

Quantitative Economic Geography *

Treb Allen and Costas Arkolakis

September 2024

Abstract

This handbook chapter presents the major advances made in the field of economic geography over the past decade. It starts by documenting a number of motivating empirical facts. It then shows how a quantitative economic geography model that combines the insights from two seminal models from an earlier generation can explain these facts. It then presents a unified quantitative economic geography framework that incorporates this and many other economic geography models. This unified framework is sufficiently tractable to characterize its equilibrium properties while sufficiently flexible to be combined with detailed spatial economic data to estimate the model parameters, conduct counterfactuals, and perform welfare analysis. The chapter concludes by discussing many extensions of the framework, some of which have already been explored and others which have not.

1 Introduction

Space matters. Life in rural New Hampshire is different than life in New York City: it is cleaner, quieter, and less crowded. But the wages are lower and—unless you are an avid outdoors person—the amenities are worse. Why do some people choose to live in New Hampshire while others choose New York City? Why are wages higher in New York City? And how much of what we observe about the spatial distribution of people and economic activity today is due to innate geographical differences (e.g. New Hampshire has mountains, whereas New York City is a natural harbor) versus geographic location (e.g. New Hampshire is far from most major cities, New York City is not) versus other economic forces (e.g. agglomeration)?

These questions form the bedrock of the study of economic geography. And while these topics have been studied for hundreds of years (see Smith (1776), von Thunen (1826), and Marshall (1890)), it took until the 1980s for the field to develop formal mathematical models for

*Allen: Department of Economics, Dartmouth College, treb.allen@dartmouth.edu. Arkolakis: Department of Economics, Yale University, costas.arkolakis@yale.edu. We thank Ayush Mathur and Qirui Wang for excellent research assistance and Ella McCoshan for various comments.

characterizing the equilibrium distribution of economic activity across space. And it was not until the past decade that these distinct mathematical frameworks have become unified into a single “quantitative” framework where the geography being considered is sufficiently realistic that it can be combined with real world data to provide precise answers to these questions. But the power of this new quantitative framework goes beyond explaining why the world looks the way it does: it also allows us to predict how changes in the world (say, to the cost of interacting across space) would change the equilibrium distribution of economic activity. In doing so, it opens up the possibility of using the framework to design better spatial policies.

The purpose of this handbook chapter is to provide an overview of this new quantitative economic geography framework. While we situate the framework relative to earlier economic models and try to highlight several exciting paths toward which the framework may evolve, the purpose of the chapter is not a literature review. Instead, our goal is to offer the reader a stand alone toolkit for using the quantitative economic geography framework in their own research. Toward that end, the focus of the chapter is two-fold: first, we try to offer the reader sufficient technical detail in order to see how (and why) the framework works the way it does without overwhelming with technicalities; second, we try to discuss how the framework can be combined with data in order to perform counterfactual analysis.

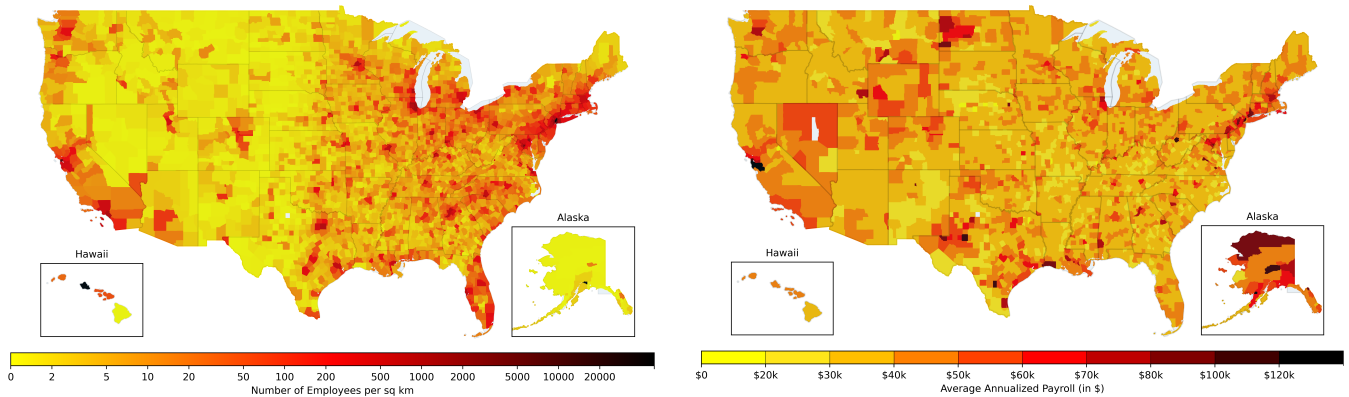
The chapter is organized as follows. We first motivate the key ingredients of the framework with a set of real world facts (Section 2). We then briefly summarize two seminal models developed in the 1980s and 1990s upon which the modern quantitative economic geography framework was built (Section 3). We next present the simplest version of the quantitative economic geography model and discuss a number of its properties (Section 4) before showing how this simple model has been extended in a number of different directions (Section 5). Section 6 presents a “universal” quantitative economic geography framework, for which all earlier models are special cases. Here we present the framework, characterize its equilibrium properties and derive expressions for its counterfactuals. Section 7 then provides a detailed description of how to combine the framework with spatial data to both estimate key model parameters and perform those counterfactuals. Finally, Section 8 discusses a number of exciting recent extensions to the framework before Section 9 identifies a number of interesting avenues for future research.

Before proceeding, we offer some advice on using this handbook chapter. Sections 2 and 3 are helpful for motivation and to see the genesis of the contemporary quantitative economic geography framework. Sections 4 and 5 offer particular micro-foundations for the “universal” framework. Section 6 is meant to offer a stand-alone description of the modern quantitative economic geography framework. This section presents the framework, characterizes its equilibrium properties, and derives expressions for its counterfactuals. Section 7 is then its empirical counterpart about how to take the framework to the data. Sections 8 and 9 offer our opinions of ideas of how the field is evolving and may evolve in the future. We have separated more

Figure 1: The spatial distribution of economic activity

(a) Workers per square kilometer

(b) Annual payroll per worker



Notes: The left panel depicts the number of workers per square kilometer; the right panel depicts the annual payroll per worker. The source for both panels is the 2017 U.S. County Business Patterns from the U.S. Census Bureau.

technical portions into appropriately labeled subsections, which can be skipped if the reader is interested in the results rather than the derivations themselves.

2 Economic geography facts

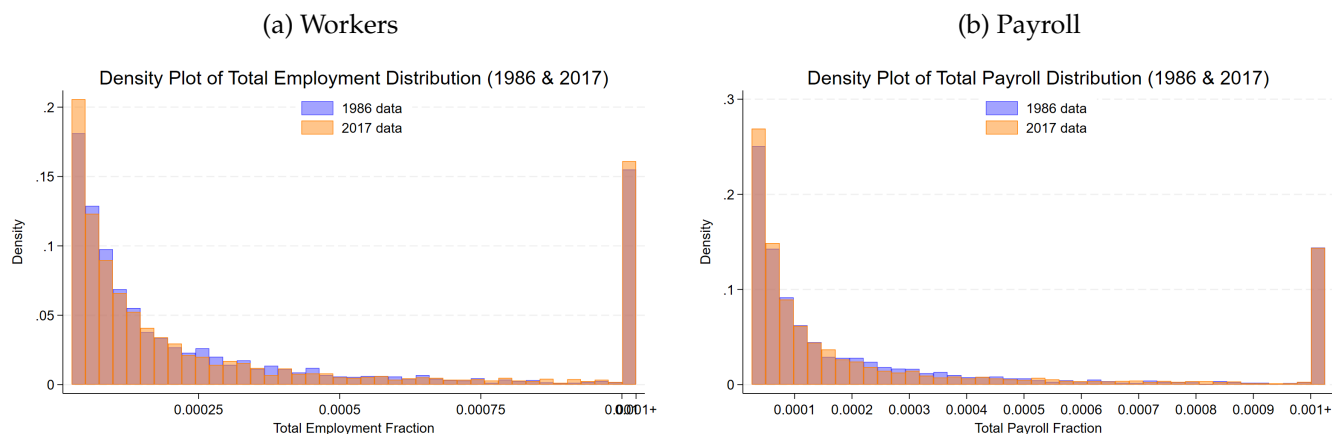
We begin by offering three facts that motivate the need for a quantitative economic geography framework.

2.1 Fact #1: Economic activity is distributed highly unequally across space

Economic activity across space is markedly unequal. Figure (1) depicts the spatial distribution of economic activity across counties in the United States (U.S.). The left panel depicts the population density, i.e. the number of workers per square kilometer. The population density differs by several orders of magnitude, with population densities below one worker per square kilometer in much of the mountain west and the population densities exceeding 1,000 workers per square kilometer along the eastern seaboard.

The right panel depicts the distribution of economic output per worker, i.e. the average annualized payroll per worker. Here too there is substantial spatial variation. Many counties around city centers, such as San Francisco, Manhattan, Seattle, and others, have an annual payroll of more than \$100,000 while in counties a few dozens kilometers away the annual payroll quickly falls below \$50,000. Because more highly populated locations have greater economic output per worker, the right panel demonstrates that total economic activity is even more spatially concentrated than the population density.

Figure 2: The spatial distribution of economic activity over time



Notes: This figure depicts the distribution across counties in employment (panel a) and output (panel b) for both 1986 and 2017. We calculate the fraction of total employment / output in each county and then plot the resulting distribution across counties. The source for both panels are the 1986 and 2017 U.S. County Business Patterns from the U.S. Census Bureau.

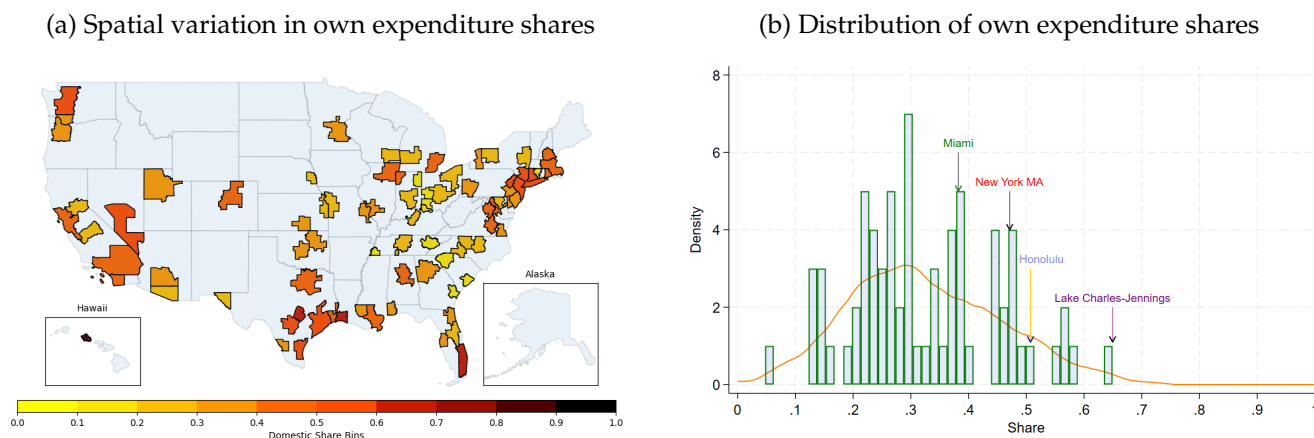
One possible explanation for the substantial spatial variation in total economic activity is that we are observing an economy in transition: for example, perhaps the population density is the largest along the eastern seaboard is simply due to those areas being older. Figure 2 shows the distribution of the fraction of total workers (left panel) and total output (right panel) across U.S. counties in both 1986 and 2017. As is evident, the spatial distribution of both population and output has remained remarkably consistent over the 30 year period, suggesting that the substantial spatial inequality we observe is not a temporary phenomenon.

