NBER WORKING PAPER SERIES

SUPPLY AND DEMAND IN SPACE

Treb Allen
Costas Arkolakis

Working Paper 30598
http://www.nber.org/papers/w30598

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022

We thank Steve Redding and our editors Erik Hurst, Nina Pavcnik, Tim Taylor, and Heidi Williams for exceptionally helpful feedback and Saptarshi Majumdar for excellent research assistance. The authors acknowledge support by the National Science Foundation under grants SES-1658838 and SES-1658875. 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.

Supply and Demand in Space
Treb Allen and Costas Arkolakis
NBER Working Paper No. 30598
October 2022
JEL No. F1,R0

ABSTRACT

What do recent advances in economic geography teach us about the spatial distribution of economic activity? We show that the equilibrium distribution of economic activity can be determined simply by the intersection of labor supply and demand curves. We discuss how to estimate these curves and highlight the importance of global geography – i.e. the connections between locations through the trading network – in determining how various policy relevant changes to geography shape the spatial economy.

Treb Allen
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
treb@dartmouth.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

Supply and Demand in Space*

Treb Allen
Dartmouth and NBER

Costas Arkolakis
Yale and NBER

October 2022

Abstract

What do recent advances in economic geography teach us about the spatial distribution of economic activity? We show that the equilibrium distribution of economic activity can be determined simply by the intersection of labor supply and demand curves. We discuss how to estimate these curves and highlight the importance of global geography – i.e. the connections between locations through the trading network – in determining how various policy relevant changes to geography shape the spatial economy.

1 Introduction

The spatial distribution of people is incredibly concentrated: 8% of the U.S. population lives in the ten largest cities in the U.S., but those cities take up less than 0.1% of the total land area. Why this concentration? More generally, what determines the distribution of people and economic activity across space?

We show that the equilibrium spatial distribution of population and economic activity can be understood through the familiar lens of supply and demand curves. We begin by applying this intuition to the famous Rosen-Roback (Rosen, 1979; Roback, 1982) framework. We then extend this same intuition to modern economic geography frameworks where locations are connected through the flow of goods based on our earlier work in Allen and Arkolakis (2014). To keep the discussion as straightforward as possible, we relegate all mathematical details

---

*Allen: treb.allen@dartmouth.edu. Arkolakis: costas.arkolakis@yale.edu. We thank Steve Redding and our editors Erik Hurst, Nina Pavcnik, Tim Taylor, and Heidi Williams for exceptionally helpful feedback and Saptarshi Majumdar for excellent research assistance. The authors acknowledge support by the National Science Foundation under grants SES-1658838 and SES-1658875. All errors are our own.

and derivations to the Appendix. We also provide a companion Matlab toolkit to help researchers apply these techniques on their own.

Spatial linkages between locations create rich interactions across space that have important implications for the predictions of the model. Now the economic fate of a location depends not only on its own "local" geography but also on the local geography of its neighbors, the impact of which is relegated by the strength of the economic ties, i.e. a "global geography". Despite this added complexity, we show the same tools based on supply and demand used to understand predictions of the Rosen-Roback framework extend readily to a globally integrated world.

The benefit of such a globally integrated framework is that it can be applied to understand both the direct and indirect impacts of real world economic policies that change either the local or global geography. To assist in such endeavors, we discuss how the framework can be applied to spatial data, highlight the most common pitfalls that can arise and offer strategies for traversing them. Finally, we provide a brief overview of the many ways in which this framework has been applied thus to better understand the spatial distribution of economic activity, as well as highlighting several interesting and as-of-yet unexplored questions for future researchers.

2 Understanding the spatial distribution of economic activity through the lens of supply and demand

Consider an economy (say, the United States) comprising many different locations, which is part of a larger world. These locations each have their own "local" geography. The "local" geography of a location includes a whole host of things, from natural geographic features like the climate, elevation, natural beauty, etc. to other less tangible characteristics of a location like the quality of its political institutions. Local geography can affect the spatial distribution of economic activity in two ways. First, it can affect the desire of people to live in a location and hence labor supply; we will call such factors "amenities." Second, it can affect how productive people are in a location and hence labor demand; we call such factors "productivities."

Labor supply and demand

Suppose there are a lot of people living in the U.S., each of whom gets to choose where they live. Wherever they choose to live, they earn a wage from producing a good and then use that wage to buy things. Let us assume that the wage they earn in any location i depends on two things: (1) the number of people living in that location; and (2) the productivities of that location according to the following labor demand curve:

$$\ln w_i = \varepsilon^D \ln L_i + \ln C_i^D, \tag{1}$$

where ε^D is the demand elasticity and C_i^D is the local productivity in region i that arises from its local geography; see Appendix A.1 for a particular micro-foundation that delivers equation (1).

Which way does the demand curve slope? Common assumptions imply that the demand elasticity is negative and the demand function downward sloping. Assuming decreasing returns to scale in production of the good (or simply the presence of a fixed factor such as capital), is perhaps the most common of all.[1] But the presence of external economies also can affect the slope of the demand function. If more workers in a location result in everyone being more productive, the demand curve can become more elastic and if these external economies are sufficiently strong the demand curve may even slope upwards. This can lead to things like multiple equilibria or "black hole" equilibria where everyone lives in one location.[2] While academically interesting, in what follows we will stick with the more common (and, arguably, empirically relevant) case of a downward sloping demand curve.

Because people get to choose where they live, each chooses their residence to be as happy as possible. What makes people happy in this framework? Two things: higher consumption (so, all else equal, workers prefer higher real wages) and living somewhere nice (i.e. a place with high amenities). If everyone is identical, this means that all inhabited locations must make people equally as happy. If prices are the same everywhere (so that the real wage is the nominal wage) and the amenity value of a location depends in part on how many other people live there, then workers' indifference across all inhabited locations generates the following labor supply curve:

$$\ln w_i = \varepsilon^S \ln L_i - \ln C_i^S, \tag{2}$$

where ε^S is the supply elasticity and C_i^S is the local amenity in region i; see Appendix A.1 for a particular micro-foundation that delivers equation (2).

Which way does the supply curve slope? We usually think of a supply curve sloping upward and it will here too as long as more people in a location make each individual less happy. The presence of a housing market (where a higher population drives up rent) or

---

[1] See, for example, Kline and Moretti (2014) and Donaldson and Hornbeck (2016).

[2] The possibility of multiple spatial equilibria is a fascinating and ongoing branch of the economic geography literature, see e.g. Krugman (1991), Matsuyama (1991), Fujita, Krugman, and Venables (1999), Davis and Weinstein (2002), Bleakley and Lin (2012), and Allen and Donaldson (2020).

idiosyncratic preferences (where a higher population means the marginal resident's match quality is worse) can also lead to upward sloping labor supply curves.[3] But it is possible for the labor supply curve to slope downward (and issues of multiplicity and black holes to arise) if the amenity value of a location is increasing in its population, e.g. through greater investments in public goods or greater variety in consumables.

The "local" spatial equilibrium

To determine the equilibrium spatial distribution of economic activity – i.e. the population and wage in every location – we simply combine the demand and supply curves in equations (1) and (2) and find the intersection. The spatial equilibrium is highlighted at point A in Figure 1.

To see how the "local" geography shapes the spatial equilibrium, consider a simple counterfactual scenario where the amenity value of residing in a location improved. For example, suppose the advent of air conditioning technology made the hot climate of the U.S. Southwest less oppressive. An improvement in amenities shifts outward the labor supply curve, moving the equilibrium from point A to point B in Figure 1. The population in the location increases, but its wage declines: the U.S. Southwest is now a better place to live, but the influx of workers depresses the wages.

The fact that we can analyze each location separately, depending on the amenity shock they receive, illustrates the somewhat paradoxical nature of the Rosen-Roback framework. It is a spatial model, but the distribution of economic activity depends only on local geography, not on what happens to other regions. Intuitive spatial features like where a location is located on a map and who its neighbors are entirely absent: it is a spatial model where space does not matter.[4] To make space matter, we need to introduce a modern economic geography model with spatial linkages. This will create the concept of "global" geography which we introduce and analyze next.

---

[3] For a discussion of heterogeneous preferences and housing market see Helpman (1998); Allen and Arkolakis (2014); Redding (2016); Ahlfeldt, Redding, Sturm, and Wolf (2015).

[4] More generally, in the Rosen-Roback framework, a change in the local geography in one location can have aggregate general equilibrium effects on e.g. the price of capital or through the aggregate labor market clearing condition. But such general equilibrium effects affect all locations equally and hence do not affect the spatial distribution of economic activity.

3 The role of global geography in the spatial distribution of economic activity

There are many ways that different locations are linked with each other: people may live in one location and work in another, people may migrate from one location, people may talk with each other leading to the spatial diffusion of ideas, etc. But perhaps the most obvious spatial linkage is through the flow of goods. Much of what an individual consumes is produced elsewhere: for example, according to the 2017 Commodity Flow Survey, intra-state trade flows in the average state was only 22% of total intra-national trade. Moreover, the pattern of trade flows are far from uniform. As panel (a) of Figure 2 highlights using the same data, nearby states trade more with each other while the total volume of trade increases with the size of the trading partners, a phenomenon originally observed in international trade flows and oftentimes referred to as "gravity" (Anderson, 2011; Head and Mayer, 2013).

How does incorporating such spatial linkages affect the spatial equilibrium? It turns out that much of the basic intuition above remains: in particular, we can still analyze the spatial equilibrium using the familiar techniques of supply and demand, albeit now augmented with a concept of both "local" and "global" geographies.

The global geography

The model discussed below is based on prior work (Allen and Arkolakis, 2014), but variations of this spatial framework with equivalent or similar mathematical formulations have recently been used in a variety of frameworks.[5] We relegate all details of the mathematical derivations to Appendix A.2. The setup retains the same features as above but now we introduce a key distinction: goods are no longer costlessly traded. Instead, there are trade relationships between different locations, governed by the presence of spatial frictions. We define T_{ij} to be the inverse of these connections, the *inverse economic distance* between regions i and j. As we will discuss in Section 4, an appealing feature of this framework is that the inverse economic distance can be measured explicitly by projecting observed bilateral trade flows on observed bilateral geographic characteristics such as distance or time of travel.

