1.4 Proof of Proposition 1

Consider a regular equilibrium satisfying equations (10) and (11). Taking the derivative of welfare in location i with respect to the population in location i from equation (35) yields:

$$\frac{dW(i)}{dL(i)} = -\frac{\gamma_1}{\sigma}\left(\frac{\left(\int_S T(i,s)^{1-\sigma} w(s) L(s)\, ds\right)^{\frac{1}{\sigma}}}{P(i)} \bar{A}(i)^{\frac{\sigma-1}{\sigma}} \bar{u}(i) L(i)^{-\frac{\gamma_1}{\sigma}-1}\right),$$

since changes in the population in location i do not affect $\frac{\left(\int_S T(i,s)^{1-\sigma} w(s) L(s) ds\right)^{\frac{1}{\sigma}}}{P(i)}$ because location i has zero measure. As a result:

$$sign\left(\frac{dW(i)}{dL(i)}\right) = -sign(\gamma_1).$$

From the definition of point-wise local stability, it immediately follows that if $\gamma_1 < 0$, the equilibrium is point-wise locally stable and if $\gamma_1 > 0$ the equilibrium is point-wise locally stable, thereby proving the Proposition.

A.1.5 Proof of Theorem 3

Combining the price index (equation (4)) and the indirect utility function (equation (5)) and imposing utility equalization yields:

$$u(i)^{1-\sigma} = W^{1-\sigma} \int_S T(s,i)^{1-\sigma} w(i)^{\sigma-1} w(s)^{1-\sigma} A(s)^{\sigma-1} ds. \tag{37}$$

In addition, equation (12) can be rearranged as follows:

$$A(s)^{\sigma-1} = \frac{1}{\phi} L(s) w(s)^{2\sigma-1} u(s)^{\sigma-1}. \tag{38}$$

Substituting equation (38) into equation (37) yields:

$$u(i)^{1-\sigma} = \frac{W^{1-\sigma}}{\phi} \int_S T(s,i)^{1-\sigma} w(i)^{\sigma-1} w(s)^{\sigma} L(s) u(s)^{\sigma-1} ds. \tag{39}$$

Define the functions $f(i) \equiv u(i)^{1-\sigma}$ and $K(s,i) \equiv \frac{W^{1-\sigma}}{\phi} T(s,i)^{1-\sigma} w(i)^{\sigma-1} w(s)^{\sigma} L(s)$. Then equation (39) can be rewritten as:

$$f(i) = \int_S K(s,i) \frac{1}{f(s)} ds. \tag{40}$$

Equation (40) arises in models of signal theory and was first studied by Nowosad (1966) in the case that $K(s,i)$ is symmetric and $S = [0,1]$. Since this equation is essentially the same as equation (33) when $\gamma_2/\gamma_1 = -1$ the argument for uniqueness in Theorem 2 directly applies here. In particular, note that because w, L, and T are continuous and bounded above and below by strictly positive numbers and $\frac{W^{1-\sigma}}{\phi}$ is strictly positive, the kernel K is continuous and $K(i,i) > 0$ for all $i \in S$. As a result, we can apply Theorem 2 and Remark 1 of Karlin and Nirenberg (1967) to equation (40), which imply that there exists a unique continuous positive function $u(i)$ satisfying equation (39) for any connected compact subspace of $\mathbb{R}^N$. The continuous and positive function A can then be determined using equation (40). Note that A and u are only identified up to scale, as the constants ϕ and W offset any changes in the normalizations of A and u, respectively.

A.2 Isomorphisms

We study two separate types of isomorphisms of our elementary gravity model with labor mobility to richer gravity trade models. First, we show that our main setup can be shown to be isomorphic to the class of gravity trade models considered by Arkolakis, Costinot, and Rodríguez-Clare (2012) if an equilibrium with labor mobility is considered in that setup. Second, we show that our setup is isomorphic to a new economic geography model as in Krugman (1991), but when an inelastic supply of housing (amenity) is introduced, as in Helpman (1998). Finally, we show an isomorphism to a gravity model where workers have idiosyncratic utility shocks for each location. In all these exercises we consider a surface S with a continuum of locations.