This fact motivates the need for a framework that can explain the persistent presence of substantial spatial variation in both where people live and how much they earn.

2.2 Fact #2: Regional trade is important

Our second stylized fact is that regions within the U.S. are strikingly economically integrated. To show this, we use the 2017 Commodity Flow Survey (CFS) Public Use File to calculate the fraction of observed spending by a Metropolitan Statistical Area (MSA) on shipments that originated in the same MSA, i.e. its “own expenditure share.” There are 70 unique MSAs in our data. An own expenditure share of 1 indicates that an MSA is entirely in autarky, whereas a value of 0 indicates that an MSA only buys goods from elsewhere. Figure 3 depicts the spatial distribution of own expenditure shares across MSAs. As is evident, nearly all MSAs spend a majority on shipments from other MSAs, i.e. across-MSA trade comprises a majority of economic activity. That being said, own expenditure shares vary considerably. Major MSAs, such as New York or Miami have much larger own shares, close to a half or more. On the contrary, many small MSAs in the Midwest have own shares closer to 0.25. Geographic location appears

Figure 3: Own expenditure shares



Notes: Own expenditure share is calculated as the ratio of the value of within-MSA transactions to the total value of MSA transactions using the 2017 Commodity Flow Survey Public Use File from the U.S. Census Bureau.

to matter too: One of the MSAs with the highest own share is Honolulu, in Hawaii, nearly four thousand kilometers away from an MSA in the mainland USA.

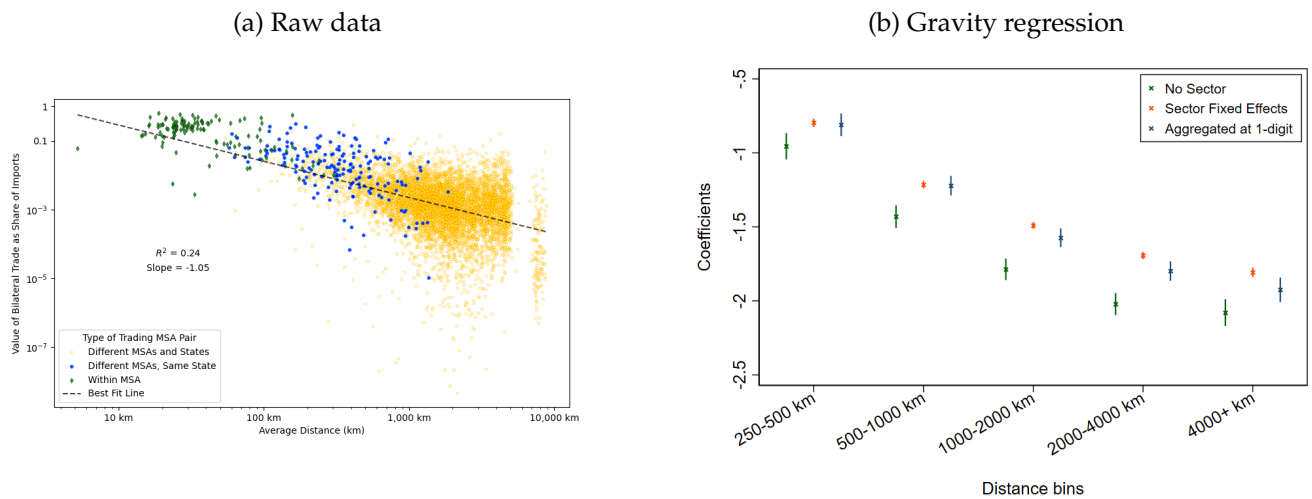
This fact motivates the need for a framework where different regions interact economically through the flow of goods.

2.3 Fact #3: Gravity holds

Our third and final stylized fact is that where a region is located and how large it is plays an important role in determining to the extent to which it interacts with other regions, i.e. gravity holds. Figure 4 reveals that the “gravity” trade pattern very familiar to international trade economists (see e.g. Anderson (1979), Eaton and Kortum (2002), and Anderson and Van Wincoop (2003)) also holds within countries: trade systematically declines with distance. We regress the common logarithm of the trade value between all the pairs of MSAs in our data to distance bins of the pair, 0-250 kilometers (km), 250-500, 500-1000, 1000-2000, 2000-4000, or more than 4000 and origin and destination fixed effects. We provide the raw observations on the left panel and the coefficients of the regression in the right panel. The left panel shows that trade (measured as the ratio of exports to all shipments to the destination MSA) is greatest within-MSA (green), followed by trade between different MSAs that belong to the same state (blue), and then unrelated MSAs. Trade flows decline substantially with distance; for example, locations that are 250-500 kilometers away have about 10% (10^{-1}) the value of trade as locations that less than 250 kilometers away, while locations that are more than 4000 kilometers away only about 1%.

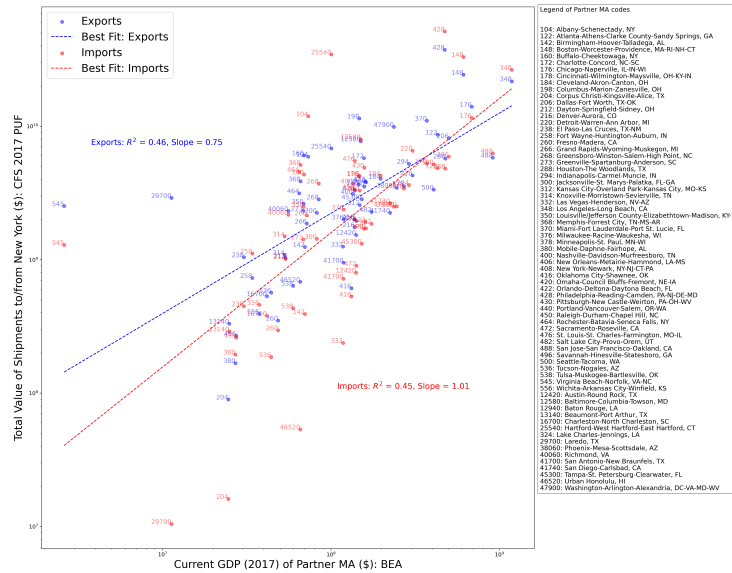
Size matters too. Following the spirit of the analysis in Head and Mayer (2014), Figure 5 plots the total trade value of trade (of both imports and exports) between the New York MSA and the Gross Domestic Product of its partner MSAs. Both incoming and outgoing trade

Figure 4: Distance matters



Notes: This figure uses the 2017 Commodity Flow Survey Dataset Public Use File to show that trade flows within the U.S. decline with distance. In the left panel, we present the scatter plot of the ratio of the value of bilateral trade flows to total MSA spending and the average distance between them. In the right panel, we report the results of a gravity regression of (log) trade flows on distance bins with origin and destination fixed effects. We construct bilateral trade flows and kilometric distance using the almost 6 million individual transactions provided by the Survey and aggregate them to bilateral MSA transactions using the weights provided by the CFS.

Figure 5: Size matters



Notes: This figure uses the 2017 Commodity Flow Survey Dataset Public Use File to show that trade flows within the U.S. increase with partner size. Each point represents the total trade between the New York MSA and another MSA; blue dots are exports from New York, red dots are imports into New York. We construct both average kilometric distance and bilateral trade value using the weights for each transaction provided by the CFS. GDP data at the MSA level are from the Bureau of Economic Analysis.

increases systematically with the size of the partner of New York state.

This fact motivates the need for a framework where the degree to which different regions interact through trade depends on their respective sizes and the distances between them.

3 Seminal models of the previous generation

We now present the two seminal models of economic geography from the 1980s and 1990s: the Rosen (1979)-Roback (1982) model of location choice and the Core-Periphery model developed by Krugman (1991). We first sketch out each model, then discuss their key insights, before noting the limitations of each framework, particularly in regard to explaining the stylized facts above.

3.1 The Rosen (1979)-Roback (1982) model of location choice

We consider first the celebrated model of location choice developed by Rosen (1979) and Roback (1982). This framework has found wide applicability in the urban and labor literature. For a more comprehensive review of the framework see Glaeser and Gottlieb, 2009 and Enrico, 2011.

3.1.1 Model sketch

Consider a world of N locations, which we index with i and j , and let $\mathcal{N} \equiv \{1, \dots, N\}$ be the set of locations, each of which produces an identical and costlessly traded numeraire good. Suppose there is a measure \bar{L} of identical and perfectly mobile agents. Agents choose in which location to live (i.e. produce and consume the numeraire good).

Agents who choose to live in location $j \in \mathcal{N}$ receive welfare:

$$W_j = w_j \times \bar{u}_j \times L_j^\beta, \quad (1)$$

where w_j is their wage, \bar{u}_j is the innate amenity value of residing in location j , L_j is the measure of agents residing in location j , and $\beta < 0$ reflects a negative congestion externality (i.e. the amenity value of residing in location j is lower the more people residing there).

Free mobility implies that welfare in all locations is equalized, i.e. there exists a scalar $W > 0$ such that $W_j = W$ for all $j \in \mathcal{N}$. Combining welfare equalization with the welfare expression (1) allows us to derive the following (inverse) *labor supply* curve:

$$L_j = w_j^{-1/\beta} \times \bar{u}_j^{-1/\beta} \times W^{1/\beta}. \quad (2)$$

Because $\beta < 0$, equation (2) shows that labor supply is upward sloping in the wage of a location with constant elasticity $-1/\beta$, with the amenity of a location acting as a (log) shifter of the supply curve. This is intuitive: the greater the innate amenity of a location, the more the labor supply curve is shifted outward, as residents need not be compensated with as high of a wage in order to choose to live there.

To produce the numeraire good, the Rosen (1979)-Roback (1982) model assumes labor and an additional fixed factor (e.g. capital) are combined in production function $Q_i = \frac{1}{1-\alpha} \bar{A}_i \bar{K}_i^\alpha L_i^{1-\alpha}$, where \bar{A}_i is the innate productivity of a region i , \bar{K}_i is the capital endowment, and α is the capital share. Inverting the first order condition of the firm with respect to labor allows us to derive the following (inverse) *labor demand* curve:

$$L_i = w_i^{-1/\alpha} \times \bar{A}_i^{1/\alpha} \bar{K}_i. \quad (3)$$

Because $\alpha > 0$, equation (3) shows that labor demand is downward sloping with constant elasticity $-1/\alpha$, with the productivity and capital endowment of a location as a (log) shifter of the demand curve. This too is intuitive: the greater the productivity and/or capital endowment of a location, the greater the marginal product of labor in that location, increasing labor demand.

