NBER WORKING PAPER SERIES

# TRADE AND THE TOPOGRAPHY OF THE SPATIAL ECONOMY

Treb Allen
Costas Arkolakis

Working Paper 19181
http://www.nber.org/papers/w19181

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2013

We thank George Alessandria, Dave Donaldson, Gilles Duranton, Pablo Fajgelbaum, Gene Grossman, Gordon Hanson, Johannes Horner, Sam Kortum, Kiminori Matsuyama, Stephen Redding, Esteban Rossi- Hansberg, Antonios Stampoulis, and Jonathan Vogel for their helpful suggestions and comments. Li Xiangliang provided excellent research assistance. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w19181.ack

Trade and the Topography of the Spatial Economy
Treb Allen and Costas Arkolakis
NBER Working Paper No. 19181
June 2013
JEL No. F10,O18,R12,R13

**ABSTRACT**

We develop a versatile 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 differential 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, amenities and the strength of spillovers in the United States. We find that geographic location accounts for 24% of the observed spatial distribution of income. Finally, we calculate that the construction of the interstate highway system increased welfare by 3.47%, roughly twice its cost.

Treb Allen
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2600
treb.allen@northwestern.edu

Costas Arkolakis
Department of Economics
Yale University, 28 Hillhouse Avenue
P.O. Box 208268
New Haven, CT 06520-8268
and NBER
costas.arkolakis@yale.edu

# 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, 2011). 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 theoretical spatial economic models has restricted their direct applicability to a narrow 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. Using this framework, we perform quantitative empirical analysis using the observed distribution of economic activity and the observed geography of the United States. We estimate that a substantial fraction – 24 percent – of the spatial variation in incomes across the United States can be explained by geographic location alone.

The first part of the paper presents our theoretical framework, which relies on distinct (but mutually compatible) *economic* and *geographic* components. 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. To incorporate the possibility of productivity or congestion externalities, we allow for the total productivity and amenity in a location to endogenously depend on its population ("spillovers"). Given this setup, we characterize conditions for the existence, uniqueness, and

[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).

stability of a spatial economic equilibrium. We also provide a simple iterative technique for calculating the equilibrium.

The geographic component provides a micro-foundation for 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, thereby providing a unique mapping between the topography of instantaneous trade costs and the resulting bilateral trade costs.

By combining the economic and geographic components, we derive a set of differential equations that succinctly relate the endogenous economic outcomes to the underlying geography of the surface and highlight the role that spillovers play in determining the equilibrium spatial distribution of economic activity. We provide several stylized examples of the mechanisms of the model, and derive a closed-form solution to the equilibrium distribution of population across a homogeneous finite line in the absence of spillovers.

The second part of the paper uses the theoretical framework to analyze the real world distribution of economic activity throughout the continental United States. We estimate the underlying geography, bilateral trade costs, productivities and amenities, of the United States. For the bilateral trade costs, we assume that the instantaneous trade costs are a function of the observable geographical features (e.g. rivers, ocean, mountains, etc.) and then estimate the relative cost of traveling over each type of land using observed bilateral state to state trade shares. The procedure is greatly facilitated by the "fast marching method" algorithm borrowed from computational physics, which allows us to efficiently compute the lowest trade cost from all locations to all other locations. Given the trade costs, we then identify the unique topography of productivities and amenities that exactly match the observed spatial distribution of wages and population given the structure of the model. In the final estimation step, we determine the strength of spillovers using the correlation between the change in the calibrated productivities and amenities and the observed change in the population distribution before and after the construction of the Interstate Highway System (IHS).

We perform two exercises using the estimated geography of the United States. First, we decompose the observed spatial distribution of economic activity into its underlying sources: differences in underlying productivity, amenities, and geographic location. We find that 24 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 the Interstate

Highway System. We estimate that the Interstate Highway System increased welfare by 3.47% – roughly twice its cost – and show that the model does a good job predicting the observed redistribution of labor and income across space.

In addition to allowing for a nearly arbitrary geography, our framework departs from the standard model of economic geography based on Krugman (1991), and extensively analyzed by Fujita, Krugman, and Venables (1999), in three 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). Third, rather than taking a stand on the source of production externalities or congestion externalities, we incorporate such spillovers by simply assuming that productivity and amenities may depend in part on the local population.[2] While ad-hoc, this assumption allows us to directly assess how the strength of spillovers affect the spatial distribution of economic activity. Despite these departures from the standard spatial framework, we show that for particular strengths of spillovers, our model becomes isomorphic to a 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).

Like Fujita, Krugman, and Venables (1999), this paper provides a theoretical treatment of the factors determining the spatial distribution of economic activity.[3] In this manner, our work is also related to Matsuyama (1999), who characterizes specialization into industries and economic activity and how this is affected by the geographical location of countries under a variety of different scenarios.[4] However, the primary goal of the paper is to provide an empirically-implementable framework to study of the role of economic geography (in the spirit suggested by Duranton (2008)). 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

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

[3]Fabinger (2012) also derives equations governing the relationship between equilibrium economic outcomes and the geography of a space, although his work is concerned with the strength of income spillovers in a world arranged along a circle or a sphere.

[4]Our model can be easily extended to allow for labor mobility within but not between countries. In this case, we can derive equilibrium expressions for the relationship between internal trade frictions and the welfare gains from international trade. In this sense, our work is also related to the work of Cosar and Fajgelbaum (2012), Ramondo, Rodríguez-Clare, and Saborio-Rodriguez (2012) and Redding (2012).

has been little empirical application of the extensive body of theoretical research on economic geography (with the notable exceptions of Redding and Sturm (2008) and Ahlfeldt, Redding, Sturm, and Wolf (2012)).

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 such transportation networks can be incorporated in our framework, we can also incorporate trade costs that do not have obvious network representations, for example geographical features like mountains and oceans.

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 four 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. Next, we combine the economic and geographic components to generate a set of differential equations governing the equilibrium distribution of economic activity on a surface with geographic trade costs. Finally, we provide several simple examples of the model highlighting the mechanisms at play.

## 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 is a continuum of locations $i \in S$, where $S$ is a compact subset of $\mathbb{R}^N$. Each location $i \in S$ produces a unique differentiated variety of a good. Trade is costly: trade costs are of the iceberg form and are described by the function $T : S \times S \to [1, \infty)$, 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 $j$.

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:

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

where $q(i)$ is the quantity consumed of the variety produced in location $i$, $\sigma \in (1,\infty)$ is the elasticity of substitution between goods $\omega$, and $u(i)$ is the local amenity.[5] We denote by $P(i)$ the consumer price index in location $i$.

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 spillovers or congestion externalities, both productivity and amenities may depend on the density of workers. In particular, we assume that 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 $\alpha$ determines the extent of the productivity spillover. Similarly, we assume that the 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 and $\beta$ determines the extent of the congestion externality. 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 monopo-

[5]While the model attains a non-trivial solution even for $\sigma \in (0,1)$ we focus on this parameter space to obtain equilibria where more productive locations attract more workers.

listically competitive framework with free entry, where the number of goods produced in a location is proportional to its population. Similarly, if $\alpha = \frac{1}{\sigma-1}$ and $\beta = -\frac{1-\gamma}{\gamma}$, our model is isomorphic to the Helpman (1998)-Redding (2012) framework with $1-\gamma$ being the budget share spent on the immobile factor.

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 $L$.

### 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 $X : S \times S \to \mathbb{R}_+$ describes equilibrium value of trade flows, i.e. $X(i,j)$ expresses the value of bilateral trade flows from location $i$ to location $j$. 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

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

is the CES price index. Define $\lambda(i,j) \equiv \frac{X(i,j)}{\int_S X(s,j)ds}$ as the relative market share of location $i$ in location $j$. Combining equations (3) and (4) yields for all $i \in S$:

$$A(i)^{\sigma-1} w(i)^{1-\sigma} = \lambda(i,i) \int_S T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma}\,ds, \tag{5}$$

where $\lambda(i,i)$ is the ratio of local goods consumed in location $i$ to the average amount of goods consumed in location $i$ from any location (so that higher values of $\lambda(i,i)$ indicate that a location is less open to trade).

### 2.1.3 Equilibrium

Trade is said to be *balanced* if for all $i \in S$:

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

When trade is balanced, the welfare of living in a location is:

$$W(i)=\frac{w(i)}{P(i)}u(i)=\frac{\left(\int_S T(i,s)^{1-\sigma}P(s)^{\sigma-1}w(s)\,L(s)\,ds\right)^{\frac{1}{\sigma}}}{P(i)}\bar{A}(i)^{\frac{\sigma-1}{\sigma}}\,\bar{u}(i)\,L(i)^{-\frac{\gamma_1}{\sigma}}, \tag{7}$$

where $\gamma_1\equiv 1-(\sigma-1)\,\alpha-\beta\sigma$. The parameter $\gamma_1$ is the partial elasticity of the welfare in a location with respect to the population in that location. Welfare is said to be *equalized* if for all $i\in S$ there exists a $W>0$ such that $W(i)\leq W$, with the equality strict 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.

Given a regular geography with parameters $\sigma$, $\alpha$, and $\beta$, we define a *spatial equilibrium* as a distribution of economic activity such that (i) trade is balanced; (ii) welfare is equalized; and (iii) the aggregate labor market clears:

$$\int_S L(s)\,ds=\bar{L}. \tag{8}$$

In what follows, we pay particular attention to two types of spatial equilibria. 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]

We now discuss sufficient conditions for the existence and uniqueness of the spatial equilibria. Using equations (3), (4), and (5) to substitute out for trade flows, $X(i,s)$, and the price index, $P(j)$, we can write the balanced trade condition (6) for all $i\in S$ as:

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

Combining equation (4) and (5) with utility equalization implies for all $i\in S$ and $s\in S$ such

[6] This concept of stability is an adaptation of the one first introduced by Krugman (1991) to a continuum of locations. Note that with a continuum of locations, changing the population in a single location does not affect the price index in that location.

that $L(i) > 0$ and $L(s) > 0$:

$$\lambda(i,i) A(i)^{1-\sigma} u(i)^{1-\sigma} = \lambda(s,s) A(s)^{1-\sigma} u(s)^{1-\sigma} = W^{1-\sigma} \tag{10}$$

Substituting equations (1), (2), and (10) into the balanced trade equation (9) yields for all $i \in S$ such that $L(i) > 0$:

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

Similarly, substituting equations (1), (2), and (4) into utility equalization yields for all $i \in S$ such that $L(i) > 0$:

$$w(i)^{1-\sigma} L(i)^{\beta(1-\sigma)} = W^{1-\sigma} \int_S T(s,i)^{1-\sigma} \bar{A}(s)^{\sigma-1} \bar{u}(i)^{\sigma-1} w(s)^{1-\sigma} L(s)^{\alpha(\sigma-1)} ds. \tag{12}$$

These two sets of equations constitutes a system that can be used to solve for $w(i)$ and $L(i)$.

When there are no productivity or amenity spillovers (i.e. $\alpha = \beta = 0$ so that $A(i) = \bar{A}(i)$ and $u(i) = \bar{u}(i)$), equations (11) and (12) are eigenfunctions for $L(i) w(i)^{\sigma}$ and $w(i)^{1-\sigma}$, respectively. 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. ■

Consider now the case with productivity or amenity spillovers, and suppose bilateral trade costs are symmetric (i.e. $T(i,s) = T(s,i)$ for all $i, s \in S$). Then substituting (10) into equations (11) and (12) yields a relationship between $L(i)$ and $w(i)$ that is consistent with both these equilibrium equations:

$$L(i) = \left(\phi \bar{A}(i)^{\sigma-1} w(i)^{1-2\sigma} \bar{u}(i)^{1-\sigma}\right)^{\frac{1}{1-(\sigma-1)(\alpha-\beta)}}, \tag{13}$$

where $\phi$ is a scalar. If equation (13) holds, then any functions $w(i)$ and $L(i)$ satisfying equation (11) will also satisfy (12) (and vice versa).[7]

[7] We prove in the subsequent theorem that for any regular equilibrium, equation (13) is the unique relationship between $L(i)$ and $w(i)$ such that equations (11) and (12) hold.

Combining equations (13) and (12) yields (after some algebra):

$$L(i)^{\tilde{\sigma}\gamma_1} = \bar{u}(i)^{(1-\tilde{\sigma})(\sigma-1)} \bar{A}(i)^{\tilde{\sigma}(\sigma-1)} W^{1-\sigma} \int_S T(s,i)^{1-\sigma} \bar{A}(s)^{(1-\tilde{\sigma})(\sigma-1)} \bar{u}(s)^{\tilde{\sigma}(\sigma-1)} \left(L(s)^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} ds, \tag{14}$$

where $\gamma_2 \equiv 1 + \alpha\sigma + (\sigma-1)\beta$ and $\tilde{\sigma} \equiv \frac{\sigma-1}{2\sigma-1}$.

Equation (14) 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 (14) has a solution for $L(i)$ and $W^{1-\sigma}$ then equilibrium wages can be determined from equation (13) using the aggregate labor clearing condition to determine the scalar $\phi$. 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 endogenous 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 part (iii) is a special case of part (ii).[8] It is straightforward to show that if $\gamma_1 = 0$ there is (generically) no regular spatial equilibrium satisfying equations (13) and (14). Finally, the following proposition characterizes when a spatial equilibria is point-wise locally stable.

**Proposition 1** *Consider a regular geography with endogenous 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.1.3. ■

Figure 1 depicts the ranges of $\alpha$ and $\beta$ and the different cases of equilibrium existence and stability with $\sigma = 4$. The graph is divided in four regions with conditions on $\alpha$ and $\beta$ that

[8] Theorem 1 and Theorem 2(iii) generalize for the case of discrete number of locations, as shown in Appendix A.1.

guarantee 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 it 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 to induce complete concentration in a single location, i.e. a black-hole. While regular equilibria exist (if $\gamma_1 < 0$), they are not point-wise locally stable.

## 2.2 Geographic component

In this subsection, we present the geographic component of our spatial framework. The geographic component provides 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 that the world is a continuum of locations $i \in S$, where $S$ is a manifold. In what follows, we focus on the cases where $S$ is a finite line, a finite circle, and a finite plane.[9]

Let $\tau : S \to \mathbb{R}_+$ be a continuous function where $\tau(i)$ gives the instantaneous (iceberg) trade cost incurred by crossing point $i \in S$. Let $t(i,j)$ be the solution to the following least cost path minimization problem:

$$t(i,j) = \min_{\gamma \in \Gamma(i,j)} \int_0^1 \tau(\gamma(t)) \left\| \frac{d\gamma(t)}{dt} \right\| dt, \tag{15}$$

where $\gamma : [0,1] \to S$ is a path and $\Gamma(i,j) \equiv \{\gamma \in C^1 | \gamma(0) = i, \gamma(1) = j\}$ 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 $i,j \in S$, $T(i,j) = e^{t(i,j)}$, we say that the bilateral trade costs are *geographic*.[10] Note that when bilateral trade costs are geographic, there exists a unique mapping from the ($n$-dimensional) instantaneous trade cost function $\tau$ to the ($n^2$-dimensional) bilateral trade costs $T$.

Equation (15) is a well studied problem that arises in a number of fields, including

[9] In fact, practically all the theoretical results that follow, apply to any finite-dimensional manifold.

[10] Note that $e^{\int_a^b f(x)dx} = \prod_a^b (1 + f(x)\, dx)$, where $\prod_a^b$ denotes a type II product integral.

fluid mechanics, image processing, and even the study of the formulation of snowflakes. Its solution is characterized by the following Eikonal (or Hamilton-Jacobi in the viscosity sense) partial differential equation (see Mantegazza and Mennucci (2003)). For any origin $i \in S$ and destination $j \in S$:

$$||\nabla t\left(i,j\right)|| = \tau\left(j\right), \tag{16}$$

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

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 $\{j|t\left(i,j\right) = C\}$ for all $C$. Equation (16) 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 (log) 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. As a result, beginning with an arbitrarily small $C$ around any initial point $i \in S$, one can trace the expansion of the contours to determine $t\left(i,j\right)$ for all $j \in S$. Figure 2 illustrates the propagation process.