Gravity models Arkolakis, Costinot, and Rodríguez-Clare (2012) consider gravity trade models with exogenous entry and free entry. Models with exogenous entry include Eaton and Kortum (2002), Chaney (2008)-Melitz (2003), and of course, the Armington (1969) setup. The gravity trade relationships and the labor market clearing conditions of these models are very similar. As long as the models are set to have the same bilateral trade costs, population, and also the elasticity of trade is set to the same value, their technology parameters can be adjusted so that they are formally isomorphic. This elasticity of trade parameter is the Frechet curvature parameter in Eaton and Kortum (2002), the Pareto curvature parameter in Chaney (2008)-Melitz (2003), and the CES demand elasticity in Armington (1969). Given this formal isomorphism, introducing labor mobility simply extends the isomorphism to a labor mobility equilibrium, as we have introduced in the main text with exogenous productivities and amenities.

The isomorphism carries on in the case of models with free entry, but allowances have to be made in order for a non-degenerate equilibrium to emerge. Models of free entry analyzed by Arkolakis, Costinot, and Rodríguez-Clare (2012) include Krugman (1980), and the Melitz (2003) model with Pareto distributed productivities considered by Arkolakis, Demidova, Klenow, and Rodríguez-Clare (2008). The assumption of free entry implies that firms need to hire f^e units of local labor to produce a unique differentiated variety in a location, and in the resulting equilibrium the number of entrants is proportional to population, $N_i \propto L_i$. When labor is allowed to move, it is straightforward to show that our setup with production spillovers and $\alpha = 1/(\sigma - 1)$ is isomorphic to the free entry models discussed above. Unfortunately, $\alpha = 1/(\sigma - 1)$ and $\beta = 0 \implies \gamma_1 = 0$, and the only equilibrium is a "black-hole" equilibrium where all production concentrates in one location. For a non-degenerate equilibrium to arise, the production externality needs to be less strong, which corresponds to allowing a negative production externality in the Arkolakis, Costinot, and Rodríguez-Clare (2012) setup with free entry, and respectively setting $\alpha < 1/(\sigma - 1)$ in our model.

Non-tradable sector (e.g. housing) A formal isomorphism can be derived with the Helpman (1998)-Redding (2012) setup. That setup assumes that workers spend a constant share δ of their income on differentiated goods and a share $1-\gamma$ to local non-tradable goods (often referred to as "housing"). A preference structure that gives rise to this is a monotonic transformation of a Cobb-Douglas aggregator with a coefficient one on the differentiated goods and $(1-\delta)/\delta$ on housing. The differentiated sector is as in Krugman (1980) and Krugman (1991) while earnings from land are equally divided by workers residing in that location. In equilibrium with labor mobility, a constant share of income is earned from wages and rents. To formally map this model to our setup we need to set $\alpha = 1/(\sigma - 1)$ and also

$\beta = -(1-\delta)/\delta$. For the equilibrium to be unique we require that

$$\alpha + \beta \geq 0 \iff \frac{1 + (\delta - 1)(\sigma)}{\delta(\sigma - 1)} \geq 0$$

This condition is discussed as a sufficient condition for the existence mobility equilibrium in an N-location world, in Redding (2012). In fact, Theorem 2 implies that this condition is sufficient for a unique labor mobility equilibrium in our setup, while the sufficient conditions for existence are weaker. It is easy to check the rest of the parts of the isomorphism, i.e. that the gravity relationship of trade is exactly the same and the trade balance condition is the same. Notice that similar isomorphisms can be derived if other free entry setups are considered instead of the Krugman (1980) one.