The equilibrium distribution of economic activity—i.e. the equilibrium population and wages in each location—can be calculated by equating labor supply equation (2) with labor demand equation (3), where the equilibrium level of aggregate welfare W is determined by the

aggregate labor market clearing constraint $\bar{L} = \sum_j L_j$. Doing so yields:

$$L_i = \bar{u}_i^{\frac{1}{\alpha-\beta}} \times (\bar{A}_i \bar{K}_i^\alpha)^{\frac{1}{\alpha-\beta}} \times W^{\frac{1}{\beta-\alpha}} \quad (4)$$

$$w_i = \bar{u}_i^{-\frac{\alpha}{\alpha-\beta}} \times (\bar{A}_i \bar{K}_i^\alpha)^{-\frac{\beta}{\alpha-\beta}} \times W^{\frac{\alpha}{\alpha-\beta}} \quad (5)$$

$$W = \left(\sum_{i \in \mathcal{N}} \bar{u}_i^{\frac{1}{\alpha-\beta}} \times (\bar{A}_i \bar{K}_i^\alpha)^{\frac{1}{\alpha-\beta}} / \bar{L} \right)^{\alpha-\beta}. \quad (6)$$

Equations (4), (5), (6) completely specify the equilibrium spatial distribution of economic activity in terms of model fundamentals.

3.1.2 Insights

Despite its simplicity, the Rosen (1979)-Roback (1982) model yields a number of insights into the spatial distribution of economic activity. Equation (4) shows that the equilibrium population is relatively greater in locations with better innate amenities, productivities, or capital endowments. Equation (5) shows that equilibrium wages are relatively greater in locations with better innate productivities or capital endowments, but relatively lower in places with higher amenities (since people are more willing to live in nice places even if the wages are low). And equation (6) demonstrates that the aggregate welfare is greater with greater amenities, productivities, and capital endowments world wide. All three equations also highlight the important role that the parameters α and β play in determining how the underlying geography of the world shapes the equilibrium distribution of economic activity.

3.1.3 Limitations

In principle, the Rosen (1979)-Roback (1982) location choice model is able to explain the high degree of inequality in the spatial distribution of economic activity highlighted in Fact 1 above by ascribing large differences in innate amenities, productivities, and endowments across locations. But it is also immediately obvious that the Rosen (1979)-Roback (1982) model is unable to capture Facts 2 and 3 for the simple reason that it abstracts entirely from the linkages between locations through the flow of goods. Given the important role that trade plays in the regional economy (Fact 2), this is a glaring omission. Moreover, without trade, there is no concept of distance or heterogeneity in who trades with whom (Fact 3). Indeed, as equations (4) and (5) highlight, apart from the endogenous scalar W , the equilibrium economic activity in one location is unaffected by the geography and endowments of every other location.

3.2 The Krugman (1991) core-periphery model

We now consider the Core-Periphery model developed by Krugman (1991) (and further elaborated by Fujita, Krugman, and Venables (1999)). Unlike the Rosen (1979)-Roback (1982) residential choice model, the core-periphery model explicitly considers the economic linkages between locations. Yet it too has several limitations, which we will discuss below.

3.2.1 Model sketch

As above, we consider a world with N locations. Unlike above, however, we assume that each location is inhabited by an (endogenous) measure N_i of infinitesimal firms, each of which produces its own distinct variety. Because consumers will have a love of variety, this will create an incentive for trade across locations.

We start with the consumer. An agent $l \in \mathcal{L} \equiv [0, \bar{L}]$ chooses (a) where to live and (b) how much to consume of each location-specific differentiated variety. As above, agents are assumed to be freely mobile, which again implies that welfare is equalized across all inhabited locations. Consumption, however, is very different than above. Agents are assumed to have constant elasticity of substitution (CES) demand and solve the following utility maximization problem:

$$\max_{j \in \mathcal{N}, \{c_{ij}\}_{i \in \mathcal{N}} \geq 0} \left(\sum_{i \in \mathcal{N}} \int_{\Omega_i} c_{ij}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \text{ s.t. } \sum_{i \in \mathcal{N}} \int_{\Omega_i} p_{ij}(\omega) c_{ij}(\omega) \leq w_j, \quad (7)$$

where $\sigma > 1$ is the elasticity of substitution between varieties from different locations, Ω_i is the set of firms in location i (with $|\Omega_i| = N_i$), p_{ij} is the price of a good produced in i and sold in j , and w_j is the wage of an agent residing in location j .

This optimization results in the following consumer demand for an individual good ω in location j :

$$y_{ij}(\omega) = p_{ij}(\omega)^{1-\sigma} P_j^{\sigma-1} w_j. \quad (8)$$

It also implies that the indirect utility function of the consumer is given by the wage divided by the price index in region j :

$$W_j = w_j / P_j, \quad (9)$$

where $P_j^{1-\sigma} \equiv \sum_{i \in \mathcal{N}} \int_{\Omega_i} p_{ij}(\omega)^{1-\sigma} d\omega$. $P_j^{1-\sigma}$ is the consumer price index. Following Redding and Venables (2004), we refer to $P_j^{1-\sigma}$ is the (*inward*) *market access* of location j . Intuitively, locations with better inward market access are those who face lower prices for the goods they purchase because they are closer (have lower trade costs) to origins with lower cost producers.

We now consider the firm. Production uses one factor, labor. Firms are homogeneous but, unlike above, they produce differentiated products under monopolistic competition. Shipments to different regions are subject to trade costs that take the “iceberg” form such that $\tau_{ij} \geq 1$

units of the goods need to be produced in i for one unit to arrive in region j . Given the demand of the consumer, the firm sets an optimal markup over the marginal cost and the price of a good in i sold in region j , yielding the following equilibrium bilateral price:

$$p_{ij} = \frac{\sigma}{\sigma - 1} \frac{w_i}{A_i} \tau_{ij}, \quad (10)$$

where, as above, w_i is the wage and A_i is the productivity.

Substituting equation (10) into equation (8) and summing up across destination we obtain total firm demand, $y_i(\omega)$,

$$y_i(\omega) = \sum_j \frac{\left(\frac{\sigma}{\sigma-1} \frac{w_i}{A_i} \tau_{ij} \right)^{1-\sigma}}{P_j^{1-\sigma}} E_j,$$

where E_j is the total expenditure in location j .

It is straightforward to show that the associated firm profits with CES are $y_i(\omega) / \sigma$. To generate their unique variety firms need to hire a fixed number of workers, f_i^e , to enter the market. The equilibrium number of firms is then pinned down by a zero-profit condition requiring that the expected profits of an entrant equal its fixed cost:

$$w_i f_i^e = \frac{1}{\sigma} \sum_j \frac{\left(\frac{\sigma}{\sigma-1} \frac{w_i}{A_i} \tau_{ij} \right)^{1-\sigma}}{P_j^{1-\sigma}} E_j. \quad (11)$$

In equilibrium, the total labor income in a location equals the sum of sales from firms of that location to all locations:

$$w_i = \frac{N_i}{L_i} \sum_j \frac{\left(\frac{\sigma}{\sigma-1} \frac{w_i}{A_i} \tau_{ij} \right)^{1-\sigma}}{P_j^{1-\sigma}} E_j. \quad (12)$$

A comparison of equations (12) and (11) immediately shows that the equilibrium number of firms in a location is proportional to its population:

$$N_i = \frac{L_i}{\sigma f_i^e}. \quad (13)$$

We now can solve for the equilibrium. Substituting the price equation (10) into the indirect utility function of the consumer in equation (9) (using equation (13) to solve for the equilibrium number of firms) and applying welfare equalization yields the following (inverse) *labor supply* curve:

$$w_i = W \left[\sum_j \frac{L_j}{\sigma f_j^e} \left(\frac{\sigma}{\sigma-1} \frac{w_j}{A_j} \tau_{ji} \right)^{1-\sigma} \right]^{1/(1-\sigma)}. \quad (14)$$

Similarly, substituting equation (13) back into equation (11) yields the following (inverse)

labor demand curve:

$$w_i = A_i^{\frac{\sigma-1}{\sigma}} \left(\frac{\sigma}{\sigma-1} \right)^{\frac{1-\sigma}{\sigma}} \left(\frac{1}{\sigma f_i^e} \right)^{\frac{1}{\sigma}} \times \left(\sum_j \frac{(\tau_{ij})^{1-\sigma}}{P_j^{1-\sigma}} E_j \right)^{\frac{1}{\sigma}}. \quad (15)$$

A quick examination of equations (14) and (15) reveals a surprising discovery: both the labor supply and demand equations in the Core-Periphery model are perfectly inelastic to the labor in a location. This is in stark contrast to the Rosen-Roback framework above, where labor supply is upward sloping (as higher wages induce greater number of workers to reside in a location) and labor demand is downward sloping (as more labor entry leads to wages falling due to competitive pressures). Why are these forces not present here? On the supply side, there is no congestion force that directly reduces the utility of an agent residing in a populated location. On the demand side, competitive wage pressures are perfectly offset by the agglomeration force through firm entry.

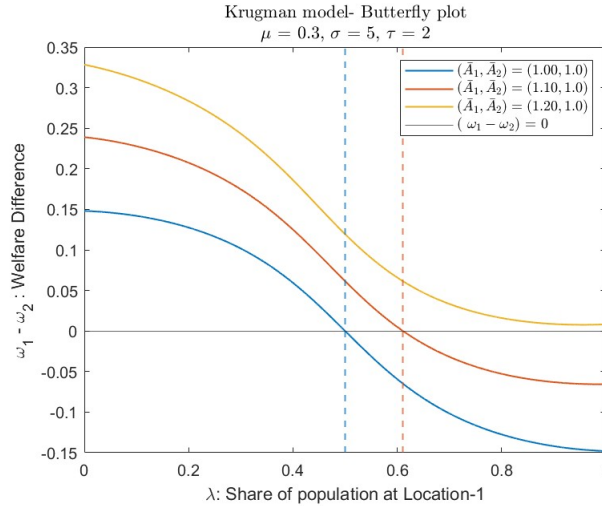
Instead of local labor, what matters here is the *global* market access. In the case of labor supply, workers are attracted to locations with better (inward) market access, as it increases their purchasing power; accordingly, such locations can offer lower wages and still attract workers. Similarly, labor demand depends on a weighted sum of the demand across all destinations $\Pi_i^{1-\sigma} \equiv \sum_j \frac{(\tau_{ij})^{1-\sigma}}{P_j^{1-\sigma}} E_j$. Analogous to above, we refer to this term as (*outward*) *market access*. Locations with better outward market access are those for whom producers face higher demand for the goods they sell because they are closer (have lower trade costs) to destinations with higher demand.

If you guessed that having a model with perfectly inelastic demand and supply curves may result in some issues, you are correct. It turns out that, absent of any additional intervention, the equilibrium is one in which there is complete concentration of all the workers in one location. Krugman (1991) solves this issue by adding an additional mechanism: the presence of an agriculture sector which is homogenous and costlessly traded across sectors. He assumes that consumers allocate a constant fraction of their income to this sector and the sector employs an exogenously given equal number of workers in each location. This additional mechanism is neither elegant nor necessary and so it has been left on the chopping block in the next generation of economic geography models.