When goods are no longer costlessly traded, two things change: first, the prices of the goods produced by workers in a location depends in part on how nearby the consumers of those products are. The closer the consumers are, the more demand for their products, and the higher the price (and hence the higher the wage) that the workers can obtain. This

---

[5] See for example Redding (2016); Donaldson and Hornbeck (2016); Allen, Arkolakis, and Takahashi (2020); Faber and Gaubert (2019). Redding and Rossi-Hansberg (2017) offer a comprehensive review of the quantitative spatial framework.

outward market access affects the labor demand curve of a location. Second, the price of goods purchased by consumers in a location depend in part on how nearby the producers of those products are. The closer the producers, the lower the price for those products, and the higher the real wage of the consumers. This *inward market access* acts as a shifter to the labor supply curve of a location.[6]

Together, the outward and inward market accesses comprise the *global geography* of a location. Following Anderson and Van Wincoop (2003); Redding and Venables (2004), the outward market access (MA_i^{out}) can mathematically be expressed as:

$$MA_i^{out} = \sum_j T_{ij} \times \frac{Y_j}{MA_j^{in}}, \tag{3}$$

where $Y_j = w_j L_j$ is the total income of location j. Intuitively, outward market access summarizes the selling potential of a market, which is greater for location i when its neighboring locations (i.e. those with high T_{ij}) are richer (i.e. have higher Y_j) or have worse alternatives for buying their own goods (i.e. have lower MA_j^{in}).

Inward market access, in turn, is similarly defined as the capacity of locations to buy from other locations that have high income to outward market access ratio, weighted again by T_{ij}:

$$MA_j^{in} = \sum_i T_{ij} \times \frac{Y_i}{MA_i^{out}} \tag{4}$$

Intuitively, inward market access is higher when a location j's neighbors (i.e. those with a higher T_{ij}) either produce a lot (i.e. have higher Y_i) or have poor alternatives for selling their goods (i.e. have a lower MA_i^{out}). Together, the global geography summarizes how each location depends on economic activity in all other locations, where closer locations are given greater weights.

Equations (3) and (4) highlight that inward and outward market accesses are intertwined, with each dependent in part on the other. Despite this feedback loop between the two, given the total income of each location and the inverse economic distance between any pair of locations, equations (3) and (4) can be jointly solved to determine the unique (to-scale) global geography of the system; the companion Matlab code provides a convenient algorithm for doing so. Panel (b) of Figure 2 depicts the (outward) market access for each U.S. states, where we proxy the inverse economic distance T_{ij} with inverse great-circle distance. States with high economic output that are close to other states with high output such as those in

---

[6] The literature sometimes refers to inward market access as "consumer" market access and outward market access as "firm" market access, see e.g. Redding and Sturm (2008) and Donaldson and Hornbeck (2016).

the Northeast have good market access; states with less economic output that are far away from states with higher economic output such as Montana have poor market access.

The global spatial equilibrium

In a world with spatial linkages, it turns out the global spatial equilibrium can be analyzed using labor supply and demand curves just as in the local spatial equilibrium above. Now, however, supply and demand will not only depend on local geography but also on global geography. In particular, the labor demand that was previously represented by equation (1) now also depends on outward market access MA_i^{out}, becoming:

$$\ln w_i = \varepsilon_{local}^D \ln L_i + \varepsilon_{global}^D \ln MA_i^{out} + \ln C_i^D. \tag{5}$$

Better outward market access acts analogously to better local productivities, C_i^D, shifting the demand curve for local labor outwards with an elasticity $\varepsilon_{global}^D \geq 0$. That elasticity is greater the less substitutable the goods produced in i are with goods produced elsewhere in the world.

Similarly, labor supply previously represented by equation (2) now depends on inward market access MA_i^{in}, becoming:

$$\ln w_i = \varepsilon_{local}^S \ln L_i + \varepsilon_{global}^S \ln MA_i^{in} - \ln C_i^S. \tag{6}$$

Better inward market access acts analogously to better local amenities C_i^S, shifting the supply curve for labor outwards with an elasticity $\varepsilon_{global}^S \leq 0$, which again is larger in magnitude the less substitutable goods produced in different locations are with each other. The two limiting cases deserve special mention. When $\varepsilon_{local}^S \to \infty$, the local population is invariant to changes in economic conditions, whereas when $\varepsilon_{local}^S \to 0$ the labor supply is infinitely elastic to local economic conditions. These special cases correspond to important cases in the literature, as we will discuss below.

Given the global geography, the global spatial equilibrium is determined just as in the local spatial equilibrium above: you simply find the wage and population in each location that equates supply with demand; point A on panel (a) of Figure 3 depicts such an equilibrium.

So what has changed in the global spatial equilibrium? The crucial insight is that *the global geography in one location depends on the spatial equilibria in all other locations*. If something changes about the local geography *anywhere* in the world, it will affect the global geography *everywhere* in the world (and it will affect nearby locations more than locations far away). Hence, the global geography puts space back into the spatial economy.

To illustrate this global spatial equilibrium, let us return to the example above. Suppose that air conditioning is invented, which makes some hot and previously inhospitable location i much more hospitable, raising the amenity of living there. Like above, this will shift outward labor supply curve in location i to point B in panel (a) of Figure 3, increasing the population in location i and reducing the wages. The story does not end here, as this change in population and wages will affect the global geography. As long as $\varepsilon^D_{local} > -1$, the income Y_i of location i will increase, raising both the inward and outward market access and resulting in an additional shift outward to both the labor demand and labor supply curves. This additional global effect further increases the population in location i and mitigates the downward fall in wages, as illustrated in point C in panel (a) of Figure 3.

At the same time, changes in the economic activity in location i affect the global geography of other locations. Consider a neighboring location j initially in equilibrium, as illustrated by point A in panel (b) of Figure 3. Because the income of location its neighbor i has improved, both its supply and demand curves will shift outwards as well. Intuitively, the greater nearby economic activity both increases the demand for the goods produced in j and increases the supply of goods consumed in j. As a result, the population in j increases too (and its wages rise), changing its equilibrium to point C in panel (b) of Figure 3, despite there being no change in its own local geography.[7]

But won't changes in the economic activity in location j have subsequent impacts on the global geography in all other locations? And won't those changes have even further impacts on the global geography, *ad infinitum*? Yes and yes: it is this infinite feedback loop between the global geography in every location that makes the global spatial equilibrium so interesting to study. In reality, point C in panels (a) and (b) of Figure 3 represents the limit of the infinite sequence of these adjustments of each locations' global geography to adjustments made in the global geography everywhere else (i.e. the "fixed point"). Indeed, this iterative process is what both the algorithm for calculating the equilibrium change in market accesses in the companion Matlab code and many tools for studying the mathematical properties of the equilibrium system is based upon.

Of course this is not the first time that infinite feedback loops in networks have been studied. Indeed, in the special case where the supply curve is infinitely elastic (i.e $\varepsilon^S \to 0$), the local and global demand elasticities are equal in magnitude (i.e. $-\varepsilon^D_{local} = \varepsilon^D_{global}$) and the inverse economic distances are symmetric (i.e. $T_{ij} = T_{ji}$), the equilibrium global economy is one in which the wages and populations of each location are (log) proportional to the

---

[7] Whether nominal wages rise or fall (i.e. whether outward or inward market access increases more) depends on the choice of the numeraire. Here we set mean wages equal to one as the numeraire, so falling wages in location i must be offset by rising wages elsewhere.

eigenvector centrality of a location in the network defined by the world geography (i.e. the economic distances, productivities, and amenities). Higher eigenvector centrality means that a node in a network is nearby to other nodes with high eigenvector centralities. Eigenvector centrality is perhaps most famously used by the Google PageRank algorithm to rank websites in searches, where websites linked to by other influential websites appear at the top of the search results. Here, locations are more populated (and wealthier) the closer they are to other more populated (and wealthy) locations. Moreover, the eigenvalue of the system corresponding to this eigenvector turns out to be the welfare of the global economy.[8]

Having shown how one can determine the global spatial equilibrium through the use of supply and demand curves, we now turn to describing the process through which this framework can be combined with spatial data to assess the impact of changes in geography on the real world spatial distribution of economic activity.

4 Estimating labor supply and demand

In the previous section, we saw how a simple supply and demand framework can be used to understand how changes in the geography affects the distribution of economic activity across spatially connected locations. But one of the most attractive aspects of the global spatial framework described above is its ability to seamlessly integrate with readily available spatial data. In this section, we describe this interplay between theory and data.

Spatial economic data

In what follows, we describe two types of spatial data: (a) data on the local economic activity of a location; and (b) data on the strength of economics linkages between locations across space.

Local economic data

Suppose that you observe both how many people reside in the location (i.e. L_i) and the total income of a location (i.e. Y_i). These data are indeed readily available; for example, in the United States, population data and income data at the county level can be constructed from the decennial census going back to the year 1840. (IPUMS' National Historical Geographic Information Systems (Manson, 2020) has provided an enormous public good in assembling

---

[8] In the more general case, the equilibrium of the spatial economy constitutes a network system of nonlinear equations. The properties of such systems remains an active field of research, see e.g. Allen, Arkolakis, and Li (2020).

these data and making them publicly available). Even in parts of the globe where spatially disaggregated income data is not readily available, one can proxy for economic activity using satellite data on the intensity of lights at nighttime, a practice pioneered by Henderson, Storeygard, and Weil (2012) and summarized in Donaldson and Storeygard (2016). Furthermore databases assembling data from various sources provide disaggregated information on economic activity at a granular geographic level, such as the G-econ database (Nordhaus and Chen, 2006) that provides proxies of income and population at the 1-arc degree.

We furthermore assume that all income accrues to labor, i.e. $Y_i = w_i L_i$, allowing us to recover wages given knowledge of income and population. While consistent with the theory above, this is a strong assumption that clearly abstracts from all other sources of income (e.g. capital, landholdings, firm profits, etc.). One could argue that all these sources of income eventually accrue to individuals; indeed, as long as the income remains in a particular location, the predictions of the global spatial framework does not change by incorporating these other sources of income. (For example, as long as individuals in a location own their own homes, a model where individuals spend money on housing is no different – we say it is "isomorphic" – to the framework described above). But in reality, not all income earned in a location accrues to the labor in that location, and such spatial flows of income would present another linkage between locations that we abstract from here.

Data on economic linkages

Now consider the second type of data: data on the economic linkages across space. These data correspond to the "inverse economic distance" T_{ij} from above. In the theory, the inverse economic distance is proportional to the value of trade flows between two locations (conditional on origin and destination fixed effects), so any observable that affects the value of bilateral trade flows can be part of the set of measures of economic linkages. As Figure (2) makes clear, one such observable is simply the geographic distance between any two locations. Indeed, one of the most robust empirical relationships in all of economics is that trade flows between locations are roughly inversely proportional to the geographic distance between them (see e.g. Disdier and Head (2008) and Chaney (2018)). Put another way, a very good start to measuring "inverse economic distance" T_{ij} is simply with "inverse geographic distance."