Land as a factor of production Suppose that land $M(i)$ and labor $L(i)$ are combined in a Cobb-Douglass production function to produce output: $Q(i) = L(i)^{\gamma} M(i)^{1-\gamma}$, where $M(i)$ is fixed. Suppose too that workers living in location $i \in S$ each own an equal share of the land, so that their total income is $w(i) + \frac{M(i)}{L(i)} r(i)$. From the first order conditions of the production function, the total income of each worker is $\frac{1}{\gamma} w(i)$. Since production in our model is $Q(i) = A(i) L(i)^{1+\alpha}$, a formal isomorphism exists where productivity $A(i) = M(i)^{1-\gamma}$, $\alpha = \gamma - 1$, and wages are scaled by $\frac{1}{\gamma}$.

Worker Heterogeneity We now build a formal isomorphism to a model where workers have idiosyncratic utility shocks in each location. Notice that this isomorphism holds for any finite number of locations. Suppose that a worker ω receives welfare $U(i, \omega)$ from living in location $i \in S$, where:

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

and $\upsilon(i, \omega)$ is distributed i.i.d. Frechet across people and locations with shape parameter θ, i.e. $Pr[\upsilon \leq u] = e^{-u^{-\theta}}$. If workers choose to live in the location with their highest idiosyncratic utility, then for any two locations $i, s \in S$, the ratio of the population densities can be written as a function of the non-idiosyncratic welfare:

$$\frac{L(i)}{L(s)} = \frac{\left(\frac{w(i)}{P(i)} u(i) \right)^{\theta}}{\left(\frac{w(s)}{P(s)} u(s) \right)^{\theta}} \iff \frac{w(i)}{P(i)} u(i) L(i)^{-\frac{1}{\theta}} = \frac{w(s)}{P(s)} u(s) L(s)^{-\frac{1}{\theta}}.$$

Since $u(i) = \bar{u}(i) L(i)^{\beta}$, the above condition is isomorphic to the utility equalization condition presented in the main text with the alternative $\tilde{\beta} = \beta - \frac{1}{\theta}$. Hence, adding heterogeneous worker preferences simply creates an additional dispersion force.

Capital and fixed factors of production, labor and capital mobility, no trade costs (Roback, 1982) Consider the following model based on Roback (1982). Suppose that there are three factors of production: labor $L(i)$, a fixed factor $F(i)$, and capital $K(i)$ which are combined together in a constant returns to scale Cobb-Douglass production function to produce a homogeneous tradable good $Q(i)$. It is assumed that labor and capital are perfectly mobile and the tradable good is traded costlessly. Finally, suppose that workers receive utility from consuming the tradable good, a local non-tradable good ("housing") $H(i)$ and a local amenity $u(i)$, where the utility function is also assumed to be Cobb-Douglass. Note that this framework nests, among others, Kline and Moretti (2014) and Diamond (2012) (with a single type of worker). Let the price of capital be $\kappa(i)$, the price of the tradable good be $P(i)$, the price of the non-tradable good be $R(i)$, and the wage be $w(i)$.

First we derive the welfare equalization condition. Because workers are perfectly mobile, welfare is equalized across locations. Given the Cobb-Douglass utility function, we can represent the utility equalization condition using the indirect utility function:

$$\ln W = \ln w(i) - (1-\zeta) \ln P(i) - \zeta \ln R(i) + \ln u(i),$$

where ζ is the fraction consumers spend on the non-tradable good. Suppose that rent can be written as:

$$\ln R(i) = \ln C(i) + \gamma \ln L(i),$$

where $\gamma > 0$ is the elasticity of rent to the population and $C(i)$ are (exogenous) local production costs. Since it is assumed that goods are perfectly traded, we have $P(i) = P$. Then we can rewrite utility equalization as:

$$\ln \tilde{W} = \ln w(i) - \zeta\gamma \ln L(i) + \ln u(i) - \zeta \ln C(i), \tag{41}$$

where $\ln \tilde{W} \equiv \ln W + (1-\zeta)$.