3.2.2 Insights

The core-periphery model was a significant addition to the economic geography literature. It offered two major innovations: first, it introduced the concept of market access arising from the linkages between locations through the trade of goods, which Fact #2 and #3 above show are a crucial part of the spatial economy. Second, the presence of firm entry introduced an

Figure 6: The distribution of spatial economic activity in the Core-Periphery Setup



Source: Author’s calculations based on the core periphery setup. We chose calibrated parameters to closely resemble the simulations in Fujita, Krugman, and Venables (1999), bearing the assymetry in productivities.

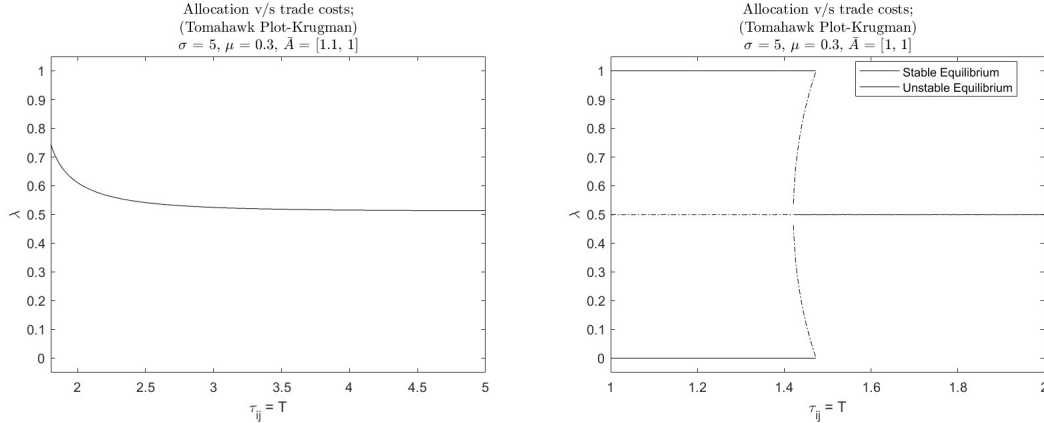
endogenous agglomeration force. Both the market access and agglomeration forces remain a cornerstone of modern quantitative economic geography, so it is worthwhile discussing briefly how they interact to determine the equilibrium distribution of economic activity in the core-periphery model.

Innate differences and economic agglomeration First, economic agglomerations can exacerbate innate difference across locations. Figure 6 plots the population allocations for two locations under different scenarios. Here we fix the productivity of location 2 to 1 and slightly vary the productivity of location 1 from 1, to 1.1, and 1.2. Of course, in the scenario where locations are symmetric, population is equally distributed between the two locations. A small change of 10% in the productivity of location 1 implies a disproportionate change in the share of population in location 1, more than 20%. In fact, if we increase the productivity to 1.2 the solution indicates that the welfare difference between the location is always positive. In other words, the agglomeration force in this scenario is so strong that with just a small change in parameters the model is lead to a corner solution where all the population moves to location 1.

This insight suggests that endogenous agglomeration forces may be part of the reason for such substantial variation in the distribution of economic activity across space presented in Fact #1 above.

Market access and economic outcomes Second, a key innovation of the Core-Periphery model is the presence of the market access term that appears because of trade costs. As trade costs increase, wage difference across locations are more pronounced. In Figure 7, left panel, we plot

Figure 7: The distribution of spatial economic activity in the Core-Periphery Setup



Source: Author’s calculations based on the core periphery setup. We chose calibrated parameters to closely resemble the simulations in Fujita, Krugman, and Venables (1999), barring the asymmetry in productivities.

the share of population in location 1 for the two locations with different productivities. and vary trade costs across a range of values. Naturally, if trade costs are too high the two locations act completely independently, but in the limit there are a few more workers in location 1, where the productivity is 10% higher. However, as the trade costs reduce the difference in productivity lead to more and more significant differences in the share of population in location 1. As trade costs decline the producer market access increases, which means higher labor demand and higher wages. This attracts population through entry and that creates an even stronger agglomeration force. The two forces coalesce at low trade costs leading to extreme levels of economic agglomeration and at trade costs lower than 2 the population share of location 1 is upward of 0.75.

This insight suggests that the strength of the economic linkages plays an important role in determining the equilibrium spatial distribution of economic activity.

Multiplicity of equilibria The final, and most intriguing feature of the Core-Periphery model is the possibility that arises for multiple equilibria. In Figure 7, right panel, we plot the equilibrium of the model with two symmetric locations, with different trade costs. This is the famous “Tomahawk graph” (see e.g. Chapter 5 in Fujita, Krugman, and Venables (1999)) that depicts the level of trade costs in the x-axis from starting from free trade and the share of population in location 1 on the y-axis. For large trade costs the equilibrium is unique and is evenly divided across the two locations. But below a threshold of trade costs the number of equilibria jumps to three, with the symmetric equilibria being still one possibility. The other two correspond to cases where population is concentrated in location 1 or population 2, due to economic agglomeration. The intuition for the emergence of the alternative equilibria with concentration in population is that as trade costs lower, economic agglomeration increases the effect of market

access to the manufacturing sector. After a critical point this effect is so strong that multiple equilibria arise. After that threshold the lower the trade costs lead to even more concentration, and eventually to only one location to be populated with manufacturing workers in equilibrium.

If multiplicity is possible, then conducting counterfactual analyses is potentially fraught, highlighting the need for a clear understanding of when the uniqueness of the spatial equilibrium can be assured.

3.2.3 Limitations

While the core-periphery model was pioneering in its introduction of spatial economic linkages and agglomeration forces, it has several notable limitations. First, because the equilibrium supply and demand only depends on market access and not local labor, even when considering simple geographies, an equilibrium may not exist or may be unrealistic (e.g. featuring complete concentration in a single location even if location fundamentals are not very different). Second, by relying entirely on firm entry to generate agglomeration forces, the model features a strong and inflexible agglomeration force that can generate economic responses to changes in the underlying geography that appear unrealistic (e.g. featuring large changes in the share of population to small changes in trade costs). Both these limitations are exacerbated when confronting real world data featuring many locations and complex geographies.

3.3 The current generation

While both the location choice model of Rosen (1979) and Roback (1982) and the Core-Periphery model of Krugman (1991) offer important new insights, neither alone offer an appropriate starting point to build a quantitative economic geography framework. But, one might wonder what could happen if we combine the richness of the economic linkages in the Core-Periphery model with the tractability of the location choice model of Rosen and Roback. Is that possible? And what could happen if we replace the extreme agglomeration forces induced by entry with alternative ways of modeling agglomeration, borrowing ideas from the literature in labor economics that uses the Rosen-Roback model (see for example the review of Enrico (2011)) or from international trade work that models the presence of Marshallian externalities (e.g. M. Grossman and Rossi-Hansberg (2010) and Kucheryavyy, Lyn, and Rodríguez-Clare (2023)). If we could add these features in a seamless way, we could revisit some of the facts in our introduction and, for example, ask how much of the increasingly unequal distribution of economic activity is due to reductions in trade costs, increases in trade openness, etc. This is what we turn to next.

4 A simple quantitative economic geography model

We now turn to presenting a simple quantitative economic geography framework. This framework, which builds off both the location choice and core-periphery models of the previous section, is based on Allen and Arkolakis (2014). As we will then discuss in detail in Section 5, it also turns out to be mathematically equivalent to many alternative frameworks.

4.1 Setup

We first describe the setup of the framework. We consider a world comprising a compact set \mathcal{N} of locations, which we index with $i, j \in \mathcal{N}$. We assume the world has perfectly competitive markets for both products and labor.

4.1.1 Geography

Each location $i \in \mathcal{N}$ endowed with (a) a technology for producing a unique variety (which we will also index by i); (b) an innate productivity \bar{A}_i for producing that variety; and (c) an innate amenity \bar{u}_i that affects the welfare of agents residing in i . The assumption that each location produces a distinct differentiated variety—proposed by Armington (1969) and applied to trade flows by Anderson (1979)—is clearly an extreme simplification. Nevertheless, the Armington assumption turns out to be both surprisingly tractable and mathematically equivalent to more complicated (but more realistic) assumptions, as we will show below. In addition, each pair of locations $i, j \in \mathcal{N}$ are endowed with a technology for trading goods, which following Samuelson (1954) we assume take the “iceberg” form, where $\tau_{ij} \geq 1$ units of the product must be shipped from $i \in \mathcal{N}$ in order for one unit to arrive in $j \in \mathcal{N}$. Together, the $\{\bar{A}_i\}_{i \in \mathcal{N}}$, $\{\bar{u}_i\}_{i \in \mathcal{N}}$, and $\{\tau_{ij}\}_{i, j \in \mathcal{N}}$ comprise the *geography* of the world.

Our interest is to determine how the geography of the world shapes the equilibrium distribution of economic activity. To do so, we begin by populating the world with a measure \bar{L} of infinitesimal agents. These agents are assumed to be perfectly mobile across locations. Whenever they choose to reside, they produce goods to exchange with other agents and consume goods that they have purchased from other agents. The particular value \bar{L} turns out to not be particularly important in this simple framework: it will not affect the fraction of agents who live in a location or the share of economic activity occurring in each location.

4.1.2 Production

Labor is the only factor of production. Given the assumption of perfect competition, the price of a good produced in $i \in \mathcal{N}$ and consumed in $j \in \mathcal{N}$ is equal to its marginal cost, namely:

$$p_{ij} = \tau_{ij} w_i / A_i, \tag{16}$$

where w_i is the wage per unit of labor and A_i is the total productivity of workers in location $i \in \mathcal{N}$. Note that the (innate) productivity \bar{A}_i and the (total) productivity A_i are distinct objects; we will discuss their relationship below.

4.1.3 Consumption

Consider an agent $l \in \mathcal{L} \equiv [0, \bar{L}]$ that supplies her unit of labor inelastically and chooses (a) where to live and (b) how much to consume of each location-specific differentiated variety in order to maximize her utility subject to her budget constraint by solving:

$$\max_{j \in \mathcal{N}, \{c_{ij}\}_{i \in \mathcal{N}} \geq 0} \left(\sum_{i \in \mathcal{N}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \times u_j \text{ s.t. } \sum_{i \in \mathcal{N}} p_{ij} c_{ij} \leq w_j, \quad (17)$$

where $\sigma \geq 0$ is the elasticity of substitution between varieties from different locations, u_j is the total amenity of residing in j , p_{ij} is the price of a good produced in i and sold in j , and w_j is the wage of an agent residing in location j . This problem is a variation of maximization in equation (7) with the additional location specific amenity term. We introduce this term to add flexibility in the model to fit the distribution of observed population in the data. As with the productivity, note that the (innate) amenity \bar{u}_j and the (total) amenity u_j are distinct objects; we will discuss their relationship below.