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 $i,j \in S$, $T\left(i,j\right) = T\left(j,i\right)$. 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 $\lim_{s\to i} T\left(s,j\right) = T\left(i,j\right)$. 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)).

## 2.3 Combining the economic and geographic components

In this subsection, we combine the economic and geographic components by considering a surface $S$ with a regular geography and geographic trade costs. We show that there exists a simple set of differential equations relating the equilibrium spatial distribution of economic activity on $S$ to the geography of $S$.

Combining utility equalization (i.e. $w\left(i\right) = WP\left(i\right)/u\left(i\right)$) with equation (13), taking the gradient with respect to $i$, and substituting in the Eikonal equation (16) yields the following relationships between the topography of the labor supply and wages and the underlying geography of productivity, amenities, and trade costs:

$$\gamma_1 \nabla \ln L(i) = (\sigma - 1) \nabla \ln \bar{A}(i) + \sigma \nabla \ln \bar{u}(i) - (2\sigma - 1) \kappa(i) \tau(i), \tag{17}$$

$$\gamma_1 \nabla \ln w(i) = -\beta(\sigma - 1) \nabla \ln \bar{A}(i) - (1 - \alpha(\sigma - 1)) \nabla \ln \bar{u}(i) + (1 - (\sigma - 1)(\alpha - \beta)) \kappa(i) \tau(i), \tag{18}$$

where $\kappa(i) \equiv \frac{\int_S \nabla t(s,i) T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma} ds}{\int_S ||\nabla t(s,i)|| T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma} ds}$. While succinct, equations (17) and (18) capture a number of important characteristics of the equilibrium distribution of economic activity. In what follows, we focus on the case $\gamma_1 > 0$ so that all spatial equilibria are regular and locally point-wise stable. First, all else equal, the equilibrium labor supply will increase as the underlying productivity and amenity of a location increase. In contrast, the equilibrium wages will decrease as the underlying productivity increases and may increase or decrease as the amenity of a location increases, depending on the sign of $1 - \alpha(\sigma - 1)$.

Second, all else equal, an increase in the remoteness of a location will cause the equilibrium labor supply to fall. To see this, note that $\nabla t(s,i)$ captures how a move from $i$ changes the distance from $s$ and $T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma}$ is proportional to the relative importance of goods from $s$ in $i$'s consumption bundle. As a result, $\kappa(i)$ captures how changes in the geographic location affect the average distance from $i$ to all its trading partners, weighted by the importance of each trading partner.[11] In other words, $\kappa(i)$ captures the gradient of the "remoteness" of a location. If moving in a particular geographic location increases the remoteness, all else equal, the equilibrium labor supply will decrease in that direction, as the price of the consumption bundle increases due to the rising trade costs. In contrast, the effect of the remoteness of a location on the equilibrium wages depends on the sign of $1 - (\sigma - 1)(\alpha - \beta)$.

Third, the instantaneous trade cost bounds how much the equilibrium labor supply and wages can change over space. In particular, note that from the triangle inequality, the length of $\kappa(i)$ is no greater than one:

$$0 \leq ||\kappa(i)|| \leq \frac{\int_S ||\nabla t(s,i) T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma}|| ds}{\int_S ||\nabla t(s,i)|| T(s,i)^{1-\sigma} A(s)^{\sigma-1} w(s)^{1-\sigma} ds} = 1,$$

so that:

$$0 \leq ||(\sigma - 1) \nabla \ln \bar{A}(i) + \sigma \nabla \ln \bar{u}(i) - \gamma_1 \times \nabla \ln L(i)|| \leq (2\sigma - 1) \tau(i)$$

[11] Indeed, it is straightforward to show that $\kappa(i) \tau(i) = \nabla \ln P(i)$.

and

$$0 \leq ||\gamma_1 \times \nabla \ln w(i) + \beta(\sigma-1)\nabla \ln \bar{A}(i) + (\alpha(1-\sigma)+1)\nabla \ln \bar{u}(i)|| \leq |1-(\sigma-1)(\alpha-\beta)|\tau(i).$$

Intuitively, the gradient of the log of the price index is bounded by the instantaneous trade costs, as at worst (best), moving in a particular direction moves you away from (toward) all other locations, increasing (reducing) all bilateral trade costs by $\tau(i)$.

Fourth, the productivity and amenity spillovers (i.e. $\alpha$ and $\beta$) only affect the topography of the labor supply directly by changing its overall elasticity to geographical factors.[12] To see this, note that $\alpha$ and $\beta$ only enter the constant $\gamma_1$ in equation (17), which simply governs the overall responsiveness of $\nabla \ln L(i)$ to $\nabla \ln \bar{A}(i)$, $\nabla \ln \bar{u}(i)$, and $\kappa(i)\tau(i)$. Thus, increasing the productivity spillovers (i.e. increasing $\alpha$) increases the elasticity of the labor supply to changes in geographical factors, while increasing the congestion externalities (i.e. decreasing $\beta$) decreases the elasticity of the labor supply to changes in geographical factors.

## 2.4 Examples: the line and the circle

To obtain an insight into the forces shaping the spatial equilibrium, we present two simple one-dimensional examples of surfaces: 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 the instantaneous trade costs be constant, i.e. $\tau(i) = \tau$ for all $i \in S$. In this case, $T(i,s) = e^{\tau|i-s|}$. Suppose that $\alpha = \beta = 0$ and $\bar{A}(i) = \bar{u}(i) = 1$, i.e. there are no spillovers and all locations have homogeneous productivities and amenities. Then the differential equation (17) becomes:

$$\frac{\partial \ln L(i)}{\partial i} = -(2\sigma-1)\tau\kappa(i), \tag{19}$$

where $\kappa(i) = \frac{\int_{-\pi}^{i} e^{(1-\sigma)\tau|i-s|}A(s)^{\sigma-1}w(s)^{1-\sigma}ds - \int_{i}^{\pi} e^{(1-\sigma)\tau|i-s|}A(s)^{\sigma-1}w(s)^{1-\sigma}ds}{\int_{-\pi}^{\pi} e^{(1-\sigma)\tau|i-s|}A(s)^{\sigma-1}w(s)^{1-\sigma}ds}$. Because $\kappa(-\pi) = -1$

[12] There is also an indirect effect of $\alpha$ and $\beta$ on $L(i)$ through $\kappa(i)$.

and $\kappa(\pi) = 1$, the following boundary conditions follow immediately from equation (19):

$$\frac{\partial \ln L(-\pi)}{\partial i} = (2\sigma - 1)\tau$$
$$\frac{\partial \ln L(\pi)}{\partial i} = -(2\sigma - 1)\tau.$$

Furthermore, by imposing symmetry, equation (19) implies that the $\frac{\partial \ln L(i)}{\partial i}$ is positive for $i < 0$ and negative for $i > 0$, i.e. the equilibrium distribution is a concentration in the center of the line, with the degree of concentration increasing in both the trade cost $\tau$ and the elasticity of substitution $\sigma$.

It turns out that there exists a closed form solution to equation (19). To see this, we differentiate equation (14) twice to show that the equilibrium satisfies the following second order differential equation:

$$\frac{\frac{\partial^2}{\partial i^2} L(i)^{\tilde{\sigma}}}{L(i)^{\tilde{\sigma}}} = k_1, \tag{20}$$

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 is characterized by the cosine function (see example 8.8.16 in Polyanin and Manzhirov (2008)):

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

where $k_2$ can be determined using the aggregate labor clearing condition.[13]

Figure 3 depicts the equilibrium labor allocation in this simple case for different values of the instantaneous trade cost. 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 productivities and amenities, wages are lower in the middle of the line to compensate for the lower price index.

Exogenous differences in productivities, amenities and spillovers also play a key role in determining the equilibrium allocation of labor and wages. We use numerical methods to

[13] In this simplest form the differential equation resulting from our system is the one also describing a system of free harmonic oscillations, which, when displaced from its equilibrium position experiences a restoring force proportional to the displacement. Mechanical examples include the pendulum, springs, and acoustical systems. 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).

compute these more general cases. Assume, for example, that there are no spillovers, but $\bar{A}(i) = e^{Ai}$. Then the differential equation (17) becomes:

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

so that the equilibrium distribution of population is shifted rightward when $A > 0$. Figure 4 depicts this reallocation of labor toward locations with higher productivities.

A different result is obtained if we increase the parameter $\alpha$ that regulates productivity externalities, but leave 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 5 depicts the population for higher values of $\alpha$, and the resulting increase in the concentration. Notice that further increases in $\alpha$, 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.

**The circle**

The example of the circle illustrates the possibility of multiplicity of spatial equilibria. All the results that we have in this case are numerical. Figure 6 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.[14]

# 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

[14]If we further increase $\alpha + \beta$ to the point that the sign of $\gamma_1$ 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.

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.

## 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 three steps. In the first step, we estimate trade costs using observable geographical features in order to best match the observed bilateral trade shares between states. In the second step, we find the unique composite productivities $A$ and amenities $u$ that generate the observed distribution of wages and population given the trade costs. Finally, we estimate the strength of spillovers $\alpha$ and $\beta$ using changes in the composite productivities and amenities over time and use these spillover estimates to back out the underlying exogenous productivities $\bar{A}$ and amenities $\bar{u}$.

### 3.1.1 Estimating trade costs

We first estimate the bilateral trade cost function $T$. To do so, we assume that $T$ is a composite function of geographic trade costs $T_g$ and non-geographic trade costs $T_{ng}$: $T = T_g + T_{ng}$. Let $G_g : S \to \mathbb{R}^d$ be an observable vector of $d$ geographical characteristics (e.g. elevation, whether $i$ is on land or water, etc.). Similarly, let $G_{ng} : S \times S \to \mathbb{R}^c$ be an observable vector of $c$ bilateral non-geographic characteristics (e.g. if $i$ and $s$ are in the same political area).[15] We assume that both the instantaneous trade costs underlying the geographic trade costs and the non-geographic trade costs can be written as a linear function of observable geographic characteristics:

$$\tau(i) = \beta_g' G_g(i)$$
$$T_{ng} = \beta_{ng}' G_{ng}.$$

[15] Technically, both $X_g$ and $X_{ng}$ are required to be continuous functions on their respective domains for the theoretical results to hold. Discontinuous observables such as political boundaries can be included by assuming borders take up space, thereby "smoothing" the discontinuity. This does not affect the empirical analysis, as the space $S$ is ultimately discretized for computational purposes.

The goal of the following procedure is to estimate $\beta_g$ and $\beta_{ng}$. To do so, we choose the $\beta_g$ and $\beta_{ng}$ to most closely match the state to state bilateral trade flows predicted by the model to the trade flows observed in the 2007 Commodity Flow Survey (CFS) using a nonlinear least squares approach. In the CFS, we observe a measure of the total value of bilateral trade flows from state $o$ to state $d$, $\tilde{X}_{od}$.[16] We assume that this data is measured with error so that the true value of bilateral trade flows $X_{od}$ can be written as:

$$X_{od} = \tilde{X}_{od}\varepsilon_{od}, \tag{21}$$

where $\varepsilon_{od}$ is i.i.d. with $E\left[\ln \varepsilon_{od}\right] = 0$.

The major difficulty of the estimation procedure is that the CFS reports state to state bilateral trade flows while our model predicts bilateral trade flows between any two locations on a surface. Formally, let $S_i$ be the set of all locations in geographic area $i$ (e.g. state $i$). Bilateral trade flows from state $o$ to state $d$ are hence:

$$X_{od} = \int_{i \in S_o} \int_{j \in S_d} X\left(i, j\right) di dj.$$

The usual estimation method would involve guessing $\beta_g$ and $\beta_{ng}$, calculating $X\left(i,j\right)$ for all locations in the world, aggregating to construct state to state trade flows $X_{od}$ and iterating to find the $\beta_g$ and $\beta_{ng}$ that minimize the distance between the $X_{od}$ and the trade flows observed in the data. Unfortunately, such a procedure is too computationally intensive to be feasible. Instead, we use the following procedure based on Anderson and Van Wincoop (2003).

It can be shown from the gravity structure of the model (equation 3) and balanced trade (equation 6) that as long as trade costs are symmetric, trade flows from any $i \in S$ to any $j \in S$ can be written as follows:

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

where $y^W \equiv \int_S w\left(s\right) L\left(s\right) ds$ is the aggregate world income. If the price index is roughly

[16] We use the value of all commodity flows between states to construct trade shares. One potential concern with this measure is that it includes shipments of intermediate goods as well as final goods. While the public CFS data only differentiates commodities at an aggregate level, we have constructed alternative measures of trade shares only using values of commodities that are likely to be final goods (e.g. "cereal grains"). The correlation of the trade shares calculated using all commodities and the trade shares calculated using only this subset of commodities is 0.98 (0.89 in logs), suggesting that the presence of intermediate goods is not substantially affecting the results.

the same for all locations within a state, i.e. $P(s) \approx P_i$ for all $s \in S_i$, then the normalized bilateral trade share can be written as:

$$\frac{X_{od}X_{do}}{X_{oo}X_{dd}} \approx \frac{\left(\int_{i\in S_o}\int_{j\in S_d} K(i,j)\,didj\right)^2}{\left(\int_{i\in S_o}\int_{j\in S_o} K(i,j)\,didj\right)\left(\int_{i\in S_d}\int_{j\in S_d} K(i,j)(j)\,didj\right)}, \tag{22}$$

where $K(i,j) \equiv T(i,j)^{1-\sigma} w(i) L(i) w(j) L(j)$. Notice that this expression says that the normalized *state-level* bilateral trade flows is approximately a weighted average of the normalized *location-level* bilateral trade flows, where the weights are determined by the product of the origin and destination location incomes. Because location incomes are observed in the data, calculating the (approximate) model predicted $\frac{X_{od}X_{do}}{X_{oo}X_{dd}}$ only requires the calculation of the bilateral trade cost function $T(i,j)$ rather than solving the entire model equilibrium.[17]

The remainder of the estimation procedure is as follows. We first guess the value of $\beta_g$ and $\beta_{ng}$. Given $\beta_g$, we use the instantaneous trade costs $\tau(i) = \beta_g' G_g(i)$ to calculate the implied geographic component of the bilateral trade cost function $\tilde{T}_g(\beta_g) : S \times S \to \mathbb{R}_{++}$. We employ the Fast Marching Method (FMM) algorithm, which calculates the bilateral trade cost by propagating the cost contour outwards from a particular origin using equation (16). FMM has proven enormously successful in quickly calculating accurate and precise approximations of bilateral cost functions on arbitrary manifolds (see Sethian (1996); Kimmel, Kiryati, and Bruckstein (1996); Kimmel and Sethian (1998)).[18] We implement FMM routines in Matlab developed by Peyre (2009) and Kroon (2009) to calculate the bilateral trade cost function $\tilde{T}_g(\beta_g)$ for all $i, j \in S_g$ and $j$ pixels.[19] The computation is quick: for a $100 \times 100$ pixel surface (which requires determining the distance between 100 million bilateral pairs), $\tilde{T}_g(\beta_g)$ can be calculated in approximately one minute on a standard personal computer. Total bilateral

[17]In practice, the correlation between the approximate $\frac{X_{od}X_{do}}{X_{oo}X_{dd}}$ and the actual $\frac{X_{od}X_{do}}{X_{oo}X_{dd}}$ predicted by the model is greater than 0.99, suggesting that the assumption of an approximately equal price index within state is not substantially affecting the results.

[18]Other trade papers (e.g. Donaldson (2012)) have calculated the distance function by approximating the distance with a discrete graph and using Dijkstra's algorithm. The Dijkstra algorithm, however, is inappropriate for approximating distances along a continuous surface as its estimated distances will not converge to the true distance as the graph approximation of the surface becomes more refined (see Mitchell (1988)). Fortunately, the FMM provides consistent estimates of the true distance using discrete approximations of the space while retaining the same operational complexity of Dijkstra's algorithm (see Mémoli and Sapiro (2001)). Indeed, the FMM algorithm is equivalent to an extension of the Dijkstra algorithm where the distance is an appropriately weighted average of several paths traveling through nodes in a gridded network (see Tsitsiklis (1995)).