More recently, researchers have begun to improve upon the distance proxy with measures of actual travel costs between locations. For example, Donaldson (2018) estimates the relative cost of traveling between locations via road, rail, and waterways by calculating the lowest cost route using Dijkstra (1959)'s algorithm (the same algorithm used e.g. by Google Maps). Allen and Arkolakis (2014) use a continuous space extension of the Dijkstra algorithm known as the Fast Marching Method (see Tsitsiklis (1995); Sethian (1999)) to calculate travel times along the optimal route between locations. Even more recently, Allen and Arkolakis (forthcoming) offers an analytical solution for the inverse economic distance as a function of the underlying transportation network. The advantages of these related approaches relative to using geographic distance directly is twofold: first, they offer more realistic measures of the actual strength of economic spatial linkages; second, they allow researchers to assess how changes in transportation infrastructure that shortens the cost or distance of travel affects the spatial distribution of economic activity.

For any observed measure(s) of the economic linkages, the inverse economic distance $\{T_{ij}\}$ can then be constructed by regressing the observed (log) value of trade flows on those measures, conditioning on the origin and destination fixed effects. The predicted values of that regression (excluding the estimated fixed effects) are the implied inverse economic distance.[9] For example, if one uses travel times as a measure of economic linkages, the inverse economic distance would be the product of the travel time and its estimated coefficient from the gravity regression.

Estimating supply and demand

Given measures of the income in each location (i.e. Y_i) and a measure of the strength of the linkages between locations (i.e. T_{ij}), we can calculate the global geography of every location – i.e. the inward and outward market accesses MA_j^{in} and MA_i^{out} – using equations (3) and (4).[10] We provide a simple iterative algorithm for solving that nonlinear system of equations in the companion Matlab code.

Now let us return to our supply and demand equations (5) and (6). We observe the left hand side price variable (i.e. the wage w_i) and the right hand side quantity variable (i.e. the population L_i) and market access variables (MA_i^{in} and MA_i^{out}). We would like to estimate the coefficients on the right hand side variables (i.e. the model elasticities ε_{local}^S, ε_{global}^S, ε_{local}^D, and ε_{global}^D) as well as the productivity and amenity residuals (i.e. $\ln C_i^D$ and $\ln C_i^S$). Or put another way, we would like to estimate a system of supply and demand where we observe the price and quantity – a problem that is very well understood!

---

[9] An alternative procedure would be to calibrate the inverse economic distance to exactly match the observed bilateral trade flows by including the regression residual in its construction. Such a procedure – which is closely related to the "exact hat algebra" pioneered by Dekle, Eaton, and Kortum (2008) and discussed in Costinot and Rodríguez-Clare (2014) – can result in an over-fitting problem when conducting counterfactuals; see Dingel and Tintelnot (2020).

[10] Recovering the global geography from the observed income and economic distances is a well behaved problem, as one can show using tools from Allen, Arkolakis, and Li (2020) that there exists a unique (to-scale) inward and outward market accesses MA_j^{in} and MA_i^{out} that solve equations (3) and (4) for any set of incomes Y_i and inverse economic distances T_{ij}.

How do we go about estimating our supply and demand curves? It might perhaps be more informative to start with what not to do. Following in the footsteps of Baldwin and Taglioni (2006), let us award medals for different types of errors that can arise, ranking them from most to least obvious.

The bronze medal error

One glaring mistake one could make in estimating supply and demand equations (5) and (6) – our "bronze medal" error – would be to use ordinary least squares (OLS). OLS is clearly not appropriate due to familiar simultaneity issues: because the right hand side population variable is determined in equilibrium from equating supply and demand, it will be correlated with both the productivity and amenity shifters. As a result, the OLS coefficient will not recover either the supply or demand elasticity.

One strategy for overcoming this bronze medal error would be to employ an instrumental variable strategy, using variation in the amenity $\ln C_i^S$ as an instrument for the equilibrium population to estimate the demand elasticity in (5) and using variation in the productivity $\ln C_i^D$ as an instrument for the equilibrium population to estimate the supply elasticity in (6). As long as the chosen variation in the amenities and productivities are uncorrelated, this will yield consistent estimates of the demand and supply elasticities.

What are examples of such instruments? One example comes from Glaeser and Gottlieb (2009) who argue that the advent of air conditioning improved the amenity of locations with warm climates. Under the assumption that the climate of a location is not also correlated with the change in the productivity of a location, the climate of a location can be used as an instrument for change in population to identify the demand elasticity ε_{local}^D ; see Allen and Donaldson (2020).

Conversely, Allen and Donaldson (2020), following Bustos, Caprettini, and Ponticelli (2016), argue that increased global demand for soy improved the productivity of locations particularly well suited for the production of soy. Under the assumption that the potential yield of soy in a location (say, relative to its potential yield for corn) did not also change the amenity of a location, the potential relative yield of soy to corn can be used as an instrument to identify the supply elasticity ε_{local}^S. Of course, the climate or agroclimatic properties are likely correlated with myriad characteristics of a location, making it unlikely these assumptions hold comparing wages and populations across locations in the cross section. As such, it is preferable to instead rely on panel variation, looking at changes in wages and populations across locations over time (or, equivalently, including location fixed effects in the estimation of the supply and demand equations).

The silver medal error

Somewhat less obviously – and our "silver medal" error – would be to ignore the spatial linkages between locations and simply estimate supply and demand using the equations (1) and (2). Doing so ignores the variation in inward and outward market access across locations, relegating that variation to the residual.[11] Notably, the instrumental variable strategy meant to address simultaneity bias from above is insufficient to address this bias. To see this, suppose you are estimating the demand equation (5), instrumenting for population with an amenity shifter. Even if that amenity shifter is uncorrelated with productivities, it will be correlated with the outward market access, biasing the estimate of the demand elasticity. Indeed, the only situation where this bias does not arise is in the special case when all locations share the same market access (as in the local spatial equilibrium).[12]

Fortunately, avoiding this mistake is straightforward: from the discussion above, one can construct measures of inward and outward market access measures from readily available spatial economic data. Including these market access measures in the supply and demand equations is a simple remedy to avoid the silver medal error. But just controlling for them alone is not sufficient (and is our final "gold" medal error).

The gold medal error

Most subtly, the market access measures are themselves correlated with the productivity and amenity of a location. This is because the market access of a location depends in part on its own economic activity, which of course depends in equilibrium on its productivity and amenity. As a result, just including the market access measures in the supply and demand equations as controls will result not only in biased estimates of the global elasticities ε^S_{global} and ε^D_{global} but also biased estimates of the local elasticities ε^S_{local} and ε^D_{local}.

To address this concern, one can again use an instrumental variables strategy, instrumenting for both the population in a location and its market access. We discussed above possible instruments for the population; what about for market access? An appropriate instrument would be correlated with market access but uncorrelated with local productivities or amenities.

---

[11] Equivalently, if you do not condition on inward and outward market accesses, the labor demand and supply elasticity will vary across locations, as e.g. emphasized by Monte, Redding, and Rossi-Hansberg (2018).

[12] Our "silver medal" error is similar in spirit to Baldwin and Taglioni (2006)'s "gold medal" error of failing to control for variation in market access in gravity equations. The two errors are distinct because unlike a gravity regression, the supply and demand regressions are not estimated using bilateral flows. As a result, their proposed solution of controlling for market access with origin and destination fixed effects does not apply here.

From equations (3) and (4), market access is a type of inverse economic distance weighted average of economic activity near a location. One possibility would be to construct an instrument based on equations (3) and (4) but excluding the own location (and perhaps also nearby locations) from the sum. But even if there is no spatial correlation in the productivity and amenity of locations, the equilibrium economic activity elsewhere depend in part on the economic activity of the own location (and hence the own productivity and amenity shifters), so such an instrument is unlikely to satisfy the exclusion restrictions.

An alternative approach has much more promise. Suppose you use observed measures of productivities and amenities along with plausible values of the model elasticities to calculate the local equilibrium of a (hypothetical) economy using equations (1) and (2). This hypothetical economy is one in which spatial linkages do not matter and the only heterogeneity in productivities and amenities across locations arise from observables. Next, combine the implied equilibrium income in each location from this hypothetical economy with the observed economic distance and use (3) and (4) to calculate what the market access would be in such a hypothetical economy. If this hypothetical market access is correlated with the actual market access (which is something you can verify), it is a valid instrument as long as you also condition on a location's own observed measures of productivities and amenities. Intuitively, the impact of market access on the supply and demand curves is being identified only on variation in productivities and amenities elsewhere (through the spatial structure of the model). Examples of such "model implied" instruments can be found in Monte, Redding, and Rossi-Hansberg (2018), Allen, Arkolakis, and Takahashi (2020), and Adao, Arkolakis, and Esposito (2019).

Taking stock

Suppose you have successfully avoided the bronze, silver, and gold medal errors above by estimating the labor supply and demand curves while appropriately instrumenting the observed population and the market access terms. Now what?

Armed with estimates of the model elasticities along with data on wages, populations, and market access terms, the residuals of the supply and demand equations (5) and (6) correspond to the productivities and amenities in each location.[13] Put another way, if you know the supply and demand elasticities, you can always find the local geography such that the observed distribution of economic activity – combined with the inverse economic distances you have constructed – is the global spatial equilibrium of the model.

---

[13] This approach of recovering the underlying geography based on the supply and demand residuals is equivalent (but perhaps easier to digest) to an approach that directly inverts the equilibrium market clearing conditions, as e.g. in Allen and Arkolakis (2014) and Redding (2016).

Because you have recovered the geography that is consistent with the observed economic activity and you know the model elasticities, you are now able to assess how any change to the geography will change the global spatial equilibrium using the techniques described in Section 3. In the next section, we will discuss the many different ways this can inform spatial policy.

5 Understanding the spatial impact of economic policies

We have seen how the global and local geographies interact through supply and demand to shape the spatial equilibrium and how those supply and demand curves can be combined with spatial data to apply the framework to the real world. Now we are equipped to describe the many types of questions that can be addressed with such a framework. We classify these questions into three types: those examining the impact of changes to the local geography, those examining the impact of changes to the global geography, and those which extend the framework above to incorporate additional spatial linkages beyond the flow of goods.