Optimization of equation (17) yields two important results. First, it allows us to write the welfare of any individual residing in location $i \in \mathcal{N}$ as an indirect function of her wage, the amenity, and the price index as follows:

$$W_i = w_i u_i / P_i. \quad (18)$$

Second, the optimization yields the following expression for the quantity consumed by agent $l \in \mathcal{L}$ residing in $j \in \mathcal{N}$:

$$c_{ij} = p_{ij}^{-\sigma} P_j^{\sigma-1} w_j, \quad (19)$$

where $P_j^{1-\sigma} = \sum_{i \in \mathcal{N}} p_{ij}^{1-\sigma}$ is the Dixit-Stiglitz price index. Multiplying the quantity consumed given in equation (19) by the price and summing across the measure L_j agents residing in $j \in \mathcal{N}$ yields the total value of goods produced in $i \in \mathcal{N}$ and sold in $j \in \mathcal{N}$, X_{ij} :

$$X_{ij} = p_{ij}^{1-\sigma} P_j^{\sigma-1} w_j L_j. \quad (20)$$

Substituting equation (16) for the prices into the bilateral demand equation (20) yields the following ‘‘gravity’’ expression:

$$X_{ij} = \tau_{ij}^{1-\sigma} (w_i / A_i)^{1-\sigma} P_j^{\sigma-1} w_j L_j. \quad (21)$$

Equation (21) states that, if goods are substitutes (i.e. $\sigma > 1$), then the value of bilateral trade flows will be decreasing in the trade costs between the locations (τ_{ij}) and the cost of production in the origin (w_i/A_i), relative to the trade costs and costs of production in all other origins, as summarized by the Dixit-Stiglitz price index (P_j), and increasing in the total income of the destination (w_jL_j). In doing so, it succinctly summarizes the rich patterns of trade flows observed empirically described in Section 2.3 above.

4.1.4 Agglomeration Forces

We now formally incorporate the agglomeration forces that we discussed in Section 3.2 above in a flexible way. To do so we link the innate productivities and amenities to the total productivities and amenities. Suppose that the total productivity of a location is a function of its (exogenous) innate productivity and its (endogenous) labor population with constant elasticity $\alpha \in \mathbb{R}$:

$$A_i = \bar{A}_i L_i^\alpha. \quad (22)$$

A positive α means that the productivity of a location is increasing with its population density. This could occur for a variety of reasons, e.g. firm entry, knowledge spillovers, better access to labor or inputs, etc. Conversely, a negative α means that the productivity of a location is decreasing with its population density, for example due to a fixed factor of production. Of course, none of these mechanisms are explicitly in this simple framework; instead we interpret α as a “reduced form” way of capturing the effects of such forces have on the prices consumers pay for tradable goods. Below, however, we will show formally that there are multiple micro-foundations incorporating such forces that justify the particular functional form of equation (22) in Section 5 below.

Similarly, suppose that the total amenity of a location is a function of its innate amenity and its labor population with constant elasticity $\beta \in \mathbb{R}$:

$$u_i = \bar{u}_i L_i^\beta. \quad (23)$$

As with the productivity spillovers, β may be positive or negative. A positive value—which indicates that amenities are increasing with population—could arise if e.g. a greater population density enables a location to sustain a greater variety of restaurants, better schools, or better parks. Conversely, a negative value could arise if a greater population density e.g. increases the cost of fixed factors of consumption like housing. As with the productivity spillovers, we interpret β as parameter meant to capture (in a reduced form way) the many possible forces through which the local population affects the amenity of living in a location other than the direct cost of tradable goods. We provide several examples or particular micro-foundations in Section 5 below.

As we mentioned in Section 3.2, an earlier generation of economic geography theory highlighted the important role that agglomeration forces played in determining how the innate geography affected the equilibrium distribution of economic activity. As will become evident below, this is still true in quantitative economic geography frameworks: the particular value and magnitude of α and β play an important role in determining the relationship between the innate geography and the equilibrium distribution of economic activity.

4.2 Equilibrium

Our goal is to determine how geography affects the equilibrium distribution of economic activity across locations. To do so, we first derive the equilibrium conditions that define this mapping.

4.2.1 Definition

We can now discuss the equilibrium of this simple quantitative framework. Formally, for any geography (which recall is the set of innate productivities $\{\bar{A}_i\}_{i \in \mathcal{N}}$, amenities $\{\bar{u}_i\}_{i \in \mathcal{N}}$, and trade costs $\{\tau_{ij}\}_{i \in \mathcal{N}}$), set of model parameters $\sigma \geq 0$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ and $\bar{L} > 0$, a *spatial equilibrium* is the distribution of labor $\{L_i\}_{i \in \mathcal{N}}$, wages $\{w_i\}_{i \in \mathcal{N}}$, and price indices $\{P_i\}_{i \in \mathcal{N}}$ such that the following four conditions hold:

1. *Goods markets clear.* In particular, the total labor income in each location is equal to its total sales to all locations:

$$w_i L_i = \sum_{j \in \mathcal{N}} X_{ij} \quad \forall i \in \mathcal{N}. \quad (24)$$

2. *Budget constraints are satisfied.* In particular, the total expenditure in each location is equal to its total income:

$$w_i L_i = \sum_{j \in \mathcal{N}} X_{ji} \quad \forall i \in \mathcal{N}. \quad (25)$$

3. *Welfare is equalized.* In particular, there exists a scalar $W > 0$ such that:

$$W_i \leq W \quad \forall i \in \mathcal{N}, \quad (26)$$

with equality if $L_i > 0$.

4. *The total population of the world is equal to its endowed amount of labor:*

$$\sum_{i \in \mathcal{N}} L_i = \bar{L}. \quad (27)$$

Note that equilibrium conditions (24) and (25) together imply that trade is balanced in every location. While this is an abstraction from reality (where trade deficits and surpluses are commonplace), the static nature of the framework makes it difficult to incorporate such imbalances, although some prior work has simply imposed exogenous transfers, see e.g. Dekle, Eaton, and Kortum (2008). In Section 8, we briefly discuss recent dynamic extensions to this framework. We also refer the interested reader to Chapter XXX “Spatial Dynamics Klaus Desmet and Fernando Parro” XXX of this volume for an in depth discussion.

In what follows, we focus on an *interior equilibrium* where every location is inhabited, so that the welfare equalization condition (26) implies that $W_i = W$ for all $i \in \mathcal{N}$. We do so for two reasons. First, an interior solution is the empirically relevant one: in an analysis of cities, states, counties, etc., all locations are inhabited. Second, the log linear functional forms in equations (22) and (23) for the agglomeration forces are not particularly realistic when locations are uninhabited, implying for a positive (negative) α or β that an uninhabited location would have zero (infinite) productivity or amenity.

4.2.2 Deriving the equilibrium conditions

Substituting the gravity equation (21) and agglomeration equations (22) and (23) into equilibrium conditions (24) and (25) and collecting like terms implies for all $i \in \mathcal{N}$:

$$w_i^\sigma L_i^{1-\alpha(\sigma-1)} = \sum_{j \in \mathcal{N}} (\bar{A}_i / \tau_{ij})^{\sigma-1} P_j^{\sigma-1} w_j L_j \quad (28)$$

$$P_i^{1-\sigma} = \sum_{j \in \mathcal{N}} (\bar{A}_j / \tau_{ji})^{\sigma-1} w_j^{1-\sigma} L_j^{\alpha(\sigma-1)} \quad (29)$$

Similarly, substituting the indirect utility equation (18) and agglomeration equations (22) and (23) into equilibrium condition (26) and solving for the price index yields for all $i \in \mathcal{N}$:

$$P_i = w_i \bar{u}_i L_i^\beta / W. \quad (30)$$

Notice that this expression can be reformulated to be written as a labor supply curve, a point we will discuss in detail below.

Finally, substituting the welfare equalization equation (30) into the market clearing equation (28) and budget constraint equation (29) yields the following system of equations for all $i \in \mathcal{N}$:

$$W^{\sigma-1} w_i^\sigma L_i^{1-\alpha(\sigma-1)} = \sum_{j \in \mathcal{N}} (\bar{A}_i \bar{u}_j / \tau_{ij})^{\sigma-1} w_j^\sigma L_j^{1+\beta(\sigma-1)} \quad (31)$$

$$W^{\sigma-1} w_i^{1-\sigma} L_i^{\beta(1-\sigma)} = \sum_{j \in \mathcal{N}} (\bar{u}_i \bar{A}_j / \tau_{ji})^{\sigma-1} w_j^{1-\sigma} L_j^{\alpha(\sigma-1)} \quad (32)$$

Equations (31) and (32) define the relationship between the world geography and model pa-

rameters on the one hand and the equilibrium distribution of economic activity on the other. They provide $2N$ equations (two for each location), which along with the aggregate labor market clearing condition (27), is the same number as the $2N + 1$ endogenous outcomes (wages w_i and labor L_i in each location, along with the overall welfare W of the world). This provides hope that this system of equations is sufficient for determining the equilibrium distribution of economic activity. This hope, as we will show below, turns out to be well founded.

The particular mathematical structure of equations (31) and (32) deserves some discussion. Both equations state that a certain log-linear combination of endogenous outcomes in a particular location i , scaled appropriately, is equal to a weighted average of a different log-linear combination of endogenous outcomes in all other locations. The log-linear combinations depend on the model parameters σ , α , and β , whereas the weights depend on the geography $\{A_i\}_{i \in \mathcal{N}}$, $\{u_i\}_{i \in \mathcal{N}}$, and $\{\tau_{ij}\}_{i,j \in \mathcal{N}}$ of the system. That the geographic weights in equation (32) are the transpose of (31) will be important below. It turns out that this particular structure is ubiquitous among quantitative economic geography models, and so in Section 6 we discuss the properties of such a system in detail. Prior to doing so, however, we first consider two special cases of the framework.

4.3 Market access, supply, and demand

We proceed by offering a slightly more intuitive expression for bilateral trade flows and, in doing so, revisit the important concept of *market access*. The following derivations come from Anderson and Van Wincoop (2003) (who referred to market access as “multilateral resistance.”) Substituting gravity equation (21) into the goods market clearing condition (24) and re-arranging yields:

$$(w_i / A_i)^{1-\sigma} = \frac{w_i L_i}{\sum_{j \in \mathcal{N}} \tau_{ij}^{1-\sigma} P_j^{\sigma-1} w_j L_j}. \quad (33)$$

We then proceed by substituting equation (33) back into the gravity equation, yielding:

$$X_{ij} = \tau_{ij}^{1-\sigma} \times \frac{w_i L_i}{\Pi_i^{1-\sigma}} \times \frac{w_j L_j}{P_j^{1-\sigma}}, \quad (34)$$

where $\Pi_i^{1-\sigma} \equiv \sum_{j \in \mathcal{N}} \tau_{ij}^{1-\sigma} P_j^{\sigma-1} w_j L_j$ is the (*outward*) *market access* and $P_j^{1-\sigma}$ was defined in equation (33) above is the (*inward*) *market access*.

The market access terms play a key role in quantitative economic geography frameworks, as they mediate how economic activity elsewhere influences economic activity in a given location. To see this, we can substitute the productivity spillover equation (22) into equation (33), take

logs, and solve for wages to construct an (inverse) labor demand curve:

$$\ln w_i = - \left(\frac{1}{\sigma} - \alpha \left(\frac{\sigma - 1}{\sigma} \right) \right) \ln L_i + \frac{1}{\sigma} \ln \Pi_i^{1-\sigma} + \frac{\sigma - 1}{\sigma} \ln \bar{A}_i. \quad (35)$$

Equation (35) highlights that, conditional on outward market access, the demand curve for labor in location $i \in \mathcal{N}$ is a straightforward log-linear function. As long as α is not too large (i.e. $\alpha < 1/(\sigma - 1)$), then the labor demand curve is downward sloping, and a larger agglomeration force attenuates the downward slope. Greater innate productivity or market access increase labor demand by shifting the demand curve upward.