[19]The routine developed by Kroon (2009) is based on the work of Hassouna and Farag (2007) and has been shown to be more accurate; however, it is not implementable on non-planar surfaces.

trade costs implied by $\beta_g$ and $\beta_{ng}$ are then simply $\tilde{T}(\beta_g, \beta_{ng}) = \tilde{T}_g(\beta_g) + \beta'_{ng} G_{ng}$.

Given the bilateral trade costs $\tilde{T}(\beta_g, \beta_{ng})$, we then calculate the implied trade shares using equation (22). Let $f_{od}(\beta)$ denote the predicted log normalized trade share given parameter $\beta \equiv [\beta_g, \beta_{ng}]$. Let $\beta^0$ be the true value so that the predicted trade shares given $\beta^0$ is equal to the actual trade shares, i.e. $f_{od}(\beta^0) = \ln\left(\frac{X_{od}X_{do}}{X_{oo}X_{dd}}\right)$. Note from equation (21) that the expected log difference between the normalized trade share predicted by the model given $\beta^0$ and the normalized trade share observed in the CFS is zero:

$$E\left[f_{od}\left(\beta^0\right) - \ln\left(\frac{\tilde{X}_{od}\tilde{X}_{do}}{\tilde{X}_{oo}\tilde{X}_{dd}}\right)\right] = E\left[\ln\left(\frac{\varepsilon_{oo}\varepsilon_{dd}}{\varepsilon_{od}\varepsilon_{do}}\right)\right] = 0,$$

which suggests the following non-linear least squares estimator:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{N^2} \sum_{o=1}^{N} \sum_{d=1}^{N} \left(f(\beta) - \ln\left(\frac{\tilde{X}_{od}\tilde{X}_{do}}{\tilde{X}_{oo}\tilde{X}_{dd}}\right)\right)^2. \tag{23}$$

To construct $G_g$, we assign each location to one of six mutually exclusive and exhaustive types: ocean, interstate highway, river (or lake), old highways (constructed by 1931), mountain (measured as elevation greater than 2000 meters), and other land, where the order indicates the precedent of each type (e.g. a location with both a river and an interstate highway is classified as an interstate highway). As a result, $\beta_g$ indicates the instantaneous trade cost of traveling over each type of location. We assume $G_{ng}$ is simply an indicator equal to one if the origin location and destination location are in different states, so that $\beta_{ng}$ indicates the ad valorem cost of trading across states.

Table 1 presents the resulting estimates from equation 23. The estimates yield a reasonable ranking of the relative costs of traveling over different types of locations: traveling via interstate highways is least costly, followed by oceans and other land, while traveling over mountains are most costly. The magnitude of the estimates suggest that trade costs are substantial: the ad valorem equivalent trade cost of shipping a good from one coast to the other ranges from 70% (if traveling only via interstate highways) to 320% (if traveling only via mountains). The non-geographic trade costs appear substantial as well; we estimate that there is a 48% ad valorem equivalent trade cost of trading with locations outside one's own state. Overall, the average ad valorem equivalent trade cost between any two locations is 74.3%, which is roughly in line with the estimates of Anderson and Van Wincoop (2004).

Figure 7 shows how well the model is able to match the normalized bilateral trade shares observed in the CFS. Overall, there is a high (0.61) correlation between the predicted and observed normalized bilateral trade shares. The model, however, is not able to fully capture the observed variance in the trade shares, as it tends to under-predict trade flows between states with the greatest amount of trade.

### 3.1.2 Identifying productivities and amenities

Suppose we observe trade costs and the equilibrium distribution of economic activity. Can we identify the underlying topography of productivities and amenities? Note that combining equation (4) with utility equalization implies:

$$u(i)^{1-\sigma} = W^{1-\sigma} \int_S T(s,i)^{1-\sigma} w(i)^{\sigma-1} w(s)^{1-\sigma} A(s)^{\sigma-1} ds. \tag{24}$$

In addition, equation (13) can be rearranged as follows:

$$A(s)^{\sigma-1} = \frac{1}{\phi} L(s) w(s)^{2\sigma-1} u(s)^{\sigma-1}. \tag{25}$$

The following theorem guarantees that for any observed distribution of economic activity, there exists a unique topography of 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.1.4. ■

In practice, amenities $u(i)$ can be determined (up to scale) by iterating equation (24) from an initial guess of the amenities, using equation (25) to substitute for $A(s)^{\sigma-1}$. Figure 8 depicts the observed spatial distribution of people (top panel) and wages (bottom) panel in the United States in the year 2000, which is available at the county level from MPC (2011).[20] Figure 9 depicts the unique spatial distribution of productivities $A(i)$ (top panel) and $u(i)$ (bottom panel) that are consistent with the observed distribution of people and

[20] Population density $L(i)$ is measured as the average number of people per square kilometer in the county. Wages $w(i)$ are measured as the total payroll per capita in each county and are trimmed at the 1%/99% level. Both wages and population density are normalized to have a mean of one.

income along with the trade costs estimated in the previous section. Conditional on the price index, locations with observed higher wages are estimated to have lower composite amenities (from utility equalization), whereas conditional on the price index and wages, locations with greater populations are estimated to have higher composite productivities (from balanced trade). As a result, cities like New York, Chicago, and Los Angeles are estimated to have higher composite productivities and lower composite amenities than nearby areas which are less densely populated and have lower wages.

It should be noted that Theorem 3 implies that only the composite productivity $A(i)$ and amenity $u(i)$ can be identified from the observed topography of wages and population. If the parameters $\alpha$ and $\beta$ governing the strength of spillovers are also known, then the underlying topography of $S$ (i.e. the functions $\bar{A}$ and $\bar{u}$) can be determined simply using equations (1) and (2) (since labor is observed). However, $\alpha$ and $\beta$ cannot be identified from the observed distribution of wages and population: for any $\alpha$ and $\beta$, 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.[21] The next subsection discusses how the strength of spillovers can be identified using time series variation resulting from a change in trade costs.

### 3.1.3 Estimating the strength of spillovers

Consider a surface $S$ with functions $\bar{A}$ and $\bar{u}$ governing the topography of exogenous productivity and amenities. Suppose that the equilibrium distribution of economic activity was observed under two alternative trade cost regimes $T_0$ and $T_1$, i.e. the set of functions $\{w_0, L_0, T_0\}$ and $\{w_1, L_1, T_1\}$ are observed. From Theorem 3, there exists unique functions $A_0$ and $u_0$ corresponding to $\{w_0, L_0, T_0\}$ and unique functions $A_1$ and $u_1$ corresponding to $\{w_1, L_1, T_1\}$. As long as the exogenous productivities $\bar{A}$ and amenities $\bar{u}$ remain unchanged, it is possible to identify the strength of spillovers since $\alpha = \frac{\ln \hat{A}}{\ln \hat{L}}$ and $\beta = \frac{\ln \hat{u}}{\ln \hat{L}}$ from equations (1) and (2), where the hat denotes the ratio between regimes, i.e. $\hat{A} \equiv \frac{A_1}{A_0}$. Intuitively, a change in the trade costs will result to a reallocation of economic activity across $S$. Because the composite amenities and productivities can be identified in the cross section and the exogenous functions $\bar{A}$ and $\bar{u}$ remain unaffected by the trade costs, any observed change in $A$ and $u$ must have arisen from changes due to spillovers.

To illustrate the process, we consider a change in trade costs arising from the construction

[21] 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.

of the U.S. interstate highway system (IHS). Construction of the IHS began in 1956 and was proclaimed complete in 1992. It now includes approximately 76,000km of roads. In Section 3.1.1, we estimated the effect of both the IHS and the old highway system on instantaneous trade costs. To determine the instantaneous trade costs prior to the construction of the IHS, we remove the interstate highway system from the topography of instantaneous trade costs, reassigning locations with interstate highways to the appropriate alternative category (using the ordering from above).

Figure 10 compares the topography of instantaneous trade costs prior to the construction of the IHS (top panel) to the instantaneous trade costs after the construction of the IHS (middle panel) using the coefficients estimated in Section 3.1.1; the decline in the instantaneous trade costs (bottom panel) depicts the complete IHS network. We then calculate the bilateral trade cost functions $T_0$ and $T_1$ given the topography of instantaneous trade costs before and after the construction of the IHS, respectively, using the FMM method described in Section (3.1.1). As an example, Figure 11 shows the bilateral trade cost from New York City to all other locations in the United States prior to the construction of the IHS (top panel), after the construction of the IHS (middle panel), and the resulting change in bilateral trade costs (bottom panel). The construction of the IHS differentially reduced the trade costs to destinations far away, as the least cost route to these destinations is along greater stretches of interstate highways.

To determine the equilibrium distribution of economic activity prior to the construction of the IHS, we use the 1930 county level U.S. census data available from MPC (2011). The population density $L_0$ is measured as the county-level average population density and the wage $w_0$ is measured as the total value of output (agriculture plus manufacturing) per capita in the county.[22] Figure 12 presents the relative population density (top panel) and wages (bottom panel) across the United States in the years 1930. Figure 13 shows the change in the topography of population densities (top panel) and wages (bottom panel) between 1930 and 2000.

Given $T_0$ and $T_1$ and the observed spatial distribution of economic activity $\{w_0, L_0\}$ and $\{w_1, L_1\}$, we follow the procedure presented in Section 3.1.2 to identify the unique topography of composite productivities and amenities $\{A_0, u_0\}$ and $\{A_1, u_1\}$ that is consistent with the observed distribution of economic activity and trade costs. Let $\varepsilon_A(i) \equiv \frac{\bar{A}_1(i)}{\bar{A}_0(i)}$ and $\varepsilon_u(i) \equiv \frac{\bar{u}_1(i)}{\bar{u}_0(i)}$ denote the change in the exogenous productivities and amenities, respectively. Then taking

[22] The different measure of wages is necessary because payroll data was not collected in 1930.

logs of equations (1) and (2) yields:

$$\ln\left(\frac{A_1(i)}{A_0(i)}\right) = \alpha \ln\left(\frac{L_1(i)}{L_0(i)}\right) + C_A + \ln \varepsilon_A(i) \tag{26}$$

$$\ln\left(\frac{u_1(i)}{u_0(i)}\right) = \beta \ln\left(\frac{L_1(i)}{L_0(i)}\right) + C_U + \ln \varepsilon_u(i), \tag{27}$$

where $C_A$ and $C_U$ capture changes in the normalization (since composite amenities and productivities are only identified up to scale).[23] If $\varepsilon_A(i)$ and $\varepsilon_u(i)$ are uncorrelated with the log change in population density, unbiased estimates of the parameters $\alpha$ and $\beta$ governing the strength of spillovers can be estimated from equations (26) and (27) using ordinary least squares (OLS). However, because $\varepsilon_A(i)$ and $\varepsilon_u(i)$ reflect changes in the exogenous productivities and amenities and the labor supply is determined in part by productivities and amenities (see equation 13), the model suggests that the OLS estimates of $\alpha$ and $\beta$ will be biased.

To overcome this problem, we estimate equations (26) and (27) using an instrumental variable (IV) procedure. We use as our instrument $\ln \hat{P}(i)$, which we define as the (log) difference between the price index in location $i$ in the year 1930 and what the price index in the year 1930 would have been if the IHS had already existed, holding fixed the observed wages and calibrated composite amenities, i.e.:

$$\ln \hat{P}(i) \equiv \frac{1}{1-\sigma} \ln \frac{\int_S T_0(s,i)^{1-\sigma} A_0(s)^{\sigma-1} w_0(s)^{1-\sigma} ds}{\int_S T_1(s,i)^{1-\sigma} A_0(s)^{\sigma-1} w_0(s)^{1-\sigma} ds}.$$

This instrument captures the increase in market access due to the decline in bilateral trade costs resulting from the construction of the IHS, i.e. it is a location-specific measure of the impact of the IHS. By holding the wages and calibrated composite amenities fixed, variation in $\ln \hat{P}(i)$ arises only from the effect of the IHS on the average trade cost location $i$ incurs with its trading partners, where the average is weighted by the importance of the trading partner in the consumption bundle of consumers in location $i$. Instrumental variables estimation will yield unbiased estimates of $\alpha$ and $\beta$ as long as $\ln \hat{P}(i)$ is uncorrelated with exogenous

[23]Because the equilibrium distribution of economic activity across space is only affected by relative differences in exogenous productivities and amenities, without loss of generality we can normalize $\int_S \bar{A}(s)\,ds = \int_S \bar{u}(s)\,ds = 1$. However, without knowledge of $\alpha$ and $\beta$, this normalization does not help determine the scale of the composite productivities and amenities. We therefore normalize $\int_S A(s)\,ds = \int_S u(s)\,ds = 1$ and include a constant to control for the change in scale.

changes in amenities and productivities conditional on the change in the population.[24] This seems a reasonable assumption, as the IHS was designed to connect major cities, connect to international road networks, and for national defense purposes, so its effect on the average bilateral trade costs of a location is unlikely to be systematically correlated to the changes in fundamentals of that location (see e.g. Michaels (2008)). Figure 14 presents a map of the impact of the IHS on the price index; while overall market access in the western United States increased the most from the construction of the IHS (as the pre-existing road density was lowest there), there exists substantial variation both within and across states.

The results of the estimation of equations (26) and (27) are presented in Table 2. The first two columns present the OLS results; the productivity spillover $\alpha$ is estimated to be 0.459, and the amenity spillover $\beta$ is estimated to be $-0.339$. The third and fourth columns present the IV results. These estimates are substantially greater in magnitude; the productivity spillover $\alpha$ is estimated to be 1.545, and the amenity spillover is estimated to be $-1.640$. Part of the reason for the substantial difference between the OLS and IV results may be because the IV regression assumes that the first-stage relationship between $\ln \hat{P}(i)$ and $\ln \hat{L}(i)$ is the same for the entire United States. In the fifth and sixth columns, we allow the first-stage relationship to vary across states.[25] In this case, the productivity spillover $\alpha$ is estimated to be 0.350, and the amenity spillover is estimated to be $-0.336$. Because of the additional flexibility in the first stage relationship, these are our preferred estimates in what follows. All estimates are precisely estimated and strongly statistically significant.

How do these estimates compare to results elsewhere? Rosenthal and Strange (2004) review the literature and find that doubling the size of city increases productivity by roughly 3 to 8 percent; our preferred IV estimate suggests a higher figure of 35 percent. However, our measure of productivity captures both the measure of variates produced in a location and the productivity per worker of each variety (see Appendix A.2). In particular, in a monopolistic competition setup with free entry with a fixed cost of labor, $\frac{1}{\sigma-1}$ of the $\alpha$ will be due to productivity gains resulting from the increased number of varieties in markets with larger population. Thus, given the assumed elasticity of substitution of nine, our estimate of $\alpha$ implies that the elasticity of productivity per worker to the population density is 22.5 percent. If amenity spillovers are interpreted as the share of income spent on non-tradable

[24] This instrumental variable strategy is similar in spirit to Ahlfeldt, Redding, Sturm, and Wolf (2012), whose identification strategy relies on the assumption that the distance to the Berlin wall was uncorrelated with changes in exogenous residential and production fundamentals.

[25] Because this requires estimating an intercept and slope coefficient for each state in the first stage, we follow Stock and Yogo (2002) and use a Limited Information Maximum Likelihood (LIML) estimator.

goods (see Appendix A.2), the IV estimates of $\hat{\beta}$ imply that 25% $(-\frac{\beta}{1-\beta})$ of income is spent on non-tradables. In comparison, the 2011 Consumer Expenditure Survey finds the average household spent 33.8% of their income on housing (BLS, 2011).

Given estimates $\hat{\alpha}$ and $\hat{\beta}$ from columns five and six of Table 2, we estimate the geography of exogenous productivities and amenities from the observed distribution of population and the calculated composite productivities and amenities using equations (1) and (2), i.e. $\bar{A}(i) = A(i) L(i)^{-\hat{\alpha}}$ and $\bar{u}(i) = u(i) L(i)^{-\hat{\beta}}$. Figures 15 and 16 present the topography of exogenous productivities (top panel) $\bar{A}$ and exogenous amenities (bottom panel) $\bar{u}$ for the years 1930 and 2000, respectively. In 1930, the exogenous productivity was high in the northeast, great plains, and west coast, while it was lower in the southwest; amenities were low in the west and high in the south and the Appalachia. In 2000, both exogenous productivity was highest in the Rocky Mountain states and lowest in the south; exogenous amenities were highest along the coasts and in the south and lowest in the Rocky Mountains.