Local Geography Shocks

Consider first the question of how changes to local geography – i.e. changes to amenities $\ln C_i^S$ which shift the supply curve or changes to productivities $\ln C_i^D$ which shift the demand curve – affect the spatial distribution of economic activity. Perhaps the most notable example of a change in amenities is the literature analyzing the implications of various housing policies (see e.g. Diamond and McQuade (2019)). Of particular interest are the implications of amenities for spatial sorting as in the work of Diamond (2016); Almagro and Dominguez-Iino (2019), summarized in the review by Diamond and Gaubert (2022). Another prominent example of amenity changes are the effects of climate change or natural disasters on the livability of a location, such as the impact of spatial flooding by Balboni (2019); Desmet, Kopp, Kulp, Nagy, Oppenheimer, Rossi-Hansberg, and Strauss (2021), hurricanes (Henkel, Kwon, and Magontier, 2022), and wildfires (Ospital, 2022). The long-run effects of conflict and war on the spatial distribution of economic activity is a third example of amenity shocks to a location, see e.g. Davis and Weinstein (2002); Redding and Sturm (2016); Chiovelli, Michalopoulos, and Papaioannou (2018).

There are also many examples of economic questions which can be viewed as shifts in the local geography that shift the local demand curve. Sectoral specialization due to productivity effects is the focus of the work of Conte, Desmet, Nagy, and Rossi-Hansberg (2021); Cruz Alvarez and Rossi-Hansberg (2021); Caliendo, Parro, Rossi-Hansberg, and Sarte (2018). Bustos, Caprettini, and Ponticelli (2016) presents evidence that the introduction of genetically modified soy beans had heterogeneous effects on agricultural productivity across areas with different soil and weather characteristics. An interesting avenue for future research is the analysis of the spatial effects of automation or new technologies that allow for remote work, following evidence presented by Dingel and Neiman (2020); Acemoglu and Restrepo (2020); Althoff, Eckert, Ganapati, and Walsh (2022). Place based policies enacted by the government can also be viewed as examples of shifts to the local demand or supply curves (depending on the particular nature of the policy), and there has been recent work (see Fajgelbaum and Gaubert (2018); Gaubert, Kline, and Yagan (2021)) examining how best to design such policies.

Global Geography Shocks

Now let us turn our attention to how changes to global geography – i.e. changes in the inverse economic distances, T_{ij}, and the resulting changes in the market access – affect the spatial distribution of economic activity.

A natural application for evaluating changes in global geography is the analysis of the effects of investment in transportation infrastructure. Examples of this work include Allen and Arkolakis (2014); Donaldson and Hornbeck (2016); Tsivanidis (2018); Severen (2019); Jaworski and Kitchens (2019); Heblich, Redding, and Sturm (2020), who study the spatial impacts of investments across a wide variety of various modes of transportation, including highways, trains, subways, and buses. One fertile topic of research is the redistributional impacts of such investments. For example, Lee (2022) presents evidence that transportation infrastructure investments may lead to racial disparities providing evidence for the tipping mechanism presented by Card, Mas, and Rothstein (2008), whereas Baum-Snow, Henderson, Turner, Zhang, and Brandt (2020) presents evidence that investment in transportation infrastructure may increase the disparities of urban and rural regions. Recent work goes beyond transportation infrastructure, incorporating other types of infrastructure including electricity transmission (Arkolakis and Walsh, 2022) and public works (Coury, Kitagawa, Shertzer, and Turner, 2022).

Another fertile topic of research is how congestion affects the value of infrastructure investment, a margin emphasized by Duranton and Turner (2011) but abstracted from in the literature above. Such a consideration is particularly important in the case of car traffic and have also risen into prominence due to the recent congestion in ports and sea routes after the COVID-19 recession. Allen and Arkolakis (forthcoming) propose a framework that allow the analysis of traffic congestion in the spatial framework, which has been extended to the global shipping network by Heiland, Moxnes, Ulltveit-Moe, and Zi (2019); Ducruet, Juhasz, Nagy, and Steinwender (2020); Ganapati, Wong, and Ziv (2021). Fan, Lu, and

Luo (2019) and Fuchs and Wong (2022) discuss the possibility of accommodating traffic in two or more coexisting transport networks with trans-modal route choices. Fajgelbaum and Schaal (2020) study optimal transportation networks in the presence of traffic congestion. Such frameworks would be ideal to be used to evaluate newly implemented city policies that aim on the reduction of traffic by imposing tolls in specific areas of the cities, such as the Singaporean traffic toll system or the congestion price system suggested for downtown Manhattan.

A classic example of changes in global geography are those that arise from changes in international trade costs and tariffs, which is commonly studied using gravity trade models, see e.g. (Anderson, 1979; Eaton and Kortum, 2002; Chaney, 2008; Anderson, 2011; Chaney, 2018). Such models are special cases of the global spatial equilibrium presented in Section 3 where the elasticity of labor supply is infinity $\varepsilon^S_{local} \to \infty$. Following the influential empirical findings of Autor, Dorn, and Hanson (2013), a large body of work has examined the impacts of changes in foreign demand (e.g. the rise of China) on the distribution of economic activity domestically (e.g. across locations within the U.S.), which in our framework, corresponds to differential changes in market access across locations in the U.S. arising from a productivity increase in China, see e.g. Caliendo, Dvorkin, and Parro (2019); Galle, RodrÃguez-Clare, and Yi (2017); Adao, Arkolakis, and Esposito (2019). Another example of a foreign demand shock differentially affecting domestic market accesses arises through tourism, see e.g. Faber and Gaubert (2019).

Alternative Spatial Linkages

The framework developed above focuses on spatial linkages between locations that arise through the trade of goods. But of course there are many ways in which people interact across space (see e.g. Christakis and Fowler (2009)), and there have been many exciting recent advances incorporating such interactions into spatial frameworks like the one developed here.

Following the seminal work of Ahlfeldt, Redding, Sturm, and Wolf (2015), a number of papers have examined the impact of spatial interactions through commuting flows in order to understand the spatial distribution of economic activity within a city, see e.g. Severen (2019); Tsivanidis (2018); Zarate Vasquez (2022). Monte, Redding, and Rossi-Hansberg (2018); Allen, Arkolakis, and Li (2015) combine trade and commuting spatial linkages in a single model.

A related literature incorporates additional spatial linkages arising through costly migration, extending the framework above to a dynamic setting, see Allen, de Castro Dobbin, and Morten (2018); Desmet, Nagy, and Rossi-Hansberg (2018); Tombe and Zhu (2019); Caliendo, Dvorkin, and Parro (2019); Allen and Donaldson (2020); Peters (2021); Kleinman, Liu, and

Redding (2021). While the steady state (or balanced growth path) of these models resemble the static framework above, they are also able to yield predictions on the adjustment path of an economy to changes in geography, which is empirically important.

Another spatial linkage garnering recent attention is through the formation of production linkages across firms. Such production networks are empirically important (see Bernard, Moxnes, and Ulltveit-Moe (2018); Bernard, Moxnes, and Saito (2019)), and a recent literature has modeled firm production networks linkages across different countries and locations (see Eaton, Kortum, and Kramarz (2022), Panigrahi (2021) Arkolakis, Huneeus, and Miyauchi (2021)).

A final spatial linkage we mention is knowledge diffusion across space. Although difficult to measure directly, recent work relying on new data sets (Couture, Dingel, Green, and Handbury, 2020; Atkin, Chen, and Popov, 2022; Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob, et al., 2022a,b) or technological innovations (Allen, 2014; Steinwender, 2018; Akerman, Leuven, and Mogstad, 2022) document the importance of the spatial spread of information on the distribution of economic activity.

Indeed the possibilities of adding additional spatial linkages or combining multiple types (or multiple layers) of linkages are limitless. Extending the framework to include such interactions brings more realism and helps to illuminate the many ways in which geography shapes the spatial economy.

6 Conclusion

This article served three purposes. First, it was meant as an introduction to the reader about how geography shapes the spatial distribution of economic activity. In the classic Rosen-Roback framework, the answer depends solely on the "local" geography of each location and the equilibrium spatial distribution can be determined through familiar analysis of supply and demand curves. The major innovation of the new generation of economic geography models is to incorporate the spatial linkages between locations – putting space into the spatial model. The equilibrium can continue to be understood using the same supply and demand curves, but appropriately augmented to incorporate the impacts of the "global" geography.

The second purpose was to guide the reader through the process of combining these spatial models with spatial data to understand how geography shapes the real world spatial economy. Detailed spatial data are now readily available and researchers can apply these data to the theory using the well understood process of estimating supply and demand curves. With spatial linkages between locations arise potential pitfalls in estimation, but we offer strategies for traversing such issues. The end result is the ability to recover the underlying local and global geography such that the theory and data exactly correspond, allowing a researcher the ability to assess the impacts of any change in geography on the real world spatial distribution of economic activity.

Finally, we demonstrate the power of this close marriage between theory and data by highlighting the many types of questions that can be addressed. The types of questions and topics that can be examined using the framework here spans an incredibly wide range of topics, spanning economic history, environmental, labor, public finance, urban, and international, to name a few. This is an exciting time to be working on spatial issues: we have a new set of tools applicable to many interesting questions, most of which have yet to be tackled.

References

Acemoglu, D., and P. Restrepo (2020): “Robots and jobs: Evidence from US labor markets,” *Journal of Political Economy*, 128(6), 2188–2244.

Adao, R., C. Arkolakis, and F. Esposito (2019): “General Equilibrium Effects in Space: Theory and Measurement,” Working Paper 25544, National Bureau of Economic Research.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): “The economics of density: Evidence from the Berlin Wall,” *Econometrica*, 83(6), 2127–2189.

Akerman, A., E. Leuven, and M. Mogstad (2022): “Information frictions, internet, and the relationship between distance and trade,” *American Economic Journal: Applied Economics*, 14(1), 133–63.

Allen, T. (2014): “Information frictions in trade,” *Econometrica*, 82(6), 2041–2083.

Allen, T., and C. Arkolakis (2014): “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*.

——— (forthcoming): “The Welfare Effects of Transportation Infrastructure Improvements,” *The Review of Economic Studies*.

Allen, T., C. Arkolakis, and X. Li (2015): “Optimal City Structure,” *Yale University, mimeograph*.

——— (2020): “On the Equilibrium Properties of Network Models with Heterogeneous Agents,” Working Paper 27837, National Bureau of Economic Research.