Similarly, we can take logs of equation (30) and solve for wages to construct an (inverse) labor supply curve:

$$\ln w_i = -\beta \ln L_i + \frac{1}{1-\sigma} \ln P_i^{1-\sigma} + \ln W - \ln \bar{u}_i. \quad (36)$$

Like with the labor demand curve, equation (36) shows that, conditional on inward market access, the labor supply curve is also simple log-linear function. As long as $\beta < 0$ (i.e. there are congestion externalities in consumption), then the labor supply curve is upward sloping, with the strength of the externalities determining the slope. Greater innate amenities or better inward market access shifts the labor supply curve downward, since either implies that workers can maintain the same level of welfare with lower nominal wages.

Note that the only place that economic activity in the rest of the world shows up in the labor demand curve (35) and labor supply curve (36) are in the outward and inward market access terms, respectively. Put another way, the market access terms together fully summarize how locations interact with each other. That means that if we could treat the market access terms as exogenous characteristics of a location, then solving for the equilibrium economic activity would be simple: we would simply equate supply and demand in each location separately and solve for the equilibrium wages and labor. But, of course, it is not so simple, as both the inward and outward market access terms are endogenous variables that depend on both the exogenous geography and the endogenous distribution of economic activity in all locations.

Before we start exploring the role of the endogenous market access terms, it is worth pointing out that we already have a small reason to celebrate. Recall from Section 3.2 that the equilibrium of the core-periphery framework was quite fragile due to the labor supply and demand curves not depending directly on local labor. By modeling the amenity and productivity spillovers, we have solved that problem, with both the labor demand curve (35) and labor supply curve (36) now depending on the local labor while still retaining the important role that market access plays in the core-periphery model. This will prove crucial for the model to remain tractable even in a world with a rich geography, making it an ideal framework for quantitative applications. Indeed, it will provide the basis of the “universal” quantitative economic

geography framework we present in Section 6.

4.4 Special case #1: No agglomeration or congestion externalities

Suppose that there are not agglomeration or congestion externalities in either production or consumption, i.e. $\alpha = \beta = 0$. In this case, the equilibrium conditions (31) and (32) simplify to the following:

$$W^{\sigma-1} w_i^\sigma L_i = \sum_{j \in \mathcal{N}} (\bar{A}_i \bar{u}_j / \tau_{ij})^{\sigma-1} w_j^\sigma L_j \quad (37)$$

$$W^{\sigma-1} w_i^{1-\sigma} = \sum_{j \in \mathcal{N}} (\bar{u}_i \bar{A}_j / \tau_{ji})^{\sigma-1} w_j^{1-\sigma}. \quad (38)$$

In the absence of agglomeration or congestion externalities, the log linear combination of endogenous outcomes is the same on both the left hand side and right hand side of each equation. This allows us to write equations (37) and (38) more succinctly in matrix notation as follows:

$$\lambda \mathbf{x} = \mathbf{T} \mathbf{x} \quad (39)$$

$$\lambda \mathbf{y} = \mathbf{T}' \mathbf{y}, \quad (40)$$

where $\lambda \equiv W^{\sigma-1}$ is an endogenous scalar, \mathbf{x} is an $N \times 1$ vector with i^{th} element $w_i^\sigma L_i$, \mathbf{y} is an $N \times 1$ vector with i^{th} element $w_i^{1-\sigma}$, and \mathbf{T} is an $N \times N$ matrix with $(i, j)^{\text{th}}$ element $(\bar{A}_i \bar{u}_j / \tau_{ij})^{\sigma-1}$. As equations (39) and (40) make clear, the endogenous distribution of economic activity in the absence of agglomeration or congestion externalities is simply determined by the left and right eigenvectors of the matrix \mathbf{T} , which summarizes the geography of the world.

Given the familiar mathematical structure of the equilibrium, a number of interesting properties of the equilibrium follow readily from well known mathematical results.

First, as long as \mathbf{T} is positive in all its elements and because \mathbf{x} and \mathbf{y} must be non-negative (since wages and populations must be weakly positive), by the Perron Frobenius theorem (Perron, 1907; Frobenius, Frobenius, Frobenius, Frobenius, and Mathematician, 1912), there exists unique (to scale) and strictly positive eigenvectors corresponding to the largest (in absolute value) eigenvalue of equations (39) and (40). Given these eigenvectors, the equilibrium wages and populations can be immediately recovered by solving the two system $x_i \propto w_i^\sigma L_i$ and $y_i \propto w_i^{1-\sigma}$ for w_i and L_i , where the proportions are pinned down by a choice of numeraire (for wages) and the aggregate labor market clearing condition equation (27) (for labor).

Second, because it is straightforward to calculate the largest eigenvalue and associated eigenvectors of a matrix (using e.g. a power iteration algorithm, see e.g. Mises and Pollaczek-Geiringer (1929)), determining the unique distribution of economic activity is very easy computationally and can be accomplished in very little time even with a large number of locations

and a rich geography.

Third, because economic activity is greater in locations with larger eigenvectors, if we view the geography matrix \mathbf{T} as a weighted graph of a network between locations, then the concept of eigenvector centrality (see e.g. Jackson et al. (2008)) tells us that economic activity will be more concentrated in locations that are more central to this network. The weights of the graph, in this case, are the elements of the matrix, T_{ij} . Since $T_{ij} = (\bar{A}_i \bar{u}_j / \tau_{ij})^{\sigma-1}$, $\sigma > 1$, more central locations will (loosely speaking) be those with better productivities and amenities and lower trade costs to other locations (and those connected to other locations with similarly advantaged geographies).

Fourth, because world welfare is (log) proportional to the largest eigenvalue of the geography matrix \mathbf{T} , then world welfare will tend to be larger the lower the trade costs and the greater the innate productivities and amenities of locations. We can formalize this result by considering what happens to the largest eigenvalue when we perturb the $(i, j)^{th}$ element of the geography matrix \mathbf{T} by a small amount. From Vahrenkamp (1976) we have that the resulting change in the largest eigenvalue is proportional to the product of the left and right eigenvectors as follows:

$$\frac{\partial \lambda}{\partial T_{ij}} = y_i x_j / \sum_{k \in \mathcal{N}} y_k x_k. \quad (41)$$

Given the relationships of λ, T_{ij}, y_j and x_i to the underlying economic variables, it is straightforward to show that equation (41) implies that the elasticity of aggregate welfare to a reduction in bilateral trade costs between locations i and j is equal to the value of trade flows between those locations, i.e.:

$$-\frac{\partial \ln W}{\partial \ln \tau_{ij}} = X_{ij} / Y^W, \quad (42)$$

where Y^W is the total world income. This result is the equivalent to the “social savings” approach proposed by Fogel (1964) for advocating the welfare effects of changes in transportation costs. Intuitively, in the absence of agglomeration or congestion forces, the competitive equilibrium is efficient, and so the first order effects of technology improvements like the reduction of trade costs on aggregate welfare are proportional to the share of aggregate economic activity on those links, as in Hulten (1978).

To summarize, in the absence of agglomeration and congestion forces, the effect of geography on the spatial distribution of economic activity is well behaved and well understood. But as emphasized in Section 3, there is much interesting economics that comes from incorporating agglomeration and congestion forces. We now make a first pass in trying to incorporate these forces.

4.5 Special case #2: Two locations

We now consider the case of two locations, symmetric trade costs, and no innate amenity differences. This is the case that can be most closely connected to the Core-Periphery model that we presented in Section 3. After some tedious derivations, equations (31) and (32) can be written as the following second order equation ratio of wages, w_1/w_2 :

$$\frac{A_2^{\sigma-1}}{A_1^{\sigma-1}} \left(\frac{w_1^{(\sigma-1)}}{w_2^{(\sigma-1)}} \right)^2 + \left(1 - \frac{A_2^{\sigma-1}}{A_1^{\sigma-1}} \right) \tau^{1-\sigma} \frac{w_1^{\sigma-1}}{w_2^{\sigma-1}} - 1 = 0.$$

The solution of this equation is the positive solution of the quadratic root of this equation, which is,

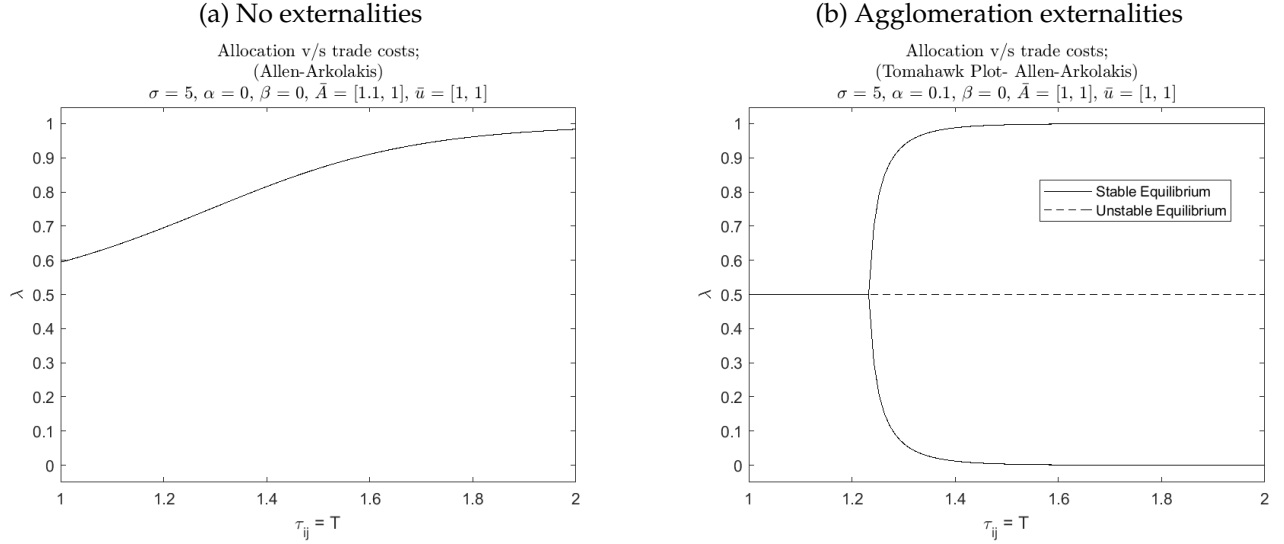
$$\frac{w_1}{w_2} = \frac{1}{2} \left(1 - \frac{A_1^{\sigma-1}}{A_2^{\sigma-1}} \right) \tau^{1-\sigma} + \frac{1}{2} \frac{A_1^{\sigma-1}}{A_2^{\sigma-1}} \sqrt{\left(\frac{A_2^{\sigma-1}}{A_1^{\sigma-1}} - 1 \right)^2 (\tau^{1-\sigma})^2 + 4 \frac{A_2^{\sigma-1}}{A_1^{\sigma-1}}}$$

We assume, without loss of generality that $A_1 > A_2$. The above expression can be shown to imply $w_2/w_1 < 1$, i.e. the more productive region offers higher wages. Notice that this ratio is decreasing in τ so long as $\sigma > 1$, i.e. the wage advantage of location 1 decreases when trade costs increase. This is intuitive as market access decreases and there is a tendency for workers to endogenously move to the location with the better access (see Allen and Arkolakis, 2018).