## 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 equation (13) with utility equalization along with some algebra yields the following expression:

$$\ln Y(i) = C + \delta_1 \ln \bar{A}(i) + \delta_2 \ln \bar{u}(i) + \delta_3 \ln P(i), \tag{28}$$

where $C$ is a scalar, and the coefficients are known functions of $\sigma$, $\alpha$, and $\beta$.[26] Equation (28) provides a log linear relationship between the observed income in location $i$, the exogenous productivities and amenities, and the price index. Note that the price index is a sufficient statistic for the effect of geographic location on income, as it is the only term that includes the bilateral trade costs. 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 (28).

There are two caveats that should be mentioned with this approach. First, the relative contribution of the two terms will depend in part on the chosen parameters $\alpha$ and $\beta$ governing productivity and amenity spillovers, respectively. For robustness, we implement the Shapley decomposition for a wide range of parameters other than our estimated values. Second, if

[26] Equation (28) can also be arrived at by integrating the differential equations (17) and (18).

the trade cost function $\tau(i)$ is mis-specified, the model may fail to capture the observed differences in income without relying on exogenous variation in productivity and amenities even if no such exogenous variation is necessary. In such an event, the estimated contribution of the price index would be biased downwards. 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.

Figures 17 and 18 report the fraction of the spatial variation in income that can be attributed to geographic location for the years 1930 and 2000, respectively, for all constellations of $\alpha \in [0,1]$ and $\beta \in [-1,0]$. The star indicates the location of the preferred IV estimates of $\hat{\alpha}$ and $\hat{\beta}$. At these estimates, 43.2% of the observed spatial variation in income is due to geographic location in the year 1930, and 24.0% of the observed spatial variation in income is due to geographic location in the year 2000. These results remain mostly unchanged for different strengths of spillovers, with the exception of spillover combinations near the black-hole threshold $\gamma_1 = 0$. Hence, the results suggest that a substantial fraction of the spatial variation in incomes across the United States can be explained by geographic location, although the importance of geographic location has declined over the past seventy years.

## 3.3 Welfare effects of the Interstate Highway System

In this section, we examine the welfare implications of the interstate highway system. This has the additional benefit of providing a check of the validity of the model, as we can compare the predicted and observed redistribution of economic activity across space. In what follows, we set parameters $\alpha = 0.35$ and $\beta = -0.35$. These values (which are well within the confidence intervals of the 2SLS estimates) ensure a unique spatial equilibrium, thereby avoiding problems of performing counterfactuals in the presence of multiple equilibria.[27]

It is first helpful to use the model to characterize the total change in welfare between the years 1930 and 2000. To do so, we calculate the welfare (up to scale) for each year using the observed wages, composite amenities, and use equation (4) to impute the price index given the estimated bilateral trade costs and composite productivities. Given the strength of spillovers, we can then determine the scale of the welfare for each year by imposing that $\int_S \bar{A}(s)\,ds = \int_S \bar{u}(s)\,ds = 1$ in both years. We find that welfare was 65.7% greater in

[27]If we use the 2SLS estimates $\alpha = 0.350$ and $\beta = -0.336$, we find the algorithm which guarantees uniform convergence for $\alpha + \beta \leq 0$ converges to an alternative equilibrium even for initial guesses of the population density near the observed distribution in 1930. It is interesting to note that this alternative equilibrium is characterized by a substantial increase in the population in the western United States (especially in California) and is associated with 17% higher welfare than the observed spatial equilibrium.

2000 than in 1930. Because the exogenous amenities and productivities are normalized to have a unit mean in both years, this welfare increase is due only to changes in the spatial distribution of $\bar{A}$ and $\bar{u}$ (and the resulting reallocation of the population) and the reduction in trade costs associated with the construction of the interstate highway system.

How much of this welfare gain can be attributed to the construction of the interstate highway system? To answer this question, we hold the topography of productivities $\bar{A}$ and amenities $\bar{u}$ fixed at the 1930 levels and recalculate the equilibrium spatial distribution of economic activity given the estimated bilateral trade costs after the construction of the IHS using equations (13) and (14). The model estimates that the construction of the IHS lead to a 3.47% increase in welfare, indicating that the majority of the welfare gain over the past seventy years is due to changes in the spatial distribution of exogenous amenities and productivities.

Given the estimate of the welfare gains from the IHS, a simple back-of-the-envelope calculation suggests that the benefits of the IHS substantially outweigh its costs. Duranton and Turner (2011) estimate that the annual total cost (interest payments on the initial construction plus maintenance costs) of each kilometer of interstate is between \$1.3 million and \$3.25 million (in 2007 dollars), depending on the population density. Given the approximately 76,000km of interstate highways, their estimates suggest a total annual cost of the IHS is between \$100 and \$250 billion. The U.S. GDP in 2007 was \$14.25 trillion; since preferences are assumed to be homothetic, if the IHS increased (static) welfare by 3.47%, its value is \$494.5 billion in 2007 dollars, suggesting a return on investment of at least 100%.

While the construction of the IHS cannot explain a majority of the change in welfare, its effect on the spatial distribution of economic activity accords well with the observed change in the spatial distribution of people and income. Figures 19 and 20 compare the observed change in the population and wages, respectively, between 1930 and 2000 (top panel) to the change arising from the construction of the IHS predicted by the model (bottom panel), where changes are measured in log differences. The model correctly predicts the increase in population density and income in the Southwest and Florida; overall, the correlation between the predicted and observed changes is 0.194 and 0.182 for population and wage changes, respectively. The model, however, substantially under-predicts the magnitude of the redistribution of population and wages. This reinforces the finding that the construction of the IHS contributed only a small portion to the changing topography of economic activity over the past seventy years.

# 4 Conclusion

We view this paper as taking a number of necessary steps toward the rigorous quantification of spatial theory. First, we develop a unified general equilibrium approach combining labor mobility, gravity, and productivity and amenity spillovers. Second, we provide a micro-foundation of trade costs as the accumulation of instantaneous trade costs over the least cost route on a surface. Combining the two allows us to determine the equilibrium spatial distribution of economic activity on any surface with (nearly) any geography. As a result, the framework can be applied directly 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: How would liberalizing a border redistribute the economic activity within each country? What will be the effect of the opening of the Northwest Passage on the global income distribution? How would removing restrictions on cross-country migration affect the equilibrium distribution of economic activity?

## 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*.

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.

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

BLS (2011): "2011 Consumer Expenditure Survey," http://www.bls.gov/cex/csxann11.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.

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

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

Cosar, K. A., and P. Fajgelbaum (2012): "Internal Geography, International Trade, and Regional Outcomes," *mimeo*.

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.

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. Turner (2011): "Urban growth and transportation," Working Paper.

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. (2012): "Trade and Interdependence in a Spatially Complex World," *mimeo, Penn State*.

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.

HASSOUNA, M., AND A. FARAG (2007): "Multistencils Fast Marching Methods: A Highly Accurate Solution to the Eikonal Equation on Cartesian Domains," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(9), 1563–1574.

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.

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

KIMMEL, R., N. KIRYATI, AND A. BRUCKSTEIN (1996): "Sub-pixel distance maps and weighted distance transforms," *Journal of Mathematical Imaging and Vision*, 6(2), 223–233.

KIMMEL, R., AND J. SETHIAN (1998): "Computing Geodesic Paths on Manifolds," *Proceedings of the National Academy of Sciences*, 95(15), 8431–8435.

KRASNOSEL'SKII, M., AND L. BORON (1964): *Positive Solutions of Operator Equations*. P. Noordhoff Groningen.

KROON, D. (2009): "Accurate Fast Marching Toolbox," Matlab Toolbox available at http://www.mathworks.com/matlabcentral/fileexchange/24531.

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.

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.

Mémoli, F., and G. Sapiro (2001): "Fast Computation of Weighted Distance Functions and Geodesics on Implicit Hyper-surfaces," *Journal of Computational Physics*, 173(2), 730–764.

Michaels, G. (2008): "The effect of trade on the demand for skill: evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 90(4), 683–701.

Mitchell, J. (1988): "An Algorithmic Approach to Some Problems in Terrain Navigation," *Artificial Intelligence*, 37(1), 171–201.

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

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

Peyre, G. (2009): "Fast Marching Toolbox," Matlab Toolbox available at http://www.mathworks.com/matlabcentral/fileexchange/6110.

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.

RAMONDO, N., A. RODRÍGUEZ-CLARE, AND M. SABORIO-RODRIGUEZ (2012): "Scale Effects and Productivity Across Countries: Does Country Size Matter?," *mimeo, University of California, Berkeley.*

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.*

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.

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

STOCK, J. H., AND M. YOGO (2002): "Testing for weak instruments in linear IV regression," .

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

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: Estimated instantaneous trade costs

| | Estimate | Standard error | Implied trade cost for a coast-to-coast trip |
|---|---|---|---|
| Ocean | 1.03 | 0.162 | 180.8% |
| River | 1.40 | 0.086 | 306.9% |
| Interstate highways | 0.53 | 0.010 | 70.1% |
| Old highways | 1.17 | 0.043 | 221.4% |
| Mountains | 1.44 | 0.785 | 320.2% |
| Other land | 1.03 | 0.028 | 179.8% |
| Different state | 0.48 | 0.005 | |

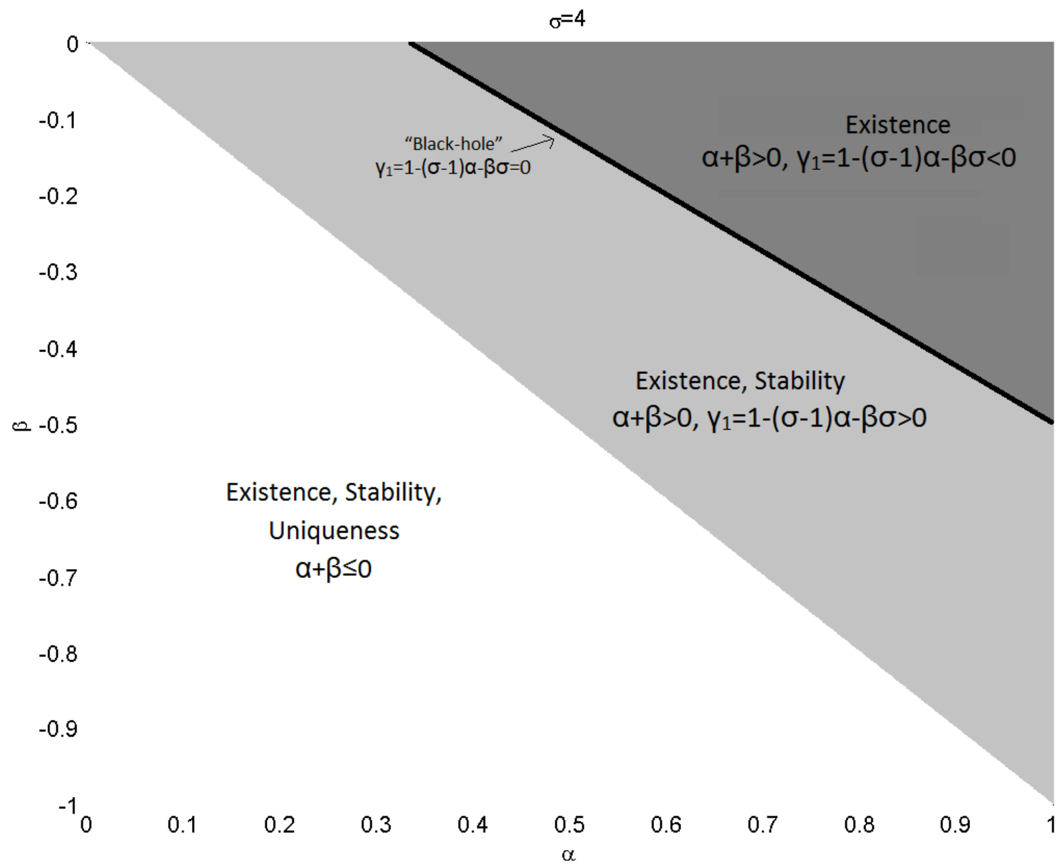
*Notes:* This table reports the estimated instantaneous trade cost of traveling over six types of locations as well as the estimated ad valorem cost of trading with other states. The distance from coast to coast is normalized to one, so that $e^x - 1$ is the total ad valorem trade cost of traveling from coast to coast across locations with instantaneous trade costs $x$. The estimation was done using a nonlinear least squares procedure to most closely match the normalized state to state bilateral trade flows observed in the 2007 Commodity Flow Survey. The elasticity of substitution ($\sigma$) is assumed to have a value of nine. The resolution of the United States is $61 \times 97$ pixels.

Table 2: Estimating the strength of spillovers

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent variable: | $\ln \hat{A}(i)$ | $\ln \hat{U}(i)$ | $\ln \hat{A}(i)$ | $\ln \hat{U}(i)$ | $\ln \hat{A}(i)$ | $\ln \hat{U}(i)$ |
| Change in relative | 0.459*** | -0.339*** | 1.545*** | -1.640*** | 0.350*** | -0.336*** |
| population density ($\ln \hat{L}(i)$) | (0.008) | (0.008) | (0.085) | (0.093) | (0.019) | (0.019) |
| Constant | 0.237*** | 0.180*** | 0.495*** | -0.129*** | 0.211*** | 0.181*** |
| | (0.008) | (0.007) | (0.025) | (0.028) | (0.010) | (0.009) |
| Identified parameter | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ |
| Estimator | OLS | OLS | 2SLS | 2SLS | LIML | LIML |
| Instrument | | | | | State-specific | State-specific |
| | | | $\ln \hat{P}(i)$ | $\ln \hat{P}(i)$ | $\ln \hat{P}(i)$ | $\ln \hat{P}(i)$ |
| First-stage F-stat | | | 291.646 | 291.646 | 223.278 | 223.278 |
| Observations | 13345 | 13345 | 13345 | 13345 | 13345 | 13345 |

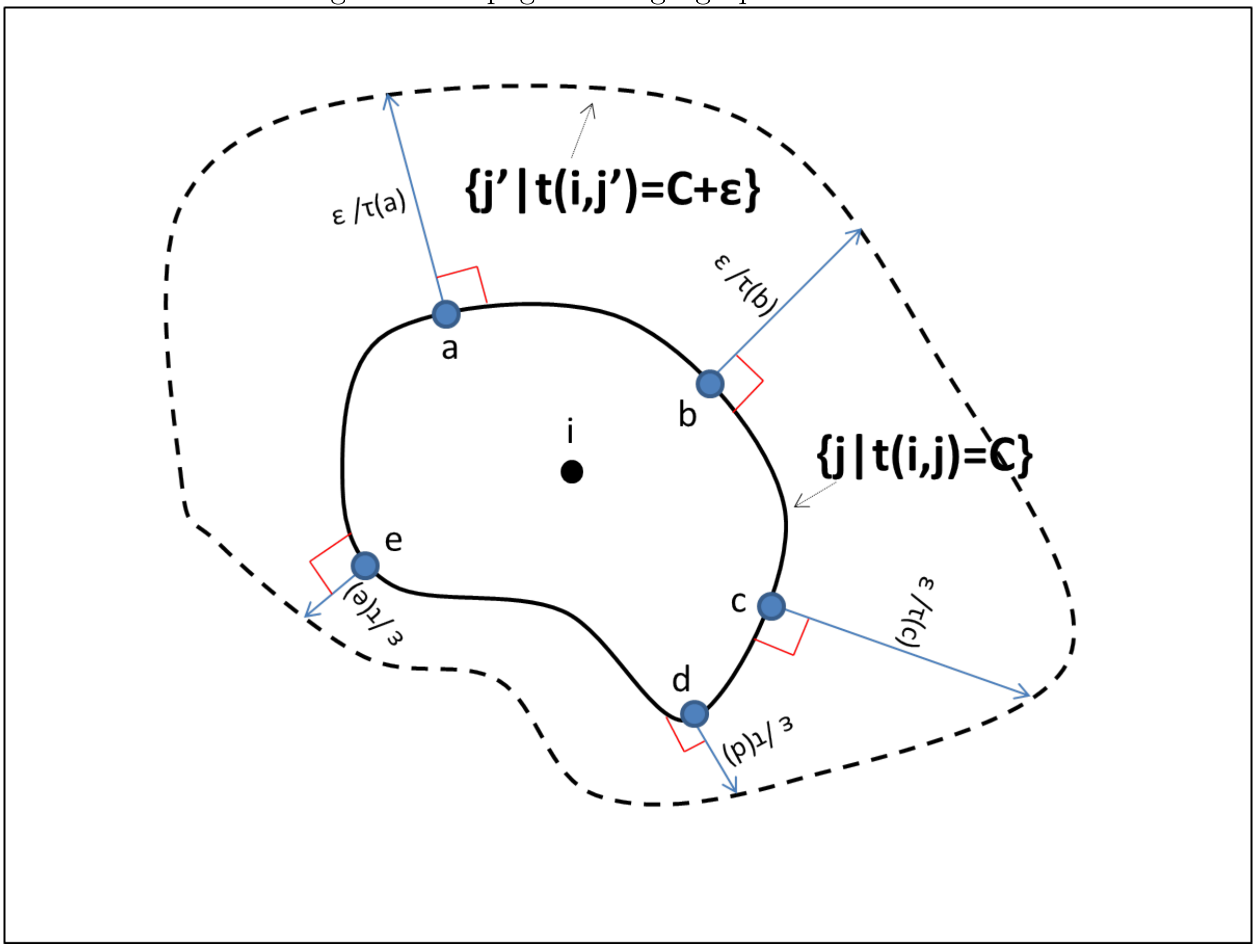
*Notes:* This figure shows the estimated strength of spillovers using the correlation between changes in the composite amenities and productivities and the change in the relative population density resulting from the construction of the U.S. Interstate Highway System. Each observation is a location in the United States; the resolution of the United States is $121 \times 193$. The instrument is the log difference between the price index in 1930 and what the price index would have been in 1930 if the Interstate Highway System had been in existence. Robust standard errors are reported in parentheses. Stars indicate statistical significance: * p<.10 ** p<.05 *** p<.01.