Allen, T., C. Arkolakis, and Y. Takahashi (2020): “Universal gravity,” *Journal of Political Economy*, 128(2), 393–433.

Allen, T., C. de Castro Dobbin, and M. Morten (2018): “Border Walls,” Working Paper 25267, National Bureau of Economic Research.

Allen, T., and D. Donaldson (2020): “Persistence and Path Dependence in the Spatial Economy,” Working Paper 28059, National Bureau of Economic Research.

Almagro, M., and T. Dominguez-Iino (2019): “Location sorting and endogenous amenities: Evidence from amsterdam,” *NYU, mimeograph*.

ALTHOFF, L., F. ECKERT, S. GANAPATI, AND C. WALSH (2022): "The geography of remote work," *Regional Science and Urban Economics*, 93, 103770.

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

——— (2011): "The Gravity Model," *Annual Review of Economics*, 3(1), 133–160.

ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1), 170–192.

ARKOLAKIS, C., F. HUNEEUS, AND Y. MIYAUCHI (2021): "Spatial production networks," *Yale University, Unpublished manuscript.*

ARKOLAKIS, C., AND C. WALSH (2022): "Clean Growth," *University of Columbia, mimeo.*

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

ATKIN, D., M. K. CHEN, AND A. POPOV (2022): "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley," Working Paper 30147, National Bureau of Economic Research.

AUTOR, D. H., D. DORN, AND G. H. HANSON (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6), 2121–68.

BALBONI, C. A. (2019): "In harm's way? infrastructure investments and the persistence of coastal cities," Ph.D. thesis, London School of Economics and Political Science.

BALDWIN, R., AND D. TAGLIONI (2006): "Gravity for Dummies and Dummies for Gravity Equations," Working Paper 12516, National Bureau of Economic Research.

BAUM-SNOW, N., J. V. HENDERSON, M. A. TURNER, Q. ZHANG, AND L. BRANDT (2020): "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 115, 103124.

BERNARD, A. B., A. MOXNES, AND Y. U. SAITO (2019): "Production networks, geography, and firm performance," *Journal of Political Economy*, 127(2), 639–688.

BERNARD, A. B., A. MOXNES, AND K. H. ULLTVEIT-MOE (2018): "Two-sided heterogeneity and trade," *Review of Economics and Statistics*, 100(3), 424–439.

Bleakley, H., and J. Lin (2012): “Portage and path dependence,” *Quarterly Journal of Economics*, 127(2), 587–644.

Bustos, P., B. Caprettini, and J. Ponticelli (2016): “Agricultural productivity and structural transformation: Evidence from Brazil,” *American Economic Review*, 106(6), 1320–65.

Caliendo, L., M. Dvorkin, and F. Parro (2019): “Trade and labor market dynamics: General equilibrium analysis of the china trade shock,” *Econometrica*, 87(3), 741–835.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2018): “The impact of regional and sectoral productivity changes on the US economy,” *The Review of economic studies*, 85(4), 2042–2096.

Card, D., A. Mas, and J. Rothstein (2008): “Tipping and the Dynamics of Segregation,” *The Quarterly Journal of Economics*, 123(1), 177–218.

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

——— (2018): “The gravity equation in international trade: An explanation,” *Journal of Political Economy*, 126(1), 150–177.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022a): “Social capital I: measurement and associations with economic mobility,” *Nature*, 608(7921), 108–121.

——— (2022b): “Social capital II: determinants of economic connectedness,” *Nature*, 608(7921), 122–134.

Chiovelli, G., S. Michalopoulos, and E. Papaioannou (2018): “Landmines and Spatial Development,” Working Paper 24758, National Bureau of Economic Research.

Christakis, N. A., and J. H. Fowler (2009): *Connected: The surprising power of our social networks and how they shape our lives.* Little, Brown Spark.

Conte, B., K. Desmet, D. K. Nagy, and E. Rossi-Hansberg (2021): “Local sectoral specialization in a warming world,” *Journal of Economic Geography*, 21(4), 493–530.

Costinot, A., and A. Rodríguez-Clare (2014): “Trade theory with numbers: Quantifying the consequences of globalization,” in *Handbook of international economics*, vol. 4, pp. 197–261. Elsevier.

Coury, M., T. Kitagawa, A. Shertzer, and M. Turner (2022): “The Value of Piped Water and Sewers: Evidence from 19th Century Chicago,” Working Paper 29718, National Bureau of Economic Research.

Couture, V., J. Dingel, A. Green, and J. Handbury (2020): “Quantifying Social Interactions Using Smartphone Data,” *Unpublished manuscript*.

Cruz Alvarez, J. L., and E. Rossi-Hansberg (2021): “The economic geography of global warming,” Becker Friedman Institute for Economics Working Paper 2021-130, University of Chicago.

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

Dekle, R., J. Eaton, and S. Kortum (2008): “Global Rebalancing with Gravity: Measuring the Burden of Adjustment,” *IMF Staff Papers*, 55(3), 511–540.

Desmet, K., R. E. Kopp, S. A. Kulp, D. K. Nagy, M. Oppenheimer, E. Rossi-Hansberg, and B. H. Strauss (2021): “Evaluating the Economic Cost of Coastal Flooding,” *American Economic Journal: Macroeconomics*, 13(2), 444–86.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018): “The geography of development,” *Journal of Political Economy*, 126(3), 903–983.

Diamond, R. (2016): “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980-2000,” *American Economic Review*, 106(3), 479–524.

Diamond, R., and C. Gaubert (2022): “Spatial sorting and inequality,” *Annual Review of Economics*, 14, 795–819.

Diamond, R., and T. McQuade (2019): “Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development,” *Journal of Political Economy*, 127(3), 1063–1117.

Dijkstra, E. W. (1959): “A note on two problems in connexion with graphs,” *Numerische mathematik*, 1(1), 269–271.

Dingel, J. I., and B. Neiman (2020): “How many jobs can be done at home?,” *Journal of Public Economics*, 189, 104235.

Dingel, J. I., and F. Tintelnot (2020): “Spatial Economics for Granular Settings,” Working Paper 27287, National Bureau of Economic Research.

Disdier, A.-C., and K. Head (2008): “The puzzling persistence of the distance effect on bilateral trade,” *The Review of Economics and statistics*, 90(1), 37–48.

Donaldson, D. (2018): “Railroads of the Raj: Estimating the impact of transportation infrastructure,” *American Economic Review*, 108(4-5), 899–934.

Donaldson, D., and R. Hornbeck (2016): “Railroads and American economic growth: A "market access" approach,” *The Quarterly Journal of Economics*, 131(2), 799–858.

Donaldson, D., and A. Storeygard (2016): “The view from above: Applications of satellite data in economics,” *Journal of Economic Perspectives*, 30(4), 171–98.

Ducruet, C., R. Juhasz, D. K. Nagy, and C. Steinwender (2020): “All aboard: the effects of port development,” CEP Discussion Papers dp1734, Centre for Economic Performance, LSE.

Duranton, G., and M. A. Turner (2011): “The fundamental law of road congestion: Evidence from US cities,” *American Economic Review*, 101(6), 2616–52.

Eaton, J., and S. Kortum (2002): “Technology, Geography and Trade,” *Econometrica*, 70(5), 1741–1779.

Eaton, J., S. S. Kortum, and F. Kramarz (2022): “Firm-to-Firm Trade: Imports, Exports, and the Labor Market,” Working Paper 29685, National Bureau of Economic Research.

Faber, B., and C. Gaubert (2019): “Tourism and economic development: Evidence from Mexico’s coastline,” *American Economic Review*, 109(6), 2245–93.

Fajgelbaum, P., and C. Gaubert (2018): “Optimal Spatial Policies, Geography and Sorting,” Working Paper 24632, National Bureau of Economic Research.

Fajgelbaum, P. D., and E. Schaal (2020): “Optimal transport networks in spatial equilibrium,” *Econometrica*, 88(4), 1411–1452.

Fan, J., Y. Lu, and W. Luo (2019): “Valuing domestic transport infrastructure: a view from the route choice of exporters,” *The Review of Economics and Statistics*, pp. 1–46.

Fuchs, S., and W. F. Wong (2022): “Multimodal Transport Networks,” *University of Oregon, mimeo.*

Fujita, M., P. Krugman, and A. J. Venables (1999): *The Spatial Economy: Cities, Regions, and International Trade.* MIT Press, Boston, Massachussetts.

Galle, S., A. RodrÃguez-Clare, and M. Yi (2017): “Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade,” Working Paper 23737, National Bureau of Economic Research.

Ganapati, S., W. F. Wong, and O. Ziv (2021): “Entrepot: Hubs, Scale, and Trade Costs,” Working Paper 29015, National Bureau of Economic Research.

Gaubert, C., P. M. Kline, and D. Yagan (2021): “Place-Based Redistribution,” Working Paper 28337, National Bureau of Economic Research.

Glaeser, E. L., and J. D. Gottlieb (2009): “The wealth of cities: Agglomeration economies and spatial equilibrium in the United States,” *Journal of economic literature*, 47(4), 983–1028.

Head, K., and T. Mayer (2013): *Gravity equations: Workhorse, toolkit, and cookbook.* Centre for Economic Policy Research.

Heblich, S., S. J. Redding, and D. M. Sturm (2020): “The making of the modern metropolis: evidence from London,” *The Quarterly Journal of Economics*, 135(4), 2059–2133.

Heiland, I., A. Moxnes, K. H. Ulltveit-Moe, and Y. Zi (2019): “Trade from space: Shipping networks and the global implications of local shocks,” *Available at SSRN 3504623.*

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

Henderson, J. V., A. Storeygard, and D. N. Weil (2012): “Measuring economic growth from outer space,” *American Economic Review*, 102(2), 994–1028.

Henkel, M., E. Kwon, and P. Magontier (2022): “The Unintended Consequences of Post-Disaster Policies for Spatial Sorting,” *MIT Center for Real Estate Research Paper*, (22/08).

Jaworski, T., and C. T. Kitchens (2019): "National policy for regional development: Historical evidence from Appalachian highways," *Review of Economics and Statistics*, 101(5), 777–790.

Kleinman, B., E. Liu, and S. J. Redding (2021): "Dynamic Spatial General Equilibrium," Working Paper 29101, National Bureau of Economic Research.

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

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

Lee, S. (2022): "When Cities Grow: Urban Planning and Segregation in the Prewar US," *University of Michigan, mimeo.*

Manson, S. (2020): "IPUMS National Historical Geographic Information System: Version 15.0," .