We now derive the equilibrium wage condition, which arises from the first order condition of profit maximizing firms. Let the production function be:

$$Q(i) = A(i) L(i)^{\delta} K(i)^{\lambda} F(i)^{1-\delta-\lambda},$$

where $A(i)$ is the local productivity. Taking the first order conditions of capital and substituting it into the first order condition for wages yields:

$$\ln w(i) = \ln\delta + \frac{\lambda}{1-\lambda}\ln\lambda + \frac{1}{1-\lambda}\ln A(i) + \left(\frac{\delta-1+\lambda}{1-\lambda}\right)\ln L(i) - \frac{\lambda}{1-\lambda}\ln\kappa(i) + \frac{1-\delta-\lambda}{1-\lambda}\ln F(i).$$

Since it is assumed capital is perfectly mobile, we have $\kappa(i) = \kappa$ so that we can write:

$$\ln w(i) = C_2 + \frac{1}{1-\lambda}\ln A(i) + \left(\frac{\delta-1+\lambda}{1-\lambda}\right)\ln L(i) + \frac{1-\sigma-\lambda}{1-\lambda}\ln F(i), \tag{42}$$

where $C_2 \equiv \ln\delta + \frac{\lambda}{1-\lambda}\ln\lambda - \frac{\lambda}{1-\lambda}\ln\kappa$.

Let us now compare this to this model. Welfare equalization (equation (5)) along with the amenity spillovers (equation (2)) yields:

$$\ln W = \ln w(i) + \ln\bar{u}(i) + \beta\ln L(i) - \ln P(i).$$

Note from equation (4) that if trade is costless, then $P(i)$ is constant. Hence we have:

$$\ln\tilde{W} = \ln w(i) + \ln\bar{u}(i) + \beta\ln L(i) - \ln P(i). \tag{43}$$

Comparing equation (43) to equation (41), the two are equivalent if $\beta = -\zeta\gamma$ and $\ln\bar{u}(i) = \ln u(i) - \zeta\ln C(i)$.

Recall that combining market clearing, welfare equalization, and imposing the symmetry of trade costs yields equation (12), which can be written as:

$$\ln w(i) = \frac{1}{2\sigma-1}\ln\phi + \frac{\sigma-1}{2\sigma-1}\ln\bar{A}(i) + \frac{(\sigma-1)(\alpha-\beta)-1}{2\sigma-1}\ln L(i) + \frac{1-\sigma}{2\sigma-1}\ln\bar{u}(i) \tag{44}$$

Comparing equation (44) to equation (42), the two are equivalent (to scale) if $\frac{\sigma-1}{2\sigma-1} = \frac{1}{1-\lambda}$, $\frac{(\sigma-1)(\alpha-\beta)-1}{2\sigma-1} = \frac{\delta-1+\lambda}{1-\lambda}$, and $\frac{1-\sigma}{2\sigma-1}\ln\bar{u}(i) = \frac{1-\delta-\lambda}{1-\lambda}\ln F(i)$. Because there are three degrees of freedom $\{\sigma,\alpha,\beta\}$, we can ensure that all three equations $\{\beta = -\zeta\gamma, \frac{\sigma-1}{2\sigma-1} = \frac{1}{1-\lambda}, \frac{(\sigma-1)(\alpha-\beta)-1}{2\sigma-1} = \frac{\delta-1+\lambda}{1-\lambda}\}$ are satisfied. However, in general, $\frac{1-\sigma}{2\sigma-1}\ln\bar{u}(i) = \frac{1-\delta-\lambda}{1-\lambda}\ln F(i)$ and $\ln\bar{u}(i) = \ln u(i) - \zeta\ln C(i)$ will not be both satisfied simultaneously. Hence, an isomorphism exists if either there are arbitrary differences in amenities or arbitrary differences in the supply of the fixed factor of production, but not both.

B Data appendix

This appendix describes the data used in Section 3.