Figure 1: Equilibria with endogenous amenities and productivity



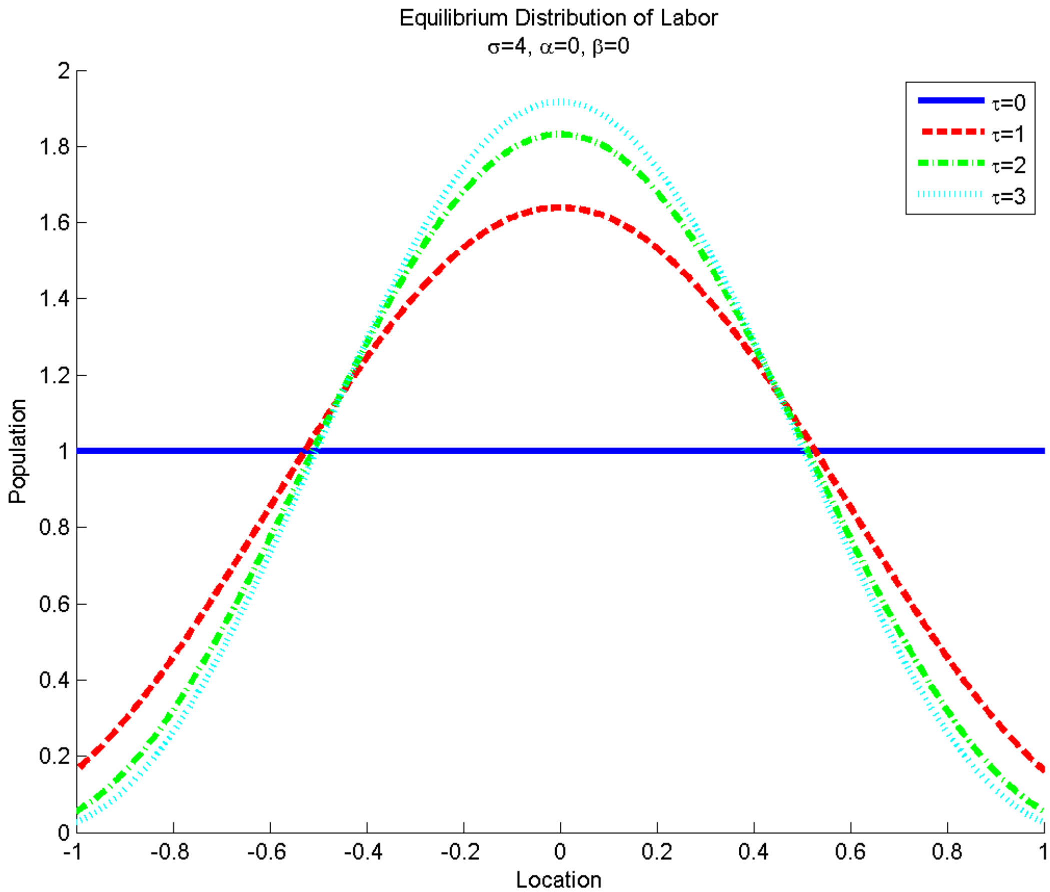

*Notes*: This figure shows the regions of values for the productivity externality $\alpha$ and the amenity externality $\beta$ for which there exists an equilibrium, for which there exists a point-wise locally stable equilibrium, and whether that equilibrium is unique. The elasticity of substitutions $\sigma$ is chosen to equal 4.

Figure 2: Propagation of geographic trade costs



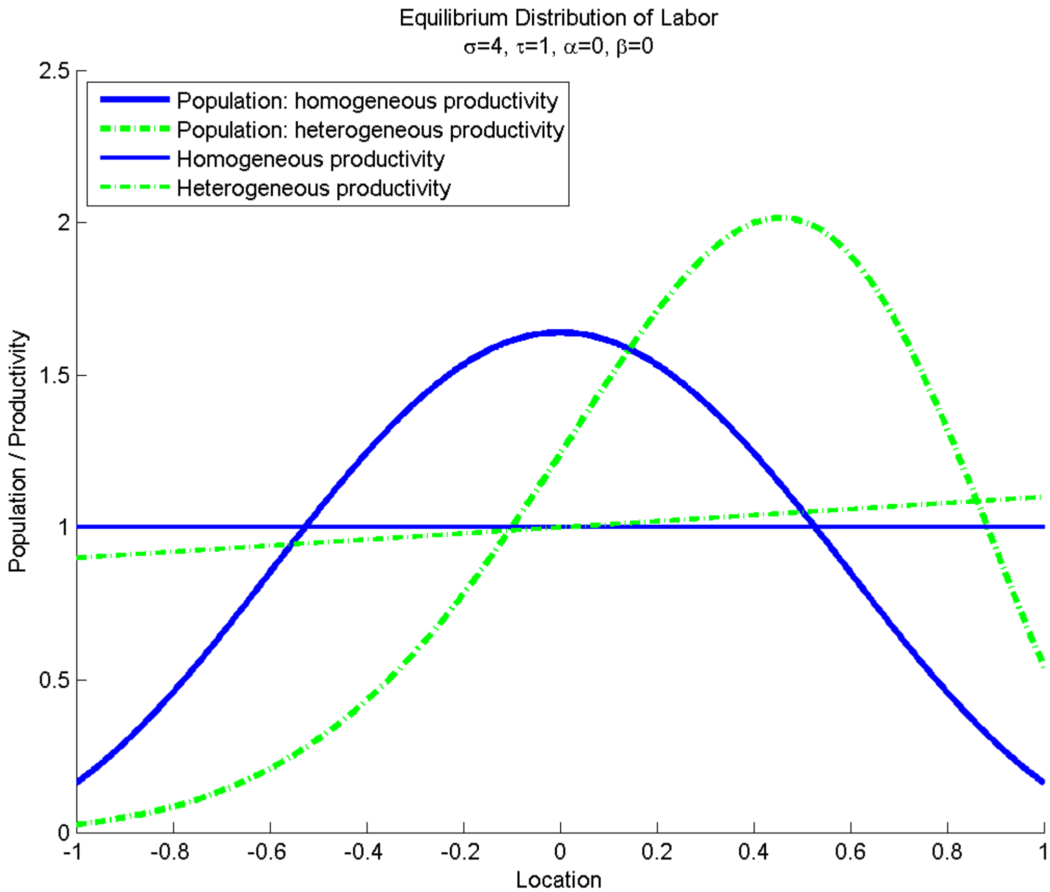

*Notes*: This figure shows how the geographic trade costs evolve across a surface. Given a contour of points on a surface such that the geographic trade cost to location $i$ is equal to a constant $C$ (the solid line), for an arbitrarily small $\varepsilon > 0$, we can construct the contour line for 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



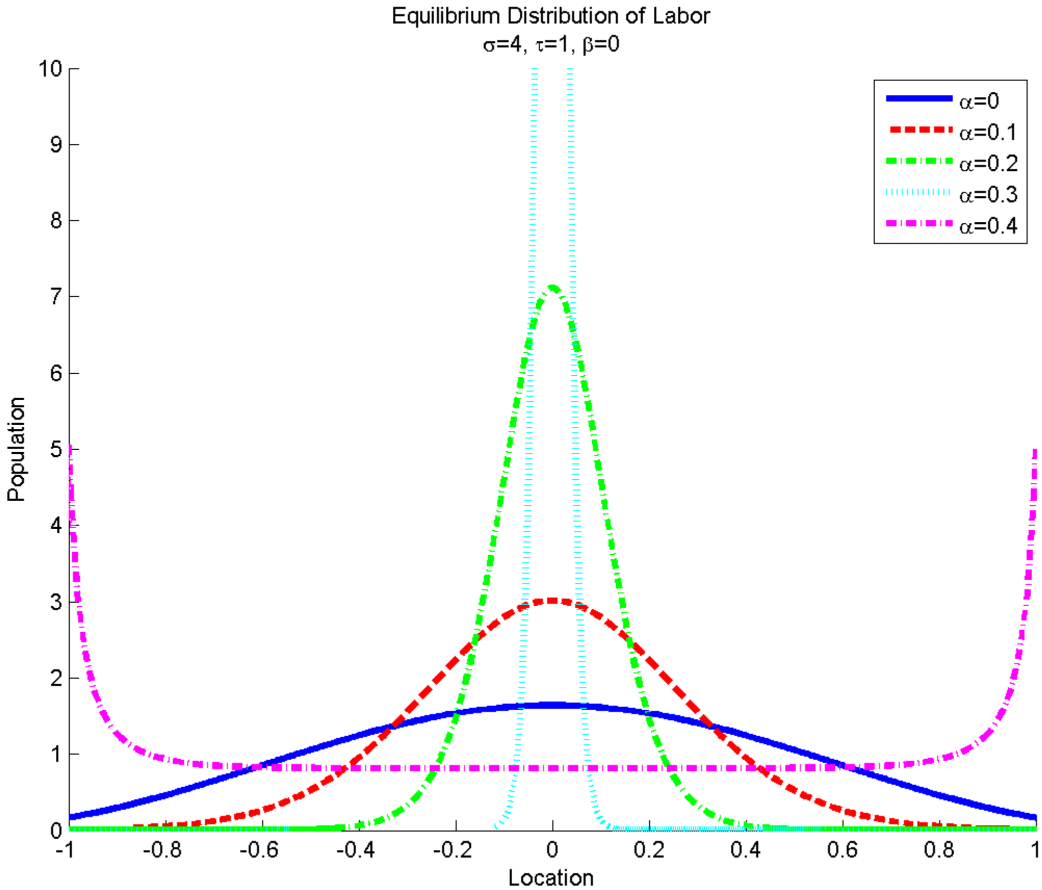

*Notes*: This figure shows how the equilibrium distribution of population along a line is affected by changes in the trade cost. When trade is costless, the population is equal along the entire line. As trade becomes more costly, the population becomes increasingly concentrated in the center of the line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

Figure 4: Economic activity on a line: Exogenous productivity differences



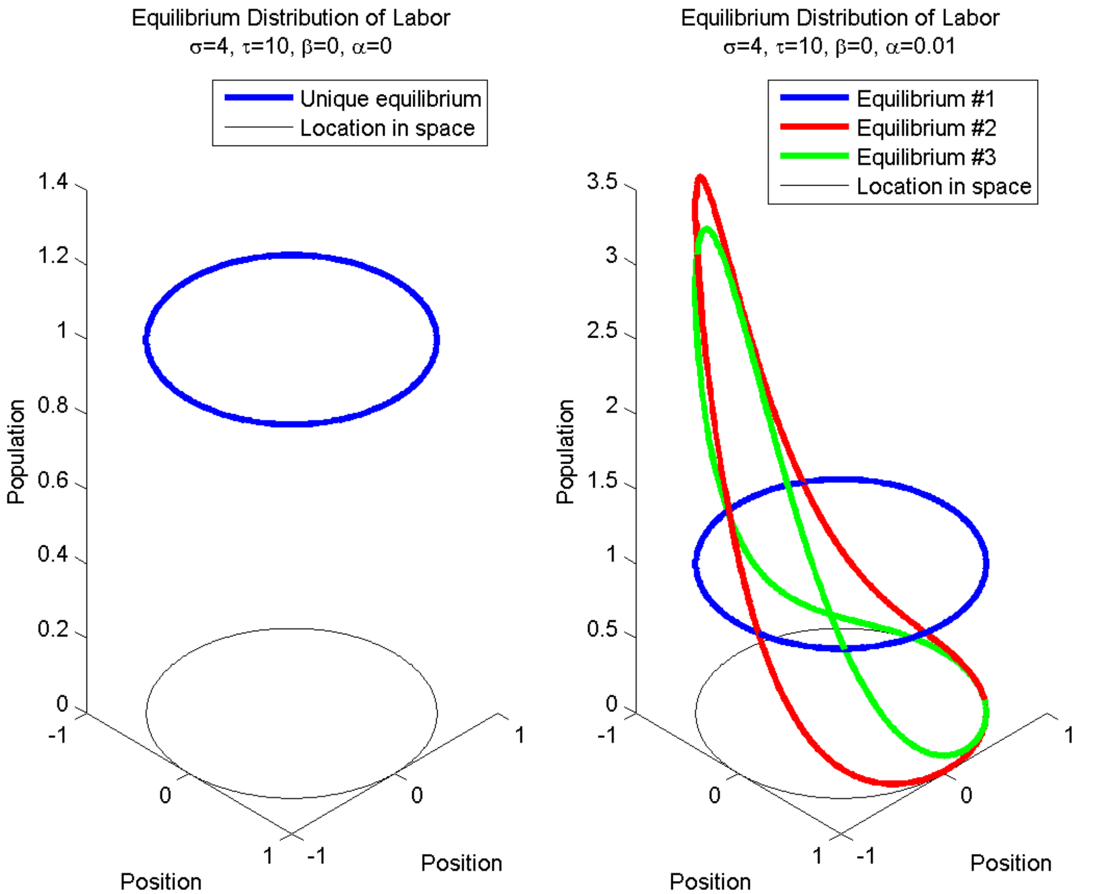

*Notes*: This figure depicts how the equilibrium distribution of population along a line is affected by exogenous differences in productivity across space. With homogeneous productivities, and positive trade costs, the population is concentrated at the center of the line. When productivity is higher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