Matsuyama, K. (1991): "Increasing returns, industrialization, and indeterminacy of equilibrium," *Quarterly Journal of Economics*, 106(2), 617–650.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, migration, and local employment elasticities," *American Economic Review*, 108(12), 3855–90.

Nordhaus, W., and X. Chen (2006): "Geographically based economic data (G-Econ)," http://gecon.yale.edu/.

Ospital, A. (2022): "Institutions and Spatial Exposure to Environmental Risk," *Unpublished manuscript.*

Panigrahi, P. (2021): "Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity," *Available at SSRN 3982029.*

Peters, M. (2021): "Market Size and Spatial Growth - Evidence from Germanys Post-War Population Expulsions," Working Paper 29329, National Bureau of Economic Research.

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., and A. J. Venables (2004): "Economic geography and international inequality," *Journal of international Economics*, 62(1), 53–82.

REDDING, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101, 148–167.

REDDING, S. J., AND E. ROSSI-HANSBERG (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

REDDING, S. J., AND D. M. STURM (2016): "Estimating Neighborhood Effects: Evidence from War-time Destruction in London," Working Papers 2016-6, Princeton University Economics Department.

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

ROSEN, S. (1979): "Wage-based indexes of urban quality of life," *Current issues in urban economics*, pp. 74–104.

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

SEVEREN, C. (2019): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *The Review of Economics and Statistics*, pp. 1–99.

STEINWENDER, C. (2018): "Real effects of information frictions: When the states and the kingdom became united," *American Economic Review*, 108(3), 657–96.

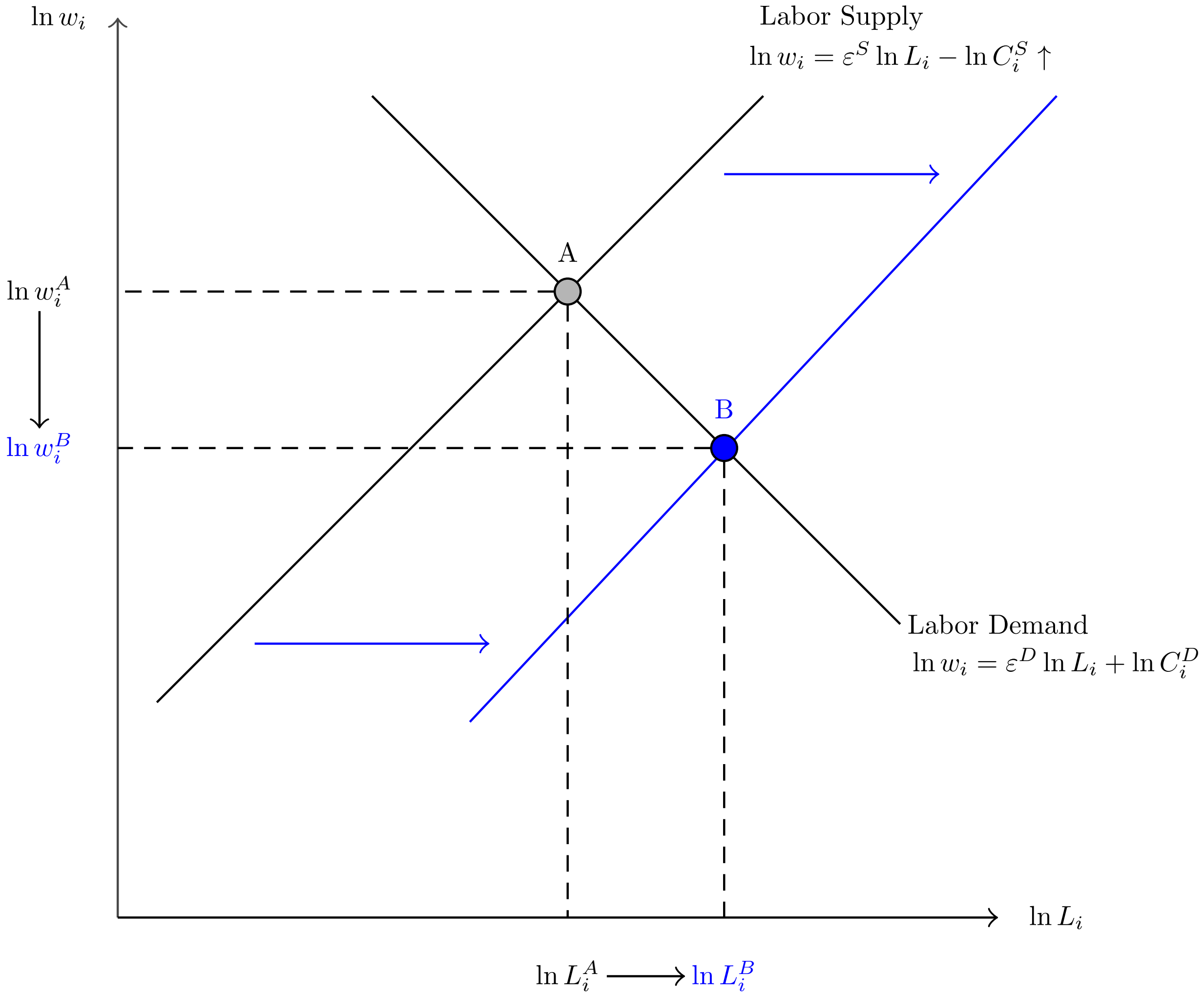
TOMBE, T., AND X. ZHU (2019): "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 109(5), 1843–72.

TSITSIKLIS, J. N. (1995): "Efficient algorithms for globally optimal trajectories," *IEEE transactions on Automatic Control*, 40(9), 1528–1538.

TSIVANIDIS, N. (2018): "The aggregate and distributional effects of urban transit infrastructure: Evidence from bogota's transmilenio," *Unpublished manuscript.*

ZARATE VASQUEZ, R. D. (2022): "Spatial Misallocation,Informality, and Transit Improvements : Evidence from Mexico City," Policy Research Working Paper Series 9990, The World Bank.

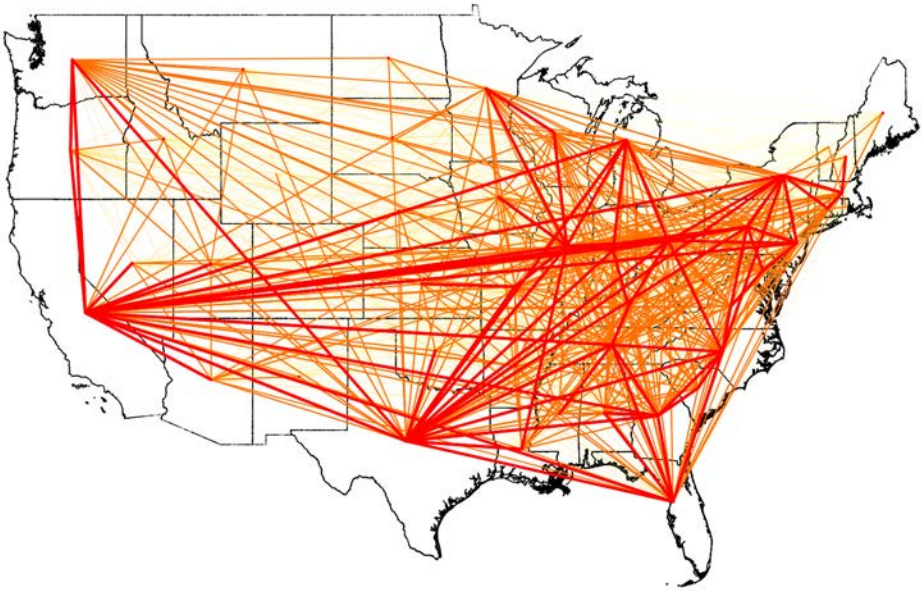
Figure 1: A supply shock in the local spatial equilibrium



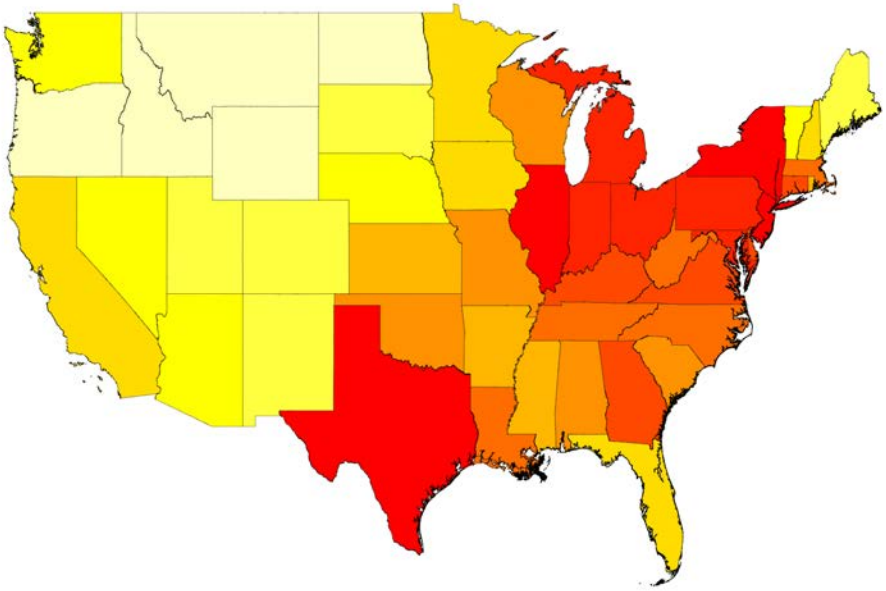
Notes: This figure illustrates the effect of an increase in the labor supply shifter on the equilibrium population and wages in a local spatial economy.