To solve for the population share we need to resort to simulations. Figure 8 in the left panel plots the share of population on location 1 when location 1 is the most productive one, when $\alpha = \beta = 0$ and the productivity in location 1 is 10% higher than the location 2. Naturally, the share of population choosing location 1 is more than half. In addition, the share of population in location 1 increases with trade costs, i.e. trade costs increase concentration of economic activity. Notice that this is in sharp contrast with predictions of the Core-Periphery framework displayed in Figure 7 where the share of the dominant region decreases with higher trade costs.

It is useful to revisit the prediction of the Core-Periphery model for the share of population as a function of trade costs, the well known Tomahawk graph depicted above in the right panel of Figure 7. To analyze that we plot our model with positive agglomeration externality, $\alpha = 0.1$, and $\beta = 0$ in the right panel of Figure 8. In the quantitative geography model there are multiple equilibria with large trade costs since this is when more concentration takes place. The stable equilibria are ones of concentration, consistent with the previous discussion, while the symmetric equilibrium is now an unstable one for high trade costs. The new figure resembles a “Pitchfork” rather than a “Tomahawk”. A different prediction, but still quite sharp!

Figure 8: The distribution of spatial economic activity in the Allen and Arkolakis (2014) model



Notes: This figure shows the fraction of labor allocated to the more productive region (y-axis) for different values of trade costs (x-axis). The left panel considers an economy without any productivity or amenity externalities; the right panel considers an economy with agglomeration forces arising through productivity externalities.

4.6 Special case #3: Symmetric trade costs

Suppose that trade costs are symmetric, i.e. for all $i \in \mathcal{N}$ and $j \in \mathcal{N}$, we have $\tau_{ij} = \tau_{ji}$. In this special case, it turns out that the inward and outward market access terms are proportional to each other, as noted by Anderson and Van Wincoop (2003) and Donaldson and Hornbeck (2016). (Similar results can also be obtained when trade costs are “quasi-symmetric”, i.e. $\tau_{ij} = \tau_{ji} \tau_i^A \tau_j^B$ for some exogenous vectors $\{\tau_i^A, \tau_i^B\}_{i \in \mathcal{N}}$; see Allen, Arkolakis, and Takahashi (2020)). To see this, we can write the inward and outward market access equations as follows:

$$P_i^{1-\sigma} = \sum_{j \in \mathcal{N}} \tau_{ji}^{1-\sigma} \Pi_j^{\sigma-1} w_j L_j \quad (43)$$

$$\Pi_i^{1-\sigma} = \sum_{j \in \mathcal{N}} \tau_{ij}^{1-\sigma} P_j^{\sigma-1} w_j L_j \quad (44)$$

Suppose that $P_i^{1-\sigma} = \kappa \Pi_i^{1-\sigma}$ for some scalar $\kappa > 0$. Then substituting out $P_i^{1-\sigma}$ in either equation for $\kappa \Pi_i^{1-\sigma}$ yields:

$$\kappa \Pi_i^{1-\sigma} = \sum_{j \in \mathcal{N}} \tau_{ij}^{1-\sigma} \Pi_j^{\sigma-1} w_j L_j, \quad (45)$$

demonstrating that equation (45) and $P_i^{1-\sigma} = \kappa \Pi_i^{1-\sigma}$ is indeed a solution to the system of equations (43) and (44). (It also turns out to be the unique solution, which we show in Section 6.5).

If $P_i^{1-\sigma} = \kappa \Pi_i^{1-\sigma}$ for some $\kappa > 0$, we can derive a simple log-linear relationship between the

population L_i and wage w_i in a location. To do so, combine the welfare equalization equilibrium condition (30) with the outward market access equation (33) using the fact that inward and outward market accesses are proportional. This results in the following expression:

$$L_i^{1+(\beta-\alpha)(\sigma-1)} \propto w_i^{(1-2\sigma)} (\bar{A}_i/\bar{u}_i)^{\sigma-1}, \quad (46)$$

where the proportionality is determined by the aggregate labor market clearing condition (27) and the choice of numeraire. By creating a log linear relationship between the equilibrium population and wages in a location, equation (46) allows us to collapse the $2N$ system of equilibrium equations given by equations (31) and (32) into N that depend only on the equilibrium population in each location:

$$\left(\frac{W}{\bar{L}^{\alpha+\beta}}\right)^{\sigma-1} l_i^{\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma)} = \sum_{j \in \mathcal{N}} \tau_{ij}^{1-\sigma} (\bar{A}_i \bar{u}_j)^{(\sigma-1)\tilde{\sigma}} (\bar{u}_i \bar{A}_j)^{\sigma\tilde{\sigma}} l_j^{\tilde{\sigma}(1+\sigma\alpha+\beta(\sigma-1))}, \quad (47)$$

where $l_i \equiv L_i/\bar{L}$ is the fraction of the total population residing in $i \in \mathcal{N}$ and $\tilde{\sigma} \equiv (\sigma - 1) / (2\sigma - 1)$. There are two things to note about equation (47): first, the aggregate labor endowment \bar{L} has no effect on the distribution of labor (it only affects the interpretation of the endogenous scalar); second, the proportionality factor κ between inward and outward market access does not affect the equilibrium, as it scales both sides of the equation equally. If we can find the equilibrium population in each location that solves equation (47), we can then use equation (46) to recover the equilibrium wages in each location. It turns out that an equilibrium exists and is unique as long as $\left| \frac{1+\sigma\alpha+\beta(\sigma-1)}{1-\alpha(\sigma-1)-\beta\sigma} \right| \leq 1$, i.e. the agglomeration forces are no stronger than the congestion forces. We provide a sketch of these results in the following subsection.

4.6.1 [Technical] Existence and Uniqueness

To find the equilibrium population in each location, we first note that the mathematical structure of equation (47) is similar to the general system of equations (31) and (32), where the population shares l_i in location $i \in \mathcal{N}$ (to some power) are proportional to a weighted sum of the populations in all other locations (to some potentially different power), where the weights depend on the geography and the powers are functions of the model parameters. If the exponents on both sides of the equation are equal, i.e. if $\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma) = \tilde{\sigma}(1+\sigma\alpha+\beta(\sigma-1))$, or equivalently $\alpha + \beta = 0$, then equation (47) is a linear system of equations and the equilibrium population is (log) proportional to the eigenvector of the geography matrix, as in Section 4.4.

Suppose instead that $\alpha + \beta \neq 0$, so that the exponents on the two sides of equation (47) are not equal. Then the system of equations is nonlinear, which requires a new set of tools. We will discuss these tools in depth in Section 6 below, but offer a glimpse here.

Let us first tackle the question of existence. Here we do not restrict ourselves to focusing

only on interior equilibria and we assume $\alpha(\sigma - 1) + \beta\sigma < 1$. (We refer the interested reader to Karlin and Nirenberg (1967) for a more complicated existence proof when $\alpha(\sigma - 1) + \beta\sigma > 1$). Define the function $F : \Delta^N \rightarrow \Delta^N$:

$$F(\mathbf{l})_i \equiv \frac{\left(\sum_{j \in \mathcal{N}} T_{ij} l_j^{\tilde{\sigma}(1+\sigma\alpha+\beta(\sigma-1))}\right)^{\frac{1}{\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma)}}}{\sum_{i \in \mathcal{N}} \left(\sum_{j \in \mathcal{N}} T_{ij} l_j^{\tilde{\sigma}(1+\sigma\alpha+\beta(\sigma-1))}\right)^{\frac{1}{\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma)}}, \quad (48)$$

where we now define $T_{ij} \equiv \tau_{ij}^{1-\sigma} (\bar{A}_i \bar{u}_j)^{(\sigma-1)\tilde{\sigma}} (\bar{u}_i \bar{A}_j)^{\sigma\tilde{\sigma}}$. Equation (48) takes as an input a vector of population shares $\{l_i\}_{i \in \mathcal{N}}$ and returns as an output another vector of population shares. It is continuous (as long as $1 - \alpha(\sigma - 1) - \beta\sigma \neq 0$). A fixed point $\mathbf{l}^* = F(\mathbf{l}^*)$ of equation (48) is a solution to equation (47), where the endogenous scalar is equal to the denominator, i.e. $\left(\frac{W}{\bar{L}^{\alpha+\beta}}\right)^{\sigma-1} = \sum_{i \in \mathcal{N}} \left(\sum_{j \in \mathcal{N}} T_{ij} l_j^{\tilde{\sigma}(1+\sigma\alpha+\beta(\sigma-1))}\right)^{\frac{1}{\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma)}}$. Because F is a continuous function that maps a convex compact space to itself, by the Brouwer's fixed-point theorem, there exists a solution to equation (47) as long as $1 - \alpha(\sigma - 1) - \beta\sigma \neq 0$.

Now let us tackle the question of uniqueness. We begin with a carefully chosen change of variables. (Re-)define $x_i \equiv (W/\bar{L}^{\alpha+\beta})^{-\frac{(\sigma-1)(1-\alpha(\sigma-1)-\beta\sigma)}{(2\sigma-1)(\alpha+\beta)}} l_i^{\tilde{\sigma}(1-\alpha(\sigma-1)-\beta\sigma)}$, so that equation (47) can be written as:

$$x_i = \sum_{j \in \mathcal{N}} T_{ij} x_j^\rho, \quad (49)$$

where $\rho \equiv \frac{1+\sigma\alpha+\beta(\sigma-1)}{1-\alpha(\sigma-1)-\beta\sigma}$. There are two things to note about the change of variables: first, the endogenous variable $\{x_i\}_{i \in \mathcal{N}}$ includes both the population L_i and the global welfare W ; and second, if we can solve for $\{x_i\}_{i \in \mathcal{N}}$ using equation (49), then we can separately recover the population and the welfare using the aggregate labor market clearing condition (27) (and then recover the equilibrium wages using equation (46)).

We now establish sufficient conditions for the uniqueness of a system of equations of the type in equation (47) by providing conditions under which the system is a contraction. The following contradiction argument is loosely based on Karlin and Nirenberg (1967). Define a (new) function $F : \mathbb{R}^N \rightarrow \mathbb{R}^N$:

$$F(\mathbf{y})_i \equiv \log \left(\sum_{j \in \mathcal{N}} T_{ij} \exp(\rho y_j) \right) \quad (50)$$

and a distance metric $d(\mathbf{y}^A, \mathbf{y}^B) \equiv \max_{i \in \mathcal{N}} |y_i^A - y_i^B|$. If there exists a $c < 1$ such that for all $\mathbf{y}^A, \mathbf{y}^B \in \mathbb{R}^N$ that $d(F(\mathbf{y}^A), F(\mathbf{y}^B)) \leq c \times d(\mathbf{y}^A, \mathbf{y}^B)$, then by the Banach fixed-point theorem, there exists a unique vector \mathbf{y}^* such that $\mathbf{y}^* = F(\mathbf{y}^*)$, which can be found by starting at any initial guess $\mathbf{y}_0 \in \mathbb{R}^N$ and iteratively updating the guess using the function, i.e. if $\mathbf{y}_n = F(\mathbf{y}_{n-1})$ then $\lim_{n \rightarrow \infty} \mathbf{y}_n = \mathbf{y}^*$. Note that once the fixed point \mathbf{y}^* to equation (50) has been found, then

the set $x_i^* \equiv \exp(y_i^*)$ will solve equation (47).