Figure 5: Economic activity on a line: Productivity spillovers



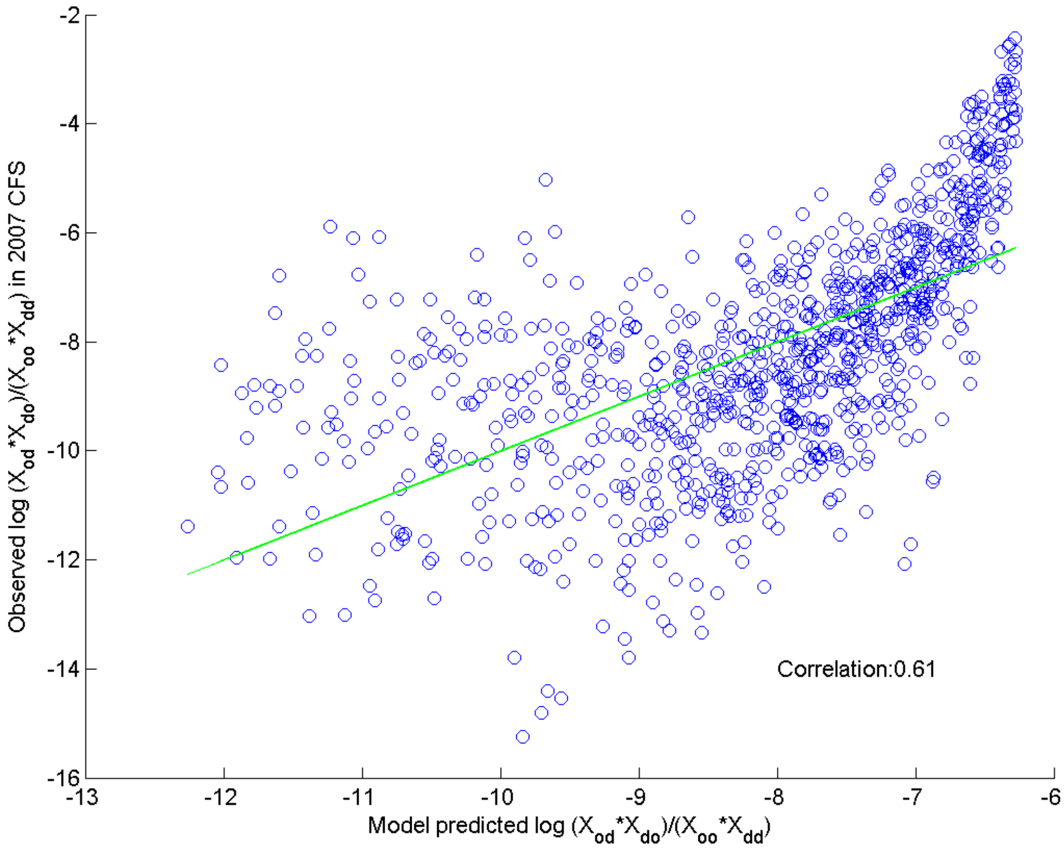

*Notes*: This figure shows how the equilibrium distribution of population along a line is affected by varying degrees of productivity spillovers. As the productivity spillovers increase, the population becomes increasingly concentrated in the center of the line. A non-degenerate equilibrium can be maintained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

Figure 6: Economic activity on a circle: Multiple equilibria



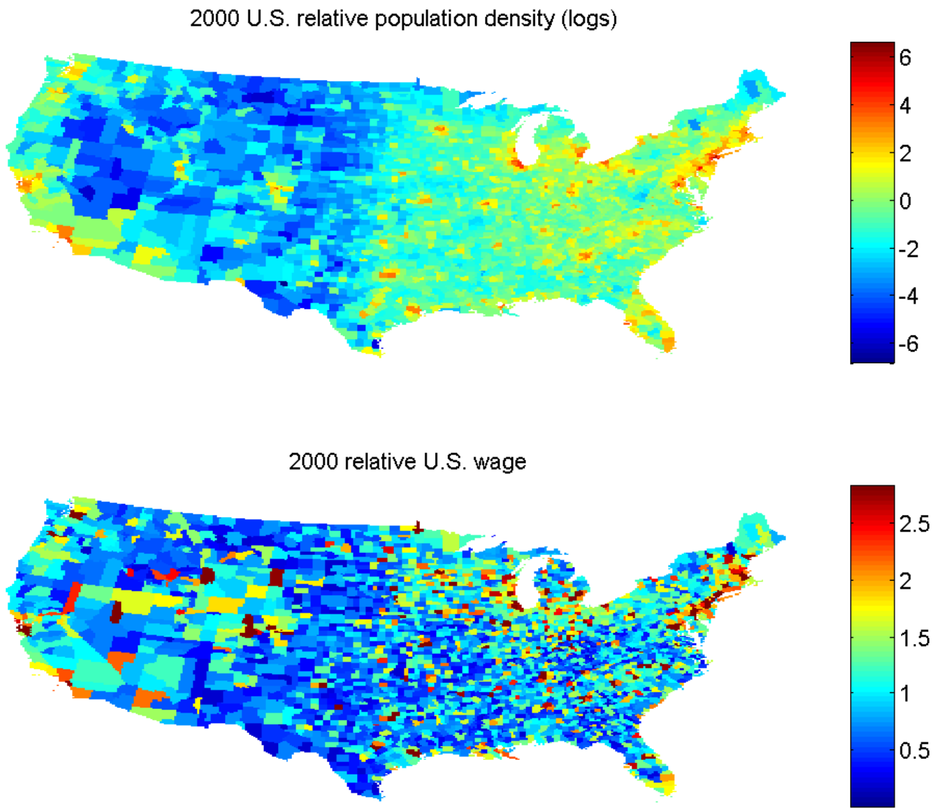

*Notes*: This figure provides an example of multiple equilibria when the surface is a one dimensional circle. The left panel shows the unique homogeneous distribution of population along the circle when $\alpha + \beta = 0$. When $\alpha + \beta > 0$ (here $\alpha = 0.01$ and $\beta = 0$), uniqueness 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 7: Estimating trade costs: Predicted versus observed state-to-state trade shares



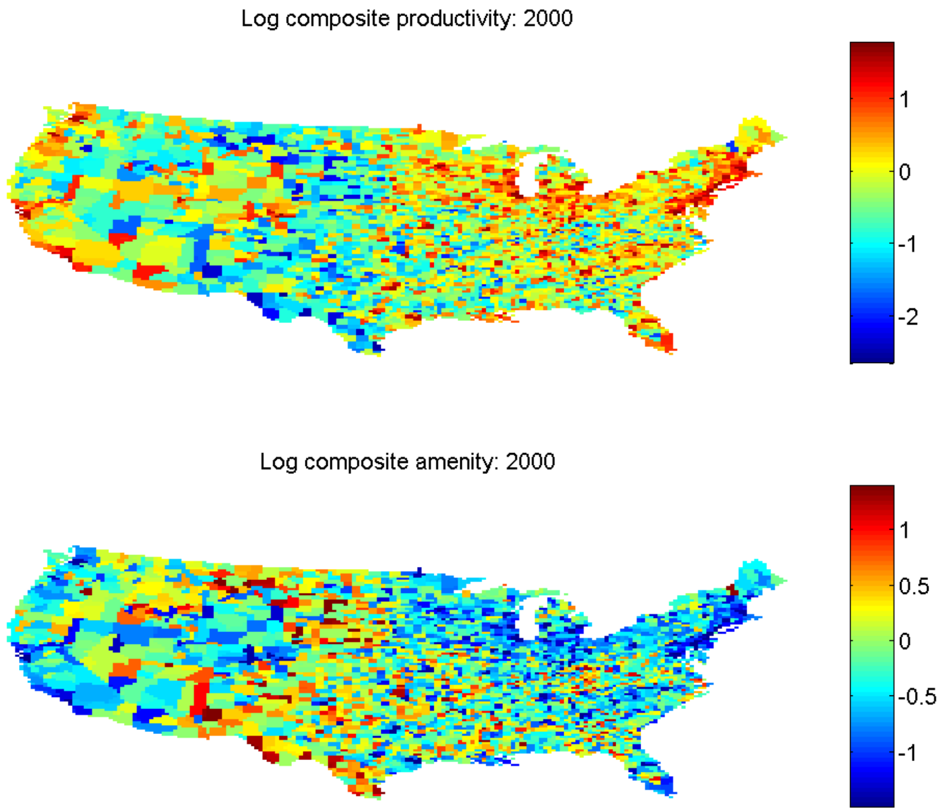

*Notes*: This figure shows how well the model captures the variation in state-to-state normalized trade shares observed in the 2007 Commodity Flow Survey given the estimated trade costs. The green line is the 45 degree line.

Figure 8: United States wages and population density in 2000



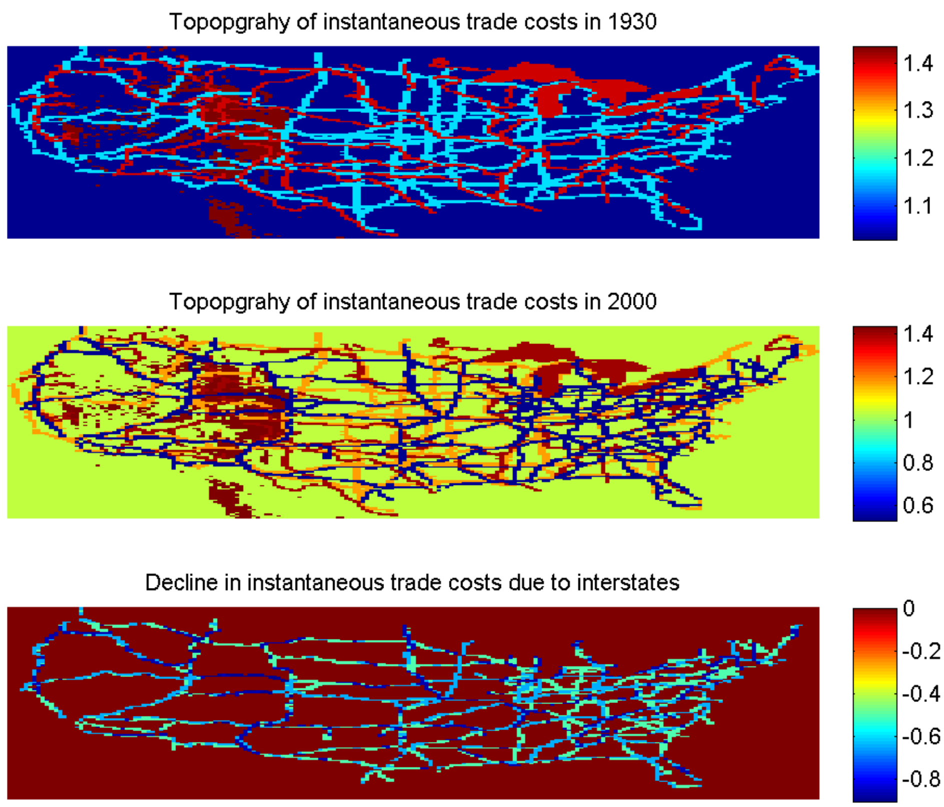

*Notes*: This figure shows the distribution of wages (measured as total payroll per capita) and population density within the United States in the year 2000. The data are reported at the county level and are normalized to have a unit mean. (Source: MPC (2011)).

Figure 9: United States: Composite productivities and amenities in the year 2000



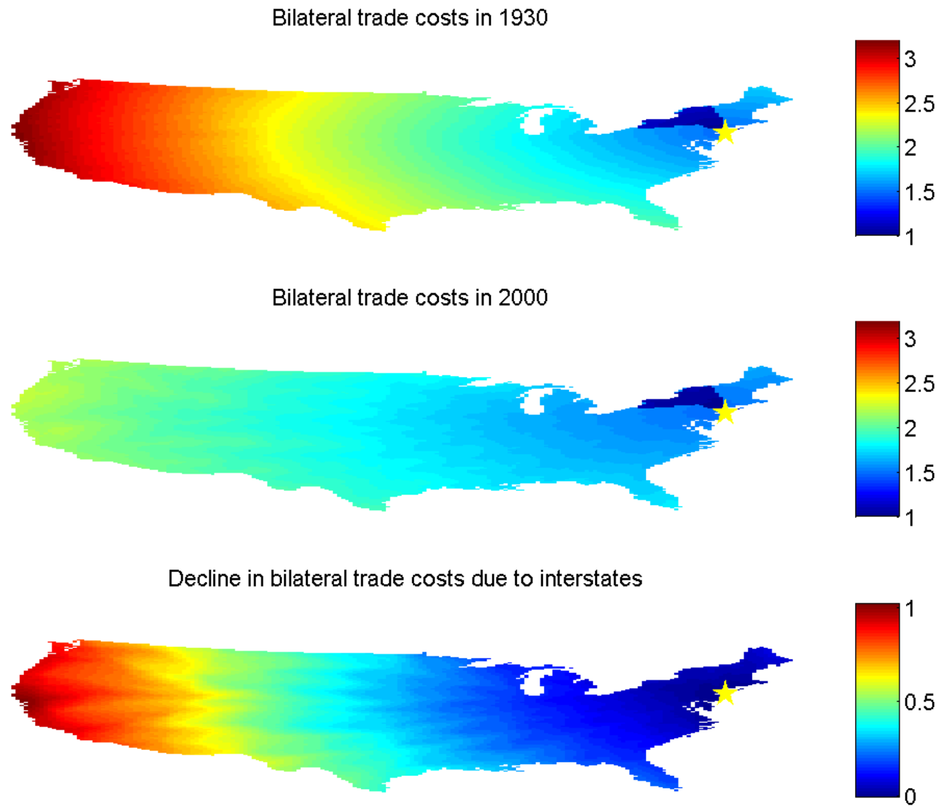

*Notes*: This figure shows the unique topography of log composite productivity (top panel) and log composite amenities (bottom panel) consistent with the observed topography of wages, population, and estimated bilateral trade costs in the year 2000.

Figure 10: United States: Instantaneous trade costs



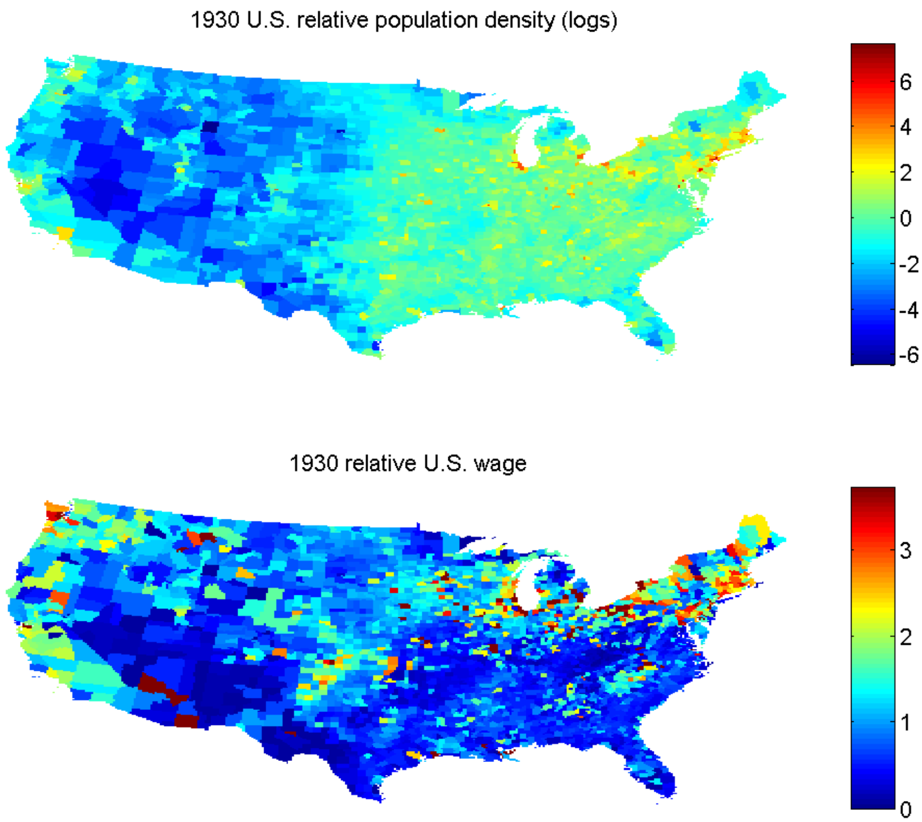

*Notes*: This figure shows the estimated topography of instantaneous trade costs for the United States. The top panel depicts the trade costs in 1930 (prior to the construction of the interstate highways), the middle panel depicts the trade costs in 2000 (after the construction of interstate highways) and the lower panel depicts the difference between the two. The values of the instantaneous trade costs reported are the log bilateral trade cost of traveling the width of the figure (e.g. an instantaneous trade cost of 4 indicates the ad valorem tax equivalent trade cost of the width of the figure is $e^4 - 1$).