B.1 Population, wage, and demographic data

The population, wage, and demographic data comes from the 2000 U.S. Census. We use county level data retrieved from the National Historic Geographic Information System MPC (2011a), of which there are $3,109$ in the continental United States. To construct the distribution of labor $L : S \to \mathbb{R}_+$, we take the total population of the county in 2000, divide it by the area of the county (which is calculated in GIS), and normalize the population so it has a mean of one. To construct the distribution of wages $w : S \to \mathbb{R}_{++}$, we divide the aggregate wage or salary income in 1999 variable by the total population of the county and normalize the wage so it has a mean of one. We use demographic data from the 2000 U.S. Census to construct non-geographic measures of bilateral trade costs. Between all bilateral county pairs, we calculate the correlation in ethnicity (using the number of people in each county in each of the 14 U.S. Census race categories), the correlation in language (using the number of people aged 18 to 64 in each county who speak English, Spanish, other Indo-European languages, Asian and Pacific Island languages, and other languages primarily at home), and the correlation in migrants (using the number of people in each county who are born in Europe, Asia, Africa, Oceania, Americas, at sea, and everywhere else). Finally, we use GIS to calculate the latitude and longitude of the centroid of every county, which we use to match the county data to the transportation network data.

B.2 Bilateral trade flow data

The bilateral trade flow data comes from the 2007 U.S. Commodity Flow Survey (CFS, 2007). For every pair of the 122 CFS areas and for each mode of travel, the CFS reports the quantity and value of goods shipped by mode of travel. We classify the modes of travel into five categories: road (corresponding to CFS categories truck, for-hire truck, and private truck), water (corresponding to CFS categories deep draft, Great Lakes, shallow draft, and water), rail (corresponding to the CFS category rail), air (corresponding to the CFS category), and other (which includes all other CFS categories, notably those reporting multiple modes of travel). We use the Google Maps API to determine the latitude and longitude of the centroid of each CFS area. This allows us to place each CFS area on the transportation network and allows us to determine the demographic data of each CFS area (which we equate the the demographic data of the nearest county). We also distinguish between CFS areas that are either parts of or entire metropolitan statistical areas (MSAs) (e.g. the IL part of the Chicago-Naperville-Michigan City MSA) versus those that are not (e.g. the remainder of IL), as the former correspond more closely to our theoretical concept of a location.

B.3 Transportation network data

The transportation network data are collected from three sources. Geographic data on the road network is taken from the National Highway Planning Network (NHPN, 2005). According to its meta data, the NHPN consists of over 400,000 miles of the nation's highways comprised of rural arterials, urban principal arterials and all National Highway System routes. In our analysis, we distinguish between interstate highways, national non-Interstate highways, and arterial roads. Geographic data on the railroad network is from the Center for Transportation Analysis Railroad Network dataset (CTA, 2003). It contains every railroad route in the U.S., Canada, and Mexico that has been active since 1993. For each railroad line, it reports a "subjective rating of line importance, and implicitly quality." In our analysis, we distinguish between class-A railroad lines, class-B railroad lines, and all other classes. Geographic data on the water network is from the U.S. Army Corps of Engineers Navigation Data Center (NDC, 1999). It includes all navigable inland waterways, off-shore water routes, and routes via the Great Lakes and the Saint Lawrence Seaway.

To determine the mode-specific normalized transportation cost between any two locations in the U.S., we construct a cost raster for each of the transportation networks. To do so, we first convert each element of each mode network into a raster file covering the continental United States using the 1983 North American GCS projection. Each raster file has resolution 1032×760, so that each pixel corresponds roughly to a 5km by 5km square. To construct the normalized road network cost raster, we assign all pixels with interstate highways a cost of one, all pixels with non-interstate highways a cost of $\frac{70}{55}$, all pixels with arterial roads a cost of $\frac{70}{35}$, and all other pixels a cost of $\frac{70}{20}$, where the costs are chosen to (roughly) reflect differences in relative travel speeds. To construct the normalized railroad cost raster, we assign all pixels with class A railroads a cost of one, all pixels with class B railroads a cost of 1.25, all pixels with other railroads a cost of 1.5, and all other pixels a cost of 3. To construct the normalized water cost raster, we assign all pixels with navigable waterways a cost of one and all other pixels a cost of 10. To construct the normalized air cost raster, we assign all pixels a cost of one (reflecting the fact that the cost of traveling through the air is (roughly) uniformly costly regardless of the location.