Figure 2: Spatial Linkages and Market Access

(a) Interstate trade flows



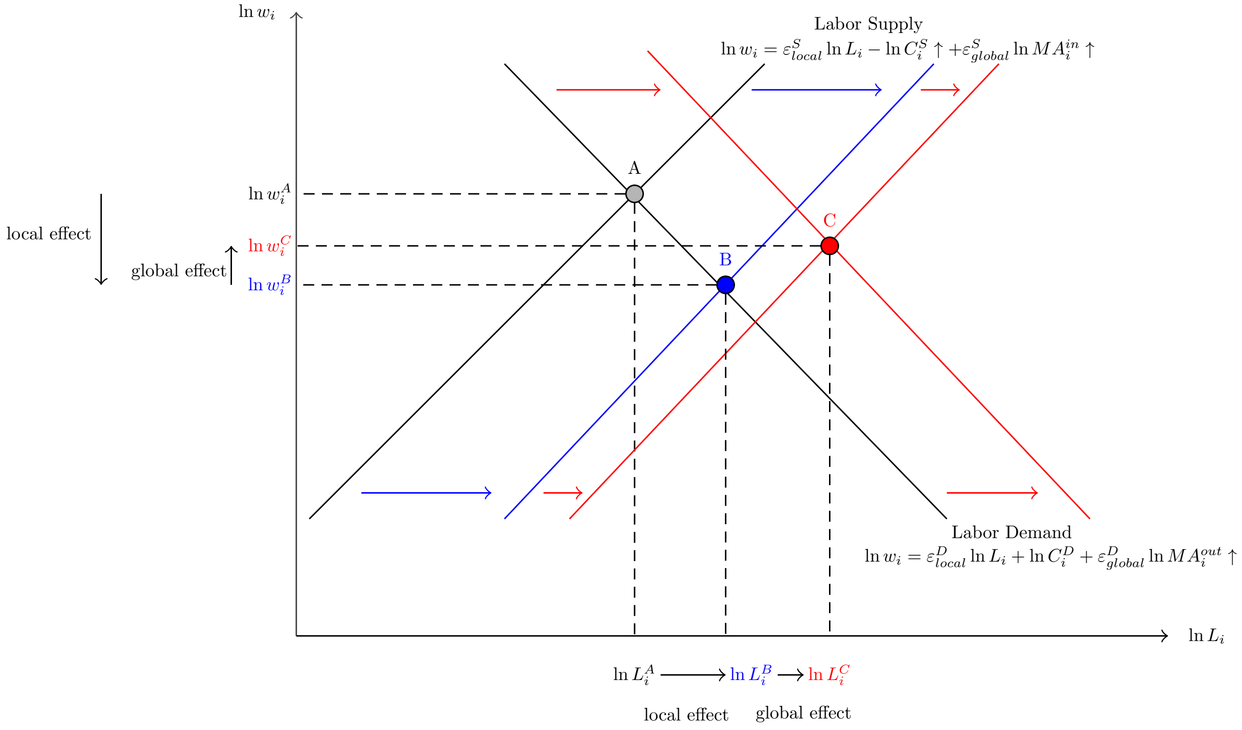
(b) Market access



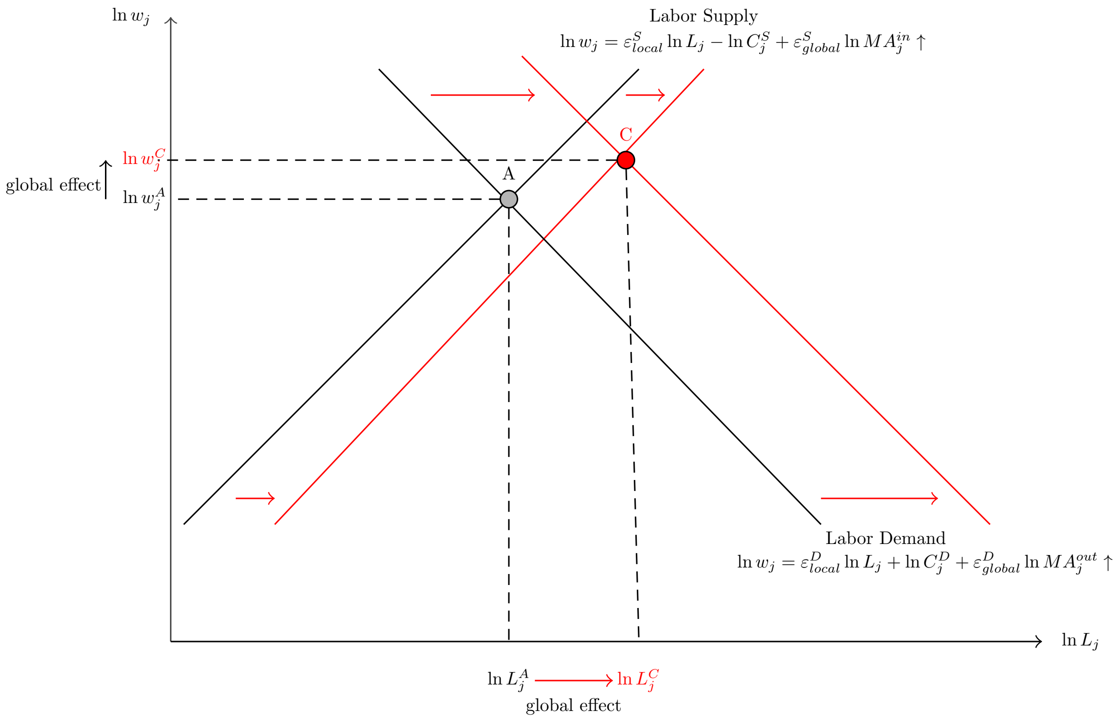
Notes: This figure illustrates the spatial linkages across U.S. states arising from trade flows. Panel (a) depicts the relative size of state-to-state bilateral trade flows, with thicker red lines indicating larger values and thinner yellow lines indicating smaller values. Panel (b) indicates the resulting (outward) market access of each state assuming trade costs T_{ij} are inversely proportional to distance.

Figure 3: A supply shock in the global spatial equilibrium

(a) The directly affected location



(b) An indirectly affected location



Notes: This figure illustrates the effect of an increase in the labor supply shifter in one location its own equilibrium population and wages (panel a) and another neighboring location (panel b).

A Appendix

In this Appendix, we provide detailed derivations of the results above.

A.1 A "Local" Spatial Economic Model

In this section, we provide a micro-foundation for the "local" supply and demand equations (1) and (2).

Consider a region comprising N different locations embedded in a larger economy. Agents can move freely across locations (or between these locations and the rest of the economy). Agents both produce and consume goods wherever they choose to live. Suppose that each location produces a homogeneous variety of good (e.g. corn). This good is freely traded across locations, but as all regions produce only that good there is no trade. We instead normalize the price of that good to one, $p = 1$, so that it provides a reference price to determine wages across locations.

The consumer problem is very simple in this context. Agents maximize welfare:

$$W_j = c_j \times u_j, \tag{7}$$

where c_j is the consumption in region j and u_j is a (non-consumption) amenity of residing in location j. The budget constraint of the consumer is simply $p \times c_j = w_j \iff c_j = w_j$, i.e. consumers consume an amount of the reference good equal to their wage.

Producers use labor and capital to produce the final good. The production function in location j is given by:

$$Y_j = A_i K_i^{\theta} L_i^{1-\theta},$$

where $\theta \in [0, 1)$. Factor markets are perfectly competitive, so the marginal productivity of labor equals the wage and the marginal productivity of capital equals the rental rate:

$$w_j = \theta A_i K_i^{\theta} L_i^{-\theta},\ r_j = (1-\theta) A_i K_i^{\theta-1} L_i^{1-\theta}.$$

It is evident that given capital the marginal productivity of labor declines with higher population and thus the wage in a location decreases when the population increases in that location. Assuming capital is fully mobile across locations so that the rental rate is equalized across locations (i.e. $r_j = r$), we have:

$$\frac{K_i}{L_i} = \left(\frac{r}{(1-\theta) A_i}\right)^{1/(\theta-1)}$$

and replacing in the wage equation yields:

$$w_i = \theta A_i \left(\frac{K_i}{L_i}\right)^{\theta} = \theta A_i \left(\frac{r}{(1-\theta)A_i}\right)^{\frac{\theta}{\theta-1}} = \theta A_i^{\frac{1}{1-\theta}} \left(\frac{r}{1-\theta}\right)^{\frac{\theta}{\theta-1}} \tag{8}$$

This model abstracts from a number of potentially important mechanisms, including other factors of production (like land), the consumption of non-tradables (like housing), possible heterogeneous preferences of different agents for different locations (e.g. I like beach and you like the mountains), economies of scale in production, etc. It turns out that a simple extension of the framework above is able to incorporate any combination of these different forces. Suppose that the productivity of a worker in a location depends in part on the total number of workers in that location:

$$A_i = \bar{A}_i L_i^{\alpha}, \tag{9}$$

where α may be positive or negative. Similarly, suppose that the amenity an agent derives from residing in a location depends in part on the total number of residents in that location:

$$u_i = \bar{u}_i L_i^{\beta}, \tag{10}$$

where again β may be positive or negative. In the model above, we have implicitly assumed $\alpha = \beta = 0$, but there are many reasons to think that α and β may be non-zero. For example, α may be negative if there is a fixed factor of production (like land) so that the more workers in a location, the less land there is per worker, driving down worker productivity. Alternatively, α may be positive if there are economies of scale in production. Similarly, β may be negative if residents also consume a local non-tradeable (like housing) that is in fixed supply, so that rent is driven up as the number of residents in a location increases. Or perhaps β is positive if greater population density induces greater supply of amenities (e.g. better parks). Allen and Arkolakis (2014) provide various micro-foundations for α and β along these lines.

Combining equations (8) and (9), we obtain the following labor demand equation:

$$\ln w_i = \frac{\alpha}{1-\theta} \ln L_i + \ln \alpha \left(\frac{r}{1-\theta}\right)^{\frac{\alpha}{1-\alpha}} \left(\bar{A}_i\right)^{\frac{\alpha}{1-\alpha}},$$

which is a special case of equation (1).

Similarly, because labor is perfectly mobile across locations, welfare equalization is equalized, i.e. $W = w_j \times u_j$. Combining welfare equalization with equation (10) yields the following labor supply equation:

$$\ln w_i = -\beta \ln L_i + \ln W \bar{u}_i^{-1},$$

which is a special case of equation (2).

A.2 A "Global" Spatial Economic Model

The goal in this section is to offer the derivations to the four equations comprising the equilibrium of the global spatial economy, namely equations (3), (4), (5) and (6). To do so, we rely on the same micro-economic foundations as in Allen and Arkolakis (2014), although as we discuss below, there are a number of alternative micro-economic foundations that also yield these equations (see e.g. Allen, Arkolakis, and Takahashi (2020)).

Consider a region comprising N different locations embedded in a larger economy. Agents can move freely across locations (or between these locations and the rest of the economy). Agents both produce and consume goods wherever they choose to live. Suppose that each location produces a distinct variety of good (e.g. French wine, Swiss cheese, ...). This assumption is called the "Armington" assumption (Armington (1969)). While clearly simplistic, it both makes the following derivations simpler and turns out to be mathematically equivalent to more realistic (but more complicated) models (see e.g. Eaton and Kortum (2002)).