Figure 11: United States: Trade costs from New York



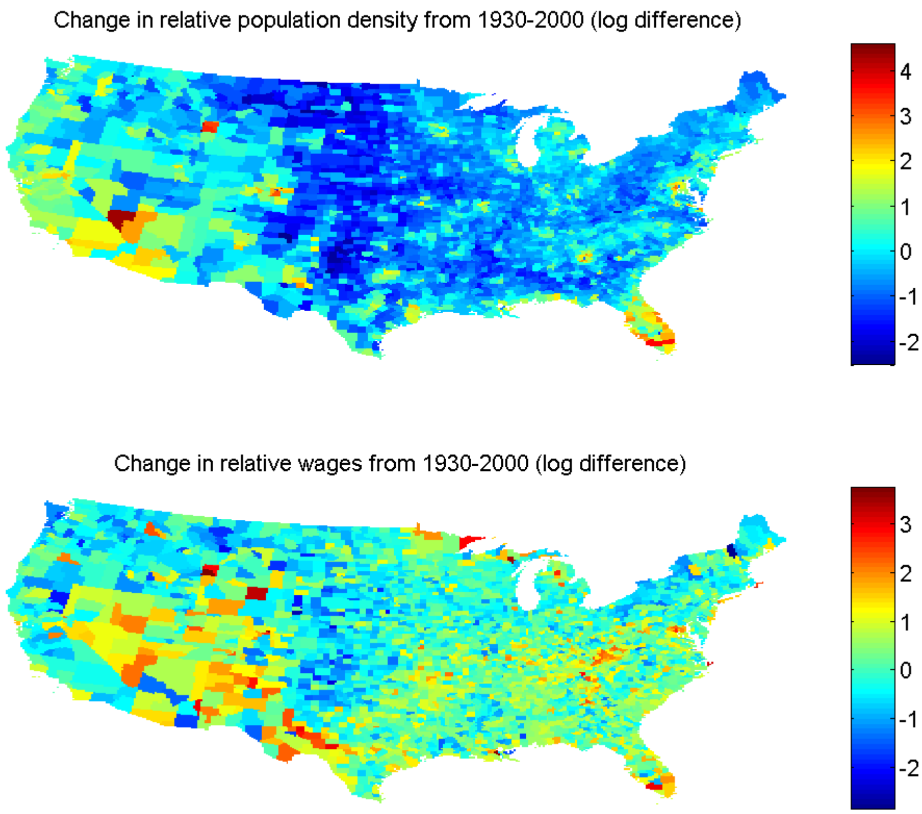

*Notes*: This figure depicts the bilateral trade costs along the optimal routes from New York, NY. The top panel reports the log bilateral trade costs in 1930 (prior to the construction of the interstate highway system), the middle panel reports the log bilateral trade costs in 2000 (after the construction of the interstate highway system), and the bottom panel reports the difference between the two.

Figure 12: United States wages and population density in 1930



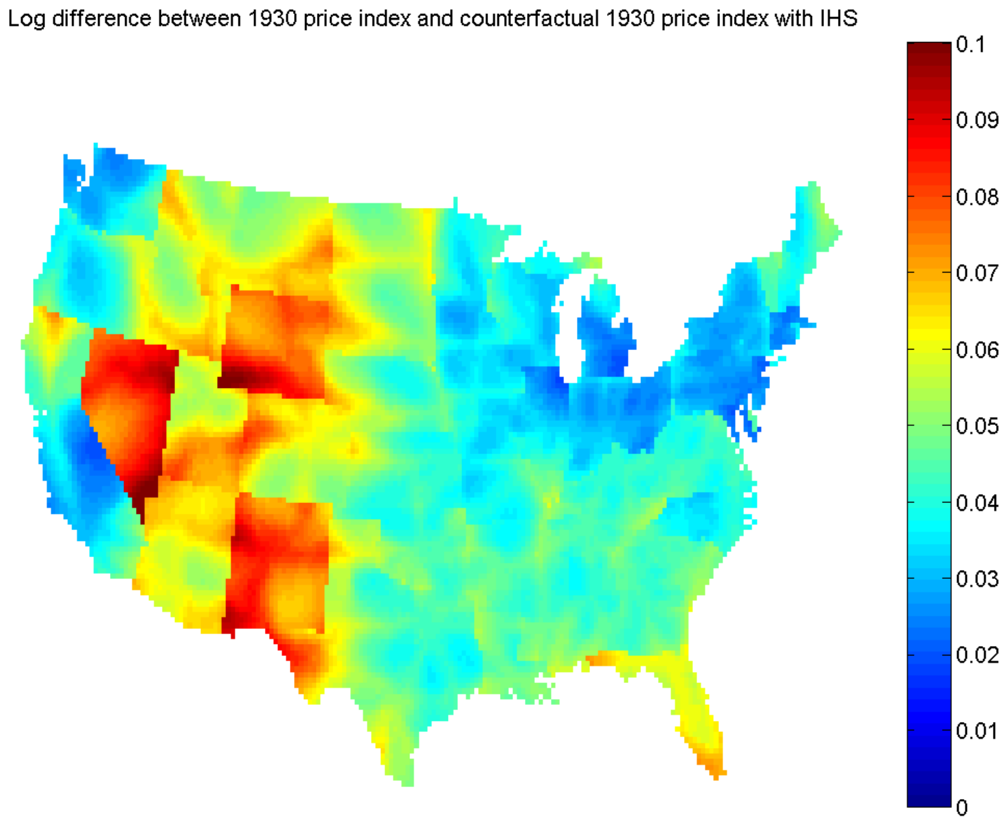

*Notes*: This figure shows the distribution of wages (measured as manufacturing and agriculture output per capita) and population density within the United States in the year 1930. The data are reported at the county level and are normalized to have a unit mean. (Source: MPC (2011)).

Figure 13: United States: Change in the population and wages



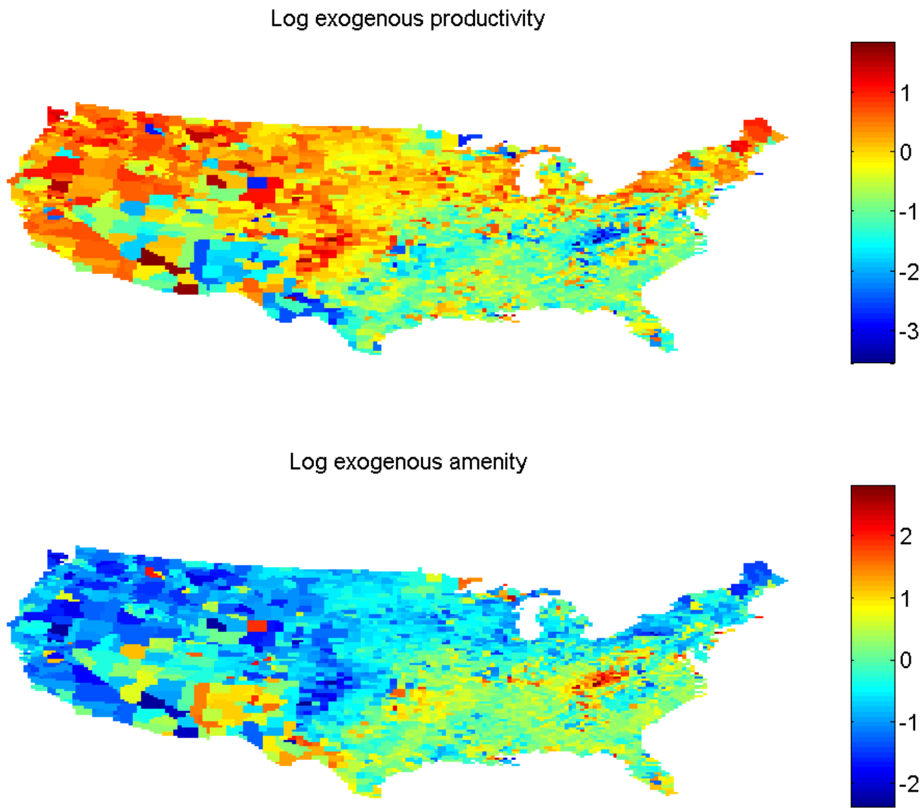

*Notes*: This figure shows the change in the relative wages and population density between 1930 and 2000. Wages and population density are normalized to have a unit mean in each year and the change is measured as the log difference in the normalized values between the two years. (Source: MPC (2011)).

Figure 14: United States: Effect of the highway system on the price index



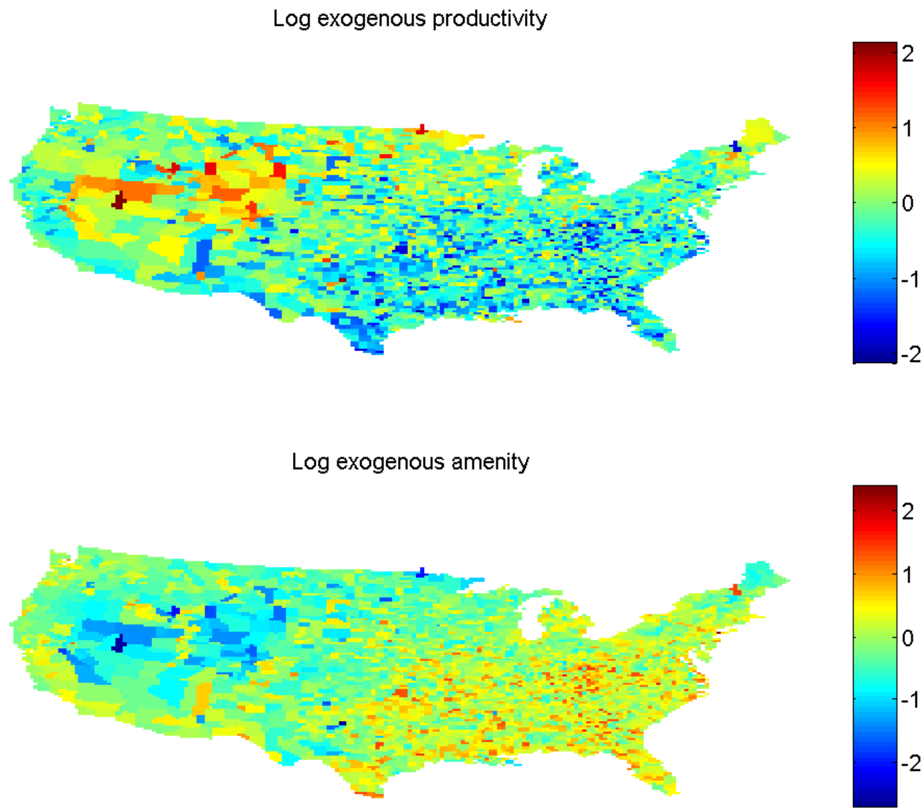

*Notes*: This figure shows the log difference in the calibrated price index in 1930 and the counterfactual price index in 1930 holding fixed the calibrated amenities and observed wages but changing the bilateral trade costs to allow for travel over interstate highways. As a result, the figure shows the relative impact of the interstate highway system on different locations, where a higher value (indicating a greater potential fall in the price index) indicates a greater impact.

Figure 15: United States: Topography of exogenous productivities and amenities in 1930



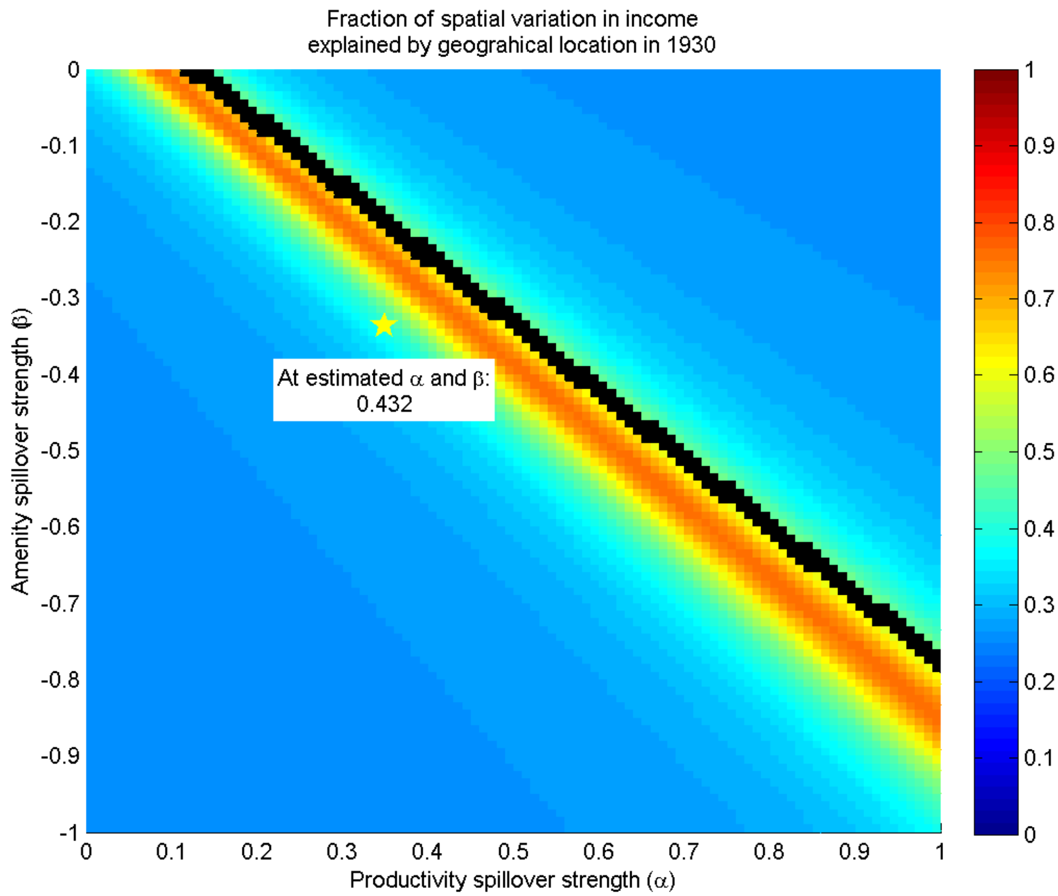

*Notes*: This figure topography of (log) exogenous productivities and amenities in the U.S. in the year 1930 using the estimated $\alpha$ and $\beta$ parameters governing the strength of productivity and amenity spillovers. Both amenities and productivities are normalized to have a mean of one.

Figure 16: United States: Topography of exogenous productivities and amenities in 2000



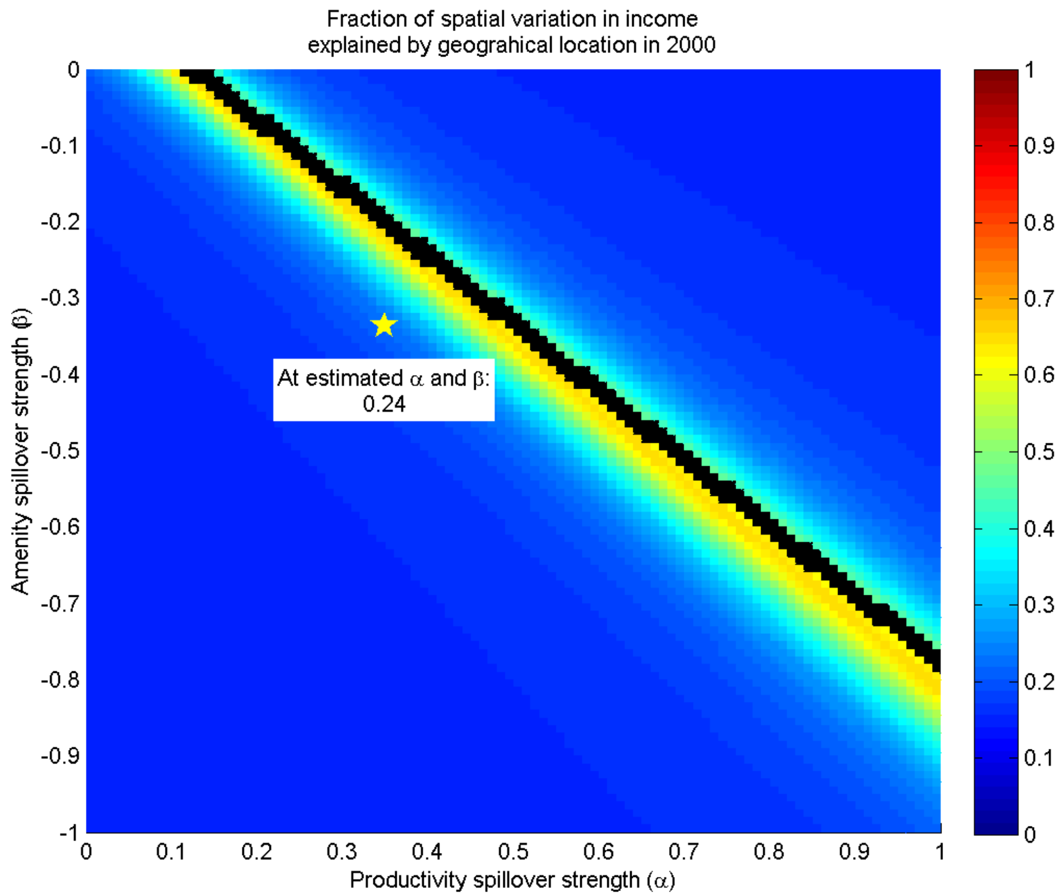

*Notes*: This figure topography of (log) exogenous productivities and amenities in the U.S. in the year 2000 using the estimated $\alpha$ and $\beta$ parameters governing the strength of productivity and amenity spillovers. Both amenities and productivities are normalized to have a mean of one.