Let us first consider the demand problem. Suppose that consumers like to consume many different varieties of goods. This "love of variety" creates an incentive for regions to trade with each other. In particular, we will assume that each agent has "constant elasticity of substitution" (CES) preferences such that if the agent lives in j and consumes quantity $\{q_{ij}\}_{j=1}^N$ of the variety of good from each location j she gets welfare:

$$W_j = \left(\sum_{i=1}^N q_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} u_j,$$

where $\sigma \geq 0$ is the elasticity of substitution (where higher σ indicates the agent is more willing to substitute one variety of good for another) and u_j is a (non-consumption) amenity of residing in location j. It is straightforward to show (but good practice to check!) that a utility-maximizing agent living in j with budget e_j and facing prices $\{p_{ij}\}_{i=1}^N$ will choose to spend $\{x_{ij}\}_{i=1}^N$, where:

$$x_{ij} = \frac{p_{ij}^{1-\sigma}}{\sum_{k=1}^N p_{kj}^{1-\sigma}} e_j.$$

and will receive welfare $W_j = e_j u_j / \left(\sum_{k=1}^N p_{kj}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$. Note that the share an agent spends of her income on each good does not depend on the level of her income, i.e. CES demand is homothetic. Since all agents living in a location face the same set of prices and CES demand is homothetic, we can can calculate the total amount spent by all agents living in location j on goods from i as:

$$X_{ij} = \frac{p_{ij}^{1-\sigma}}{\sum_{k=1}^N p_{kj}^{1-\sigma}} E_j, \tag{11}$$

where E_j is the total expenditure in j.

Now let us consider the supply problem. Suppose that the production of the differentiated varieties uses only labor as a factor of production and that each worker in location i can produce A_i units of the variety. Suppose too that it requires $t_{ij} \geq 0$ units of labor to ship a good from i to j. Finally, suppose that there is perfect competition, so that the price of goods are equal to their marginal cost of production. We then have that the price of a differentiated variety produced in i and sold in j is:

$$p_{ij} = \tau_{ij} \left(\frac{w_i}{A_i}\right), \tag{12}$$

where $\tau_{ij} \equiv 1 + t_{ij} A_i$ is the iceberg trade cost. Combining equations (11) and (12) yields the following gravity equation for trade flows:

$$X_{ij} = \frac{\tau_{ij}^{1-\sigma} \times \left(\frac{w_i}{A_i}\right)^{1-\sigma}}{\sum_k \tau_{kj}^{1-\sigma} \times \left(\frac{w_k}{A_k}\right)^{1-\sigma}} E_j. \tag{13}$$

Equation (13) is known as a *trade gravity equation* because it says that trade between locations is (a) proportional to the economic "size" of the origin and location; and (b) inversely proportional to the economic "distance" between the origin and destination. These two properties of trade flows are even more obvious when you re-write the equation as:

$$X_{ij} = T_{ij} \times \left(\frac{Y_i}{MA_i^{out}}\right) \times \left(\frac{E_j}{MA_j^{in}}\right), \tag{14}$$

where $T_{ij} \equiv \tau_{ij}^{1-\sigma}$ is the measure of (inverse) economic distance and we call $MA_j^{in} \equiv \sum_k \tau_{kj}^{1-\sigma} \times \left(\frac{w_k}{A_k}\right)^{1-\sigma}$ the *inward market access* and $MA_i^{out} \equiv \left(\frac{w_i}{A_i}\right)^{\sigma-1} Y_i$ the *outward market access.* As discussed in the main text, consumers in locations with higher inward market access benefit by being closer to the sellers of the goods they consume, whereas producers in locations with higher outward market access benefit from being closer to the buyers of the goods they produce. The new variant of the gravity equation in (14) highlights that the appropriate measure of economic size combines both the total income or expenditure of a location and its market access.

As noted in the main text, the inward and outward market accesses are closely related. To see this, we introduce two accounting identities. First, the income Y_i of each location i is equal to its total sales, i.e. $Y_i = \sum_{j=1}^N X_{ij}$, which when combined with the gravity equation (14) yields:

$$MA_i^{out} = \sum_{j=1}^{N} T_{ij} \times \left(\frac{E_j}{MA_j^{in}} \right). \tag{15}$$

Second, the expenditure E_j of each location j is equal to its total purchases, i.e. $E_j = \sum_{i=1}^N X_{ij}$, which when combined with the gravity equation (14) yields:

$$MA_j^{in} = \sum_i T_{ij} \times \left(\frac{Y_i}{MA_i^{out}} \right). \tag{16}$$

Equations (15) and (16) correspond to equations (3) and (4) in the main text. One neat thing about equations (15) and (16) is that given observed data on income and expenditures and estimates of (inverse) economic distances T_{ij}, you can use the two equations to uniquely identify (up-to-scale) the equilibrium inward and outward market access for every location. Another neat thing about the equations is that if the (inverse) economic distance is symmetric, i.e. $T_{ij} = T_{ji}$ and income is equal to expenditure, i.e. $Y_i = E_i$, then the inward and outward market access are equal up to scale, i.e. $MA_j^{in} \propto MA_j^{out}$ (which may be why oftentimes there is talk of "market access" without specifying if it is "inward" or "outward").

Equations (15) and (16) let you calculate the inward and outward market access given information on income and expenditure. But how you figure out the equilibrium income and expenditure in each location? To close the model, we impose three market clearing conditions. The first market clearing conditions has to do with the demand for labor in a location. We require that the the income earned in a location is paid out to labor, i.e. $w_i L_i = Y_i$. This is straightforward in this model, as labor is the only factor of production and there is perfect competition, although the condition would have to be modified in models with multiple factors of production or with market power and firm profits.

Combining this equilibrium condition with the definition of outward market access (i.e. $MA_i^{out} \equiv \left(\frac{w_i}{A_i}\right)^{\sigma-1} Y_i$) yields the following labor demand equation:

$$\ln w_i = -\frac{1}{\sigma} \ln L_i + \frac{\sigma - 1}{\sigma} \ln A_i + \frac{1}{\sigma} \ln MA_i^{out}, \tag{17}$$

which is a special case of equation(5).

The second market clearing condition has to do with the supply for labor in a location. We assume that workers are equally happy to live in all locations and that level of happiness is in turn equal to the happiness they would achieve by living elsewhere in the economy. This comes from the assumption that workers are freely mobile across different locations: if workers can move wherever, why would anyone live in a location that makes them less happy? Of course, in reality, there may be many reasons that workers may live in locations with low levels of happiness, e.g. idiosyncratic preferences for different locations (more on this below) or the cost of moving between locations (which requires extending the static framework here into a dynamic one, see e.g. Desmet, Nagy, and Rossi-Hansberg (2018); Caliendo, Parro, Rossi-Hansberg, and Sarte (2018); Allen and Donaldson (2020)). Let us suppose that the level of happiness W_i an agent gets from residing in location i depends on both her utility from consumption and from a local amenity u_i. Finally, let us normalize the level of happiness in the rest of the world to one. While it may seems like a consequential choice to treat our set of locations as a small region in a large global economy, it actually is not: the equilibrium distribution of economic activity (i.e. the relative populations and incomes in all locations) is identical to a setting where the aggregate population is fixed.

Combining this equilibrium condition with the definition of inward market access (i.e. $MA_i^{in} \equiv \sum_k \tau_{ki}^{1-\sigma} \times \left(\frac{w_k}{A_k}\right)^{1-\sigma}$) yields the following labor supply equation:

$$W_i = 1 \iff \ln w_i = -\ln u_i - \frac{1}{\sigma - 1} \ln MA_i^{in}, \tag{18}$$

which is a special case of equation (6), albeit one where the labor supply is perfectly elastic.

Finally, we impose that income is equal to expenditure, i.e. $E_i = Y_i$. This implies that the value of goods being sent out of each location is equal to the value of goods being sent into each location, i.e. that trade is balanced. Since the model is a static one, this makes sense (although it highlights that the model is not well suited to explaining trade deficits observed in the data, which presumably arise due to dynamic considerations). Together, the labor demand equation (17), the labor supply equation (18), and the market access equations (15) and (16) can be solved together to determine the equilibrium population and wages in all locations.[14]

The model above provides an explanation for the market access equations (3) and (4), as well as special cases of the general labor supply and demand equations (5) and (6). As in the Rosen-Roback framework described in Appendix A.1, we can incorporate the presence

---

[14] Because the equilibrium holds for any choice of units of wages, one also must choose a numeraire. In the companion Matlab code, we impose that the average wage across locations is equal to one.

of productivity and amenity spillovers of the form given in equations (9) and (10) to derive a more general form of the supply and demand curves.

Substituting equation (9) into the labor demand equation (17) yields the following modified labor demand curve:

$$\ln w_i = -\frac{1}{\sigma}\left(1-\alpha\left(\sigma-1\right)\right)\ln L_i + \frac{\sigma-1}{\sigma}\ln \bar{A}_i + \frac{1}{\sigma}\ln MA_i^{out}. \tag{19}$$

The more positive α, the flatter the downward sloping labor demand curve is, up to the point that $\alpha = \frac{1}{\sigma-1}$, at which point further increases in α actually cause the labor demand curve to shift upward! It is at this point that a "black hole" equilibrium becomes possible where all population is concentrated in a single location, see Fujita, Krugman, and Venables (1999).

Similarly, substituting equation (10) into the labor supply equation (18) yields the following modified labor supply curve:

$$\ln w_i = -\beta \ln L_i - \ln \bar{u}_i - \frac{1}{\sigma-1}\ln MA_i^{in}. \tag{20}$$

If β is negative (i.e. more people in a location reduce the amenity value of residing in that location), then the labor supply curve has a positive slope: to compensate perfectly mobile individuals for the amenity loss of the greater population, wages have to rise. As above, given the labor supply and demand equations along with the market access equations (15) and (16), we can solve the model to determine the equilibrium population and wages in all locations. But this has an interesting (and somewhat surprising conclusion): conditional on the slope of the labor supply and demand curves, the particular micro-foundation for a non-zero α and β do not matter for the equilibrium spatial distribution of economic activity. Or put another way, two different micro-foundations that both yield the same labor supply and demand curves will have the exact same implications for the equilibrium spatial distribution of economic activity.