Figure 17: United States: Fraction of spatial inequality of income due to geographic location in 1930



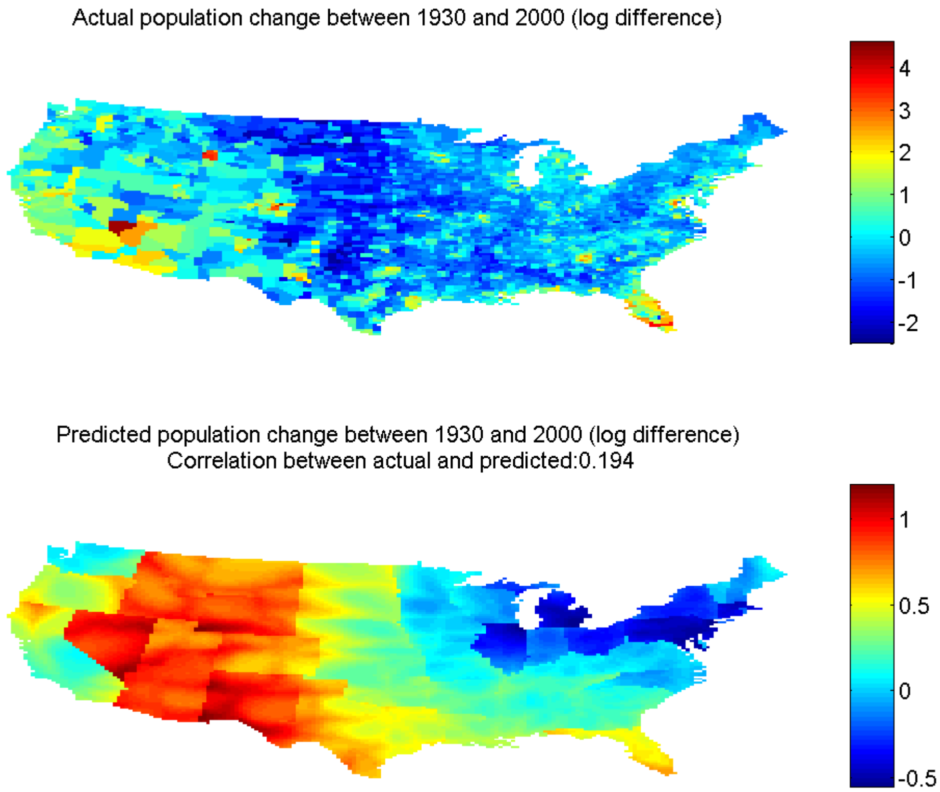

*Notes*: This figure shows the fraction of the observed variation in income across space in the U.S. in the year 1930 that is due to geographic location. The decomposition is calculated for all constellations of productivity spillover strength $\alpha \in [0, 1]$ and $\beta \in [-1, 0]$ (except for those sufficiently near the "black-hole" condition; see Figure 1). The star indicates the estimated spillover strength using the construction of the interstate highway system.

Figure 18: United States: Fraction of spatial inequality of income due to geographic location in 2000



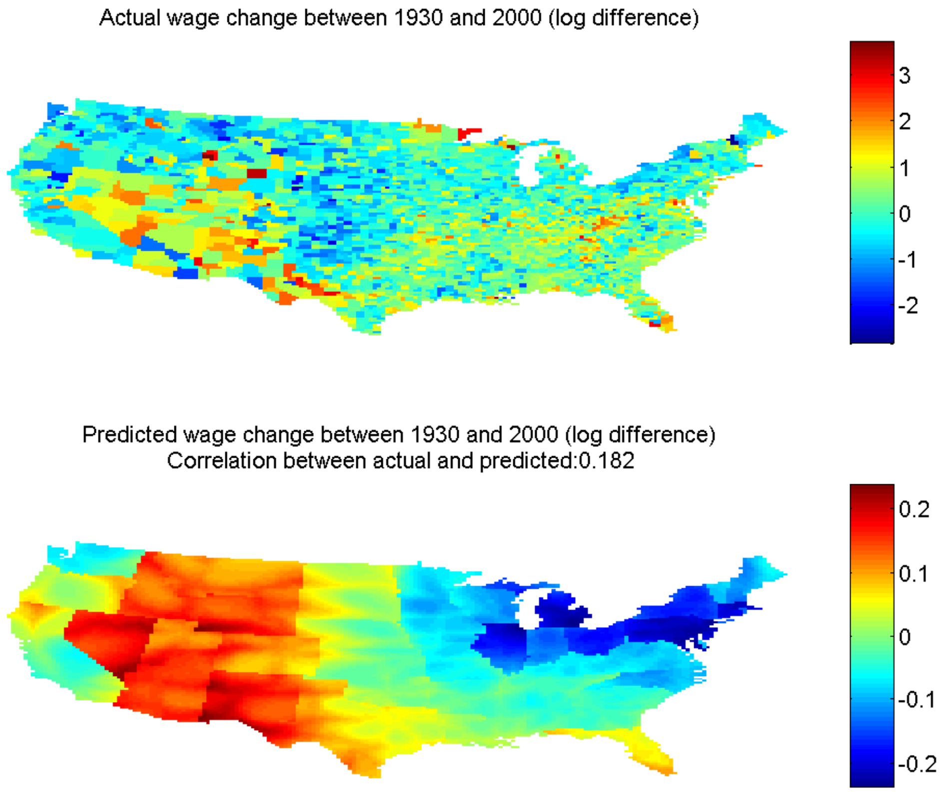

*Notes*: This figure shows the fraction of the observed variation in income across space in the U.S. in the year 2000 that is due to geographic location. The decomposition is calculated for all constellations of productivity spillover strength $\alpha \in [0,1]$ and $\beta \in [-1,0]$ (except for those sufficiently near the "black-hole" condition; see Figure 1). The star indicates the estimated spillover strength using the construction of the interstate highway system.

Figure 19: United States: Predicted versus actual effect of IHS on population density




*Notes*: The top panel of this figure shows the observed change in the relative population (measured in log differences) between the years 1930 and 2000. The bottom panel of the figure shows the predicted change in the population (measured in log differences) as a result of the construction of the Interstate Highway System.

Figure 20: United States: Predicted versus actual effect of IHS on wages




*Notes*: The top panel of this figure shows the observed change in the relative wages (measured in log differences) between the years 1930 and 2000. The bottom panel of the figure shows the predicted change in the wages (measured in log differences) as a result of the construction of the Interstate Highway System.

# A 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 (11) can be written as:

$$g\left(i\right) = \lambda \int_S K\left(s,i\right) g\left(s\right) ds, \tag{29}$$

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 (12) in an identical form:

$$f\left(i\right) = \lambda \int_S K\left(i,s\right) f\left(s\right) ds, \tag{30}$$

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 (29) and (30) 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 $\mathbb{R}$.[28]), there exists a unique (to-scale) strictly positive function $g(i)$ and constant $\lambda_1$ that solves equation (29) and a unique (to-scale) strictly positive function $f(i)$ and constant $\lambda_2$ that solves equation (30).

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)[29]), because the kernel of equation (30) is the transpose of the kernel equation (29), $\lambda_1$ is a characteristic value of equation (30) and $\lambda_2$ is a characteristic value of equation (29). In addition, notice that the constants $\lambda_1$ and $\lambda_2$ correspond to the smallest characteristic values of equations (29) and (30), respectively.[30] Suppose that $\lambda_1 > \lambda_2$, we will arrive at a contradiction. In that case equation (30) has a characteristic value smaller than $\lambda_2$, 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 (29) and (30). 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

$$f_{n+1}(i) = \frac{\int_S K(i,s)\, f_n(s)\, ds}{\int_S \int_S K(i,s)\, f_n(s)\, ds di}, \tag{31}$$

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 (31) 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

[28]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.

[29]Note that because $\lambda_1$ is real, the complex conjugate of $\lambda_1$, $\bar{\lambda}_1$, is equal to $\lambda_1$ (and likewise for $\lambda_2$ and $\bar{\lambda}_2$).

[30]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 (31) 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.

**Discrete number of locations** Theorem 1 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 is coming from the fact that for a square matrix its eigenvalue is the same as the eigenvalue of its transpose. The matrix operator in this case is ergodic in the sense that an iterative approach as the one in (31) converges to the true solution. The algorithm in this case is the same as for the case of continuous variables.

### A.1.2 Proof of Theorem 2

We first show that if there exists a regular spatial equilibrium, then equation (13) is the unique relationship between $w(i)$ and $L(i)$ that satisfies equations (11) and (12). 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 (11) and (12). 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 (11) and (12) we have:

$$\phi(i) = \frac{\int_S T(i,s)^{1-\sigma}\, \bar{u}(s)^{\sigma-1}\, L(s)^{1+\beta(\sigma-1)}\, w(s)^{\sigma}\, ds}{\int_S T(s,i)^{1-\sigma}\, \bar{A}(s)^{\sigma-1} w(s)^{1-\sigma}\, L(s)^{\alpha(\sigma-1)}\, ds} \iff$$
$$\phi(i) = \frac{\int_S F(s,i)\phi(s)^{\beta}\, ds}{\int_S F(s,i)\phi(s)^{\beta-1}\, ds}, \tag{32}$$

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 (32) as

$$\frac{\phi(i)^{\beta}}{\int_S F(s,i)\phi(s)^{\beta}\,ds} = \frac{\phi(i)^{\beta-1}}{\int_S F(s,i)\phi(s)^{\beta-1}\,ds}. \tag{33}$$

Define $\gamma(i) \equiv \frac{\phi(i)^{\beta}}{\int_S F(s,i)\phi(s)^{\beta}ds}$. 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 (33) as the following set of equations:

$$g_1(i) = \int_S \lambda(i)F(s,i)g_1(s)ds \tag{34}$$

$$g_2(i) = \int_S \lambda(i)F(s,i)g_2(s)ds \tag{35}$$

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 (34) and (35), 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 (13). Hence equation (13) is the unique relationship between $w(i)$ and $L(i)$ that satisfies equations (11) and (12) for a regular spatial equilibrium.

Since equation (13) holds for any regular equilibrium, it is sufficient to consider it along with equation (14) to determine existence and uniqueness of a regular equilibrium rather than equations (11) and (12) directly. Note that we can rewrite equation (14) as a nonlinear integral equation

$$f(i) = \lambda \int_S K(s,i)\,f(s)^{\frac{\gamma_2}{\gamma_1}}\,ds, \tag{36}$$

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 (36) 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

$$\widetilde{f}(i) = \int_S K(s,i)\left[\tilde{f}(s)\right]^{\frac{\gamma_2}{\gamma_1}} ds \iff \tag{37}$$
$$f(i)\,\lambda^{\frac{1}{\frac{\gamma_2}{\gamma_1}-1}} = \int_S K(s,i)\left[f(s)\,\lambda^{\frac{1}{\frac{\gamma_2}{\gamma_1}-1}}\right]^{\frac{\gamma_2}{\gamma_1}} ds,$$

which is equivalent to (36). In our case $f(i) \equiv L(i)^{\gamma_1}$ and the labor market clearing constraint implies

$$\bar{L} = \lambda^{-\frac{1}{\gamma_2-\gamma_1}} \int_S \widetilde{f}(s)^{1/\gamma_1} ds, \tag{38}$$

and thus, for each solution for $\widetilde{f}(s)$, a unique solution for $\lambda$.[31] Therefore, finding a solution for $\widetilde{f}(i)$ gives us the solution for $f(i)$ and the eigenvalue of the system $\lambda$, 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 (36). 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 $\phi$ 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

[31] 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 (37), 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.

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.[32]

**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 notice that if $\gamma_1 > 0$ expression (7) 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 $\lambda$, 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 (38), 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

[32] 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.

$\widetilde{f}(s)\, s^c$ is increasing for $c \in (0,1)$ then there there exists a unique positive solution to equation (37). Furthermore, that solution is the uniform limit of the successive approximations:

$$\widetilde{f}_{n+1}(i) = \int_S K(s,i)\, \widetilde{f}_n(s)^{\frac{\gamma_2}{\gamma_1}}\, ds. \tag{39}$$

for any arbitrary non-zero, non-negative $\widetilde{f}_0(i)$.[33] 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 (37) and uniqueness is proved when $\left|\frac{\gamma_2}{\gamma_1}\right| < 1$

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 (36) 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 $\mathbb{R}^N$.

**Discrete number of locations** The uniqueness proof for Theorem 2 also applies to the case of a discrete number of locations for $\gamma_2/\gamma_1 \in (0,1]$. 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.

### A.1.3 Proof of Proposition 1

Consider a regular equilibrium satisfying equations (13) and (14). Taking the derivative of welfare in location $i$ with respect to the population in location $i$ from equation (7) yields:

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

[33]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.

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:

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

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.1.4 Proof of Theorem 3

Substituting equation (25) into equation (24) yields:

$$u\left(i\right)^{1-\sigma} = \frac{W^{1-\sigma}}{\phi}\int_S T\left(s,i\right)^{1-\sigma} w\left(i\right)^{\sigma-1} w\left(s\right)^{\sigma} L\left(s\right) u\left(s\right)^{\sigma-1} ds. \tag{40}$$

Define the functions $f\left(i\right) \equiv u\left(i\right)^{1-\sigma}$ and $K\left(s,i\right) \equiv \frac{W^{1-\sigma}}{\phi} T\left(s,i\right)^{1-\sigma} w\left(i\right)^{\sigma-1} w\left(s\right)^{\sigma} L\left(s\right)$. Then equation (40) can be rewritten as:

$$f\left(i\right) = \int_S K\left(s,i\right)\frac{1}{f\left(s\right)} ds. \tag{41}$$

Equation (41) arises in models of signal theory and was first studied by Nowosad (1966) in the case that $K\left(s,i\right)$ is symmetric and $S = [0,1]$. Since this equation is essentially the same as equation (37) 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 $K\left(i,i\right) > 0$ for all $i \in S$. As a result, we can apply Theorem 2 and Remark 1 of Karlin and Nirenberg (1967) to equation (41), which imply that there exists a unique continuous positive function $u\left(i\right)$ satisfying equation (40) for any connected compact subspace of $\mathbb{R}^N$. The continuous and positive function $A$ can then be determined using equation (41). Note that $A$ and $u$ are only identified up to scale, as the constants $\phi$ and $W$ offset any changes in the normalizations of $A$ and $u$, respectively.

## A.2 Isomorphism of the perfect competition and the monopolistic competition environments

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 externalities 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.

**Gravity models with two sectors** A formal isomorphism can be derived with the Helpman (1998)-Redding (2012) setup. That setup assumes that workers spend a constant share $\gamma$ 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 $(1-\gamma)/\gamma$ 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 $\beta = -(1-\gamma)/\gamma$. For the equilibrium to be unique we require that

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

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.

**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 $\omega$ 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 $\theta$, 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.