Choose any $\mathbf{y}^A, \mathbf{y}^B \in \mathbb{R}^N$. Then we have:

$$d\left(F\left(\mathbf{y}^A\right), F\left(\mathbf{y}^B\right)\right) = \max_{i \in \mathcal{N}} \left| \log \left(\sum_{j \in \mathcal{N}} C_{ij} \exp\left(\rho\left(y_j^A - y_j^B\right)\right) \right) \right|, \quad (51)$$

where $C_{ij} \equiv T_{ij} \exp\left(\rho y_j^B\right) / \sum_k T_{ik} \exp\left(\rho y_k^B\right)$. Because $\sum_{j \in \mathcal{N}} C_{ij} = 1$, we can bound the distance as follows:

$$\max_{i \in \mathcal{N}} \left| \log \left(\sum_{j \in \mathcal{N}} C_{ij} \exp\left(\rho\left(y_j^A - y_j^B\right)\right) \right) \right| \leq |\rho| \max_{i \in \mathcal{N}} \left| y_j^A - y_j^B \right|, \quad (52)$$

or equivalently:

$$d\left(F\left(\mathbf{y}^A\right), F\left(\mathbf{y}^B\right)\right) \leq |\rho| d\left(\mathbf{y}^A, \mathbf{y}^B\right). \quad (53)$$

As a result, if $|\rho| < 1$, then there exists a unique interior equilibrium.

What does it mean that $|\rho| < 1$? Recall from above that $\rho \equiv \frac{1+\sigma\alpha+\beta(\sigma-1)}{1-\alpha(\sigma-1)-\beta\sigma}$. Focusing on the case where $\sigma \geq 1$ and $\alpha, \beta \in [-1, 1]$, this condition will hold as long as $\alpha + \beta < 0$, i.e. the externalities are net congestive rather than net agglomerative. Loosely speaking, there is a unique equilibrium as long as individuals would prefer to spread out across locations.

What happens if $|\rho| > 1$? First, from above, there will still exist interior equilibria (as long as $1 - \alpha(\sigma - 1) - \beta\sigma \neq 0$); however, if the ρ becomes too large (so that $1 - \alpha(\sigma - 1) - \beta\sigma < 0$), then a “black hole” equilibrium where all agents reside in a single location becomes possible (as noted by Fujita, Krugman, and Venables (1999)) and any interior equilibria becomes unstable (in the sense that welfare would increase if individuals all moved to the same location). Second, there will exist certain geographies for which there are guaranteed to be multiple interior equilibria; in the words of Allen, Arkolakis, and Li (2023), the condition is sufficient and “globally” necessary. For example, Allen and Arkolakis (2014) show that when $\rho > 1$, there are an infinite number of equilibria featuring greater economic activity near an arbitrary chosen location when identical locations are arrayed around a ring. But what remains an open question in the field is how the particular geography of the world determines whether or not there will be multiplicity if $|\rho| > 1$. Recent analysis by Kucheryavy, Lyn, and Rodríguez-Clare (2021) of a two location model, for example, has shown that uniqueness can be maintained as long as trade costs are sufficiently low.

5 New models, same old equilibrium

The simple quantitative economic geography model presented in the previous section relies on a number of strong assumptions (e.g. perfect labor mobility, identical agents, perfect competition, etc.). But it turns out to be remarkably general. This is because a number of interesting

departures from these strong assumptions are mathematically equivalent to the mathematical structure developed above. This is because these departures act identically to the productivity and/or amenity externalities in the baseline model. We now briefly describe seven alternative models that are isomorphic (and an eighth that is not). The goal is to understand better how different economic assumptions relate to agglomeration or dispersion forces present in the quantitative economic geography model, thereby affecting the labor supply and demand curves determining the spatial equilibrium.

5.1 Ricardian comparative advantage

In the simple model in Section 4, trade arises because consumers have a love of variety and each location produces a differentiated variety. However, one can instead develop a mathematically equivalent model with perfect competition and Ricardian comparative advantage. The key insight—developed by Eaton and Kortum (2002)—is to assume there is a continuum of goods sold competitively by each region. Each good is provided by the least cost supplying country in each destination. If productivities of these products ω follow a Frechet distribution, i.e. $\varepsilon_i(\omega) \sim e^{-A_i z^{-\theta}}$, then Eaton and Kortum (2002) show that expenditure share of country j on goods produced in i is given by:

$$\lambda_{ij} = \frac{\left(\frac{w_i \tau_{ij}}{A_i}\right)^{-\theta}}{\sum_i \left(\frac{w_i \tau_{ij}}{A_i}\right)^{-\theta}}. \quad (54)$$

Equation (54) is isomorphic to the gravity equation (21) with $\theta = \sigma - 1$. As a result, the Armington justification for trade in the baseline model presented in Section 4 and this Ricardian trade model is the interpretation of the trade elasticity.

5.2 Non-tradable sector

In the simple model in Section 4, all goods produced in the economy are tradable. But in reality, some goods cannot be (easily) traded. To do so, we follow Helpman et al. (1995). The key idea is that the workers spend a constant fraction of their income δ on the differentiated good and the rest to a locally produced good that is not traded with the rest of the locations, such as services or housing. The non-tradable sector is perfectly competitive and its returns are equally distributed across workers that reside in each location. This change affects only the labor supply of the economy. The non-tradable sector is akin to introducing a congestion amenity externality and is equivalent to our baseline model so long as $\beta = -(1 - \delta) / \delta$, i.e. the degree of congestion in consumption equals the negative of the ratio of non-tradeable to tradeable good shares.

5.3 Fixed factors of production

In the simple model in Section 4, labor is the only factor of production. To incorporate factors of production such as land and capital, we follow Donaldson and Hornbeck (2016). The key idea is that the production function features an additional fixed factor, e.g capital, $Y_i = A_i K_i^{\tilde{\alpha}} L_i^{1-\tilde{\alpha}}$, and $\tilde{\alpha}$ is the share of the fixed factor. Under that configuration the new model is isomorphic to the baseline model presented in the previous section if $\tilde{\alpha} = -\alpha$, i.e. the fixed factor acts as a negative productivity externality.

5.4 Endogenous labor supply

In the simple model in Section 4, agents inelastically supply the entirety of their labor endowment. To incorporate an endogenous labor supply decision, suppose agents have preferences over both their consumption and their labor supply as follows:

$$U(C_i, L_i) = C_i u_i - \frac{(L_i)^{1+\eta}}{1+\eta}, \quad (55)$$

where the first term is the CES aggregator as specified in equation (17), with budget constraint:

$$\sum_{i \in \mathcal{N}} p_{ij} c_{ij} = P_j C_j = w_j L_j. \quad (56)$$

Agents now decide how many labor units to devote in work. Straightforward application of first order conditions of the constrained maximization problem of equations (55) and (56) yields a labor supply function $L_i = (w_i u_i / P_i)^{1/\eta}$. Expressing this equation in the form of equation (36), we obtain an exact equivalence if we set $-\beta = \eta$, i.e. an endogenous labor supply acts isomorphically to a congestion externality on amenities.

5.5 Idiosyncratic preferences

Rather than assuming that all agents have identical preferences across locations, we can extend the model to allow agents to have heterogeneous preferences across locations, as in Redding (2016). Further discussion of this approach and its use in urban settings can be found in Ahlfeldt, Redding, Sturm, and Wolf (2015) and are discussed in detail in Chapter XXX Steve's chapter XXX of this handbook.

The key idea of incorporating heterogeneity in agent's preferences is to assume that the welfare of each worker ω choosing to reside in location j is $\frac{w_j}{P_j} u_j \times \varepsilon_j(\omega)$, where $\varepsilon_j(\omega)$ is assumed to be an idiosyncratic preference term that is agent-specific. If we assume that distribution is Frechet, then following Eaton and Kortum (2002), it is straightforward to show that the share

of workers that choose i as their location is

$$L_i/\bar{L} = \frac{(w_i u_i / P_i)^\theta}{\sum_j (w_j u_j / P_j)^\theta}. \quad (57)$$

Equation (57) is isomorphic to the labor supply equation (36) in the baseline model with the strength of the amenity externality set to $\beta = -1/\theta$. Intuitively, the more people who reside in a location, the lower on average their idiosyncratic preferences are for that location, which acts in a mathematically equivalent way to a congestion externality in amenities.

5.6 Idiosyncratic productivities

Rather than assuming that all agents are equally productive in all locations, we can also extend the model to allow agents to be heterogeneous in their productivities, as in Bryan and Morten (2019). The approach works in a similar way as incorporating heterogeneous preferences, where we assume that the welfare of each worker ω choosing to reside in location j is $\frac{w_j}{P_j} u_j \times \varepsilon_j(\omega)$, but we now interpret $\varepsilon_j(\omega)$ as the idiosyncratic productivity of agent ω in location j . In this case, the number of workers choosing a location still follows equation (57). However, because the workers who choose a location will tend to be more productive in that location, the effective units of labor in location i , \tilde{L}_i , are:

$$\tilde{L}_i = L_i^{\frac{\theta-1}{\theta}}. \quad (58)$$

As it is the effective units of labor in a location that determines the amount of goods produced and the payments to labor, we can write the supply and demand curves in terms of effective units of labor. Using equation (58), it is straightforward to show that these labor and supply curves are equivalent to the baseline model when $\beta = -1/(\theta - 1)$, albeit with a modified aggregate effective labor market clearing constraint $\sum_{i \in \mathcal{N}} \tilde{L}_i^{\frac{\theta}{\theta-1}} = \bar{L}^{\frac{\theta-1}{\theta}}$. Intuitively, heterogeneity in productivities, like heterogeneity in preferences, acts as a congestion externality in amenities.

5.7 Monopolistic competition with free entry

In the simple model in Section 4, we assumed perfect competition. However, that model is also isomorphic to a model featuring monopolistic competition and free entry, as in the Krugman (1991) model discussed in Section 3.2. Comparing the labor supply and demand equations (14) and (15) from Section 3.2 to the labor supply and demand equations (35) and (36) from Section 4, we see that the former is isomorphic to the latter when $\alpha = 1/(\sigma - 1)$, i.e. monopolistic competition with free entry acts equivalently to a productivity agglomeration force.

5.8 Round-about production

The previous models emphasize the flexibility of the simple model in Section 4 in relaxing a variety of assumptions. We conclude with an example of an extension that is *not* isomorphic. In the simple model, we abstracted from intermediate goods production. Following Krugman and Venables (1995); Eaton and Kortum (2002), suppose we incorporate intermediate goods in the following simple way: we assume that goods are used for final consumption but also as intermediate inputs in the production of other goods. In particular, we assume that firms produce their products using a Cobb-Douglas combination of labor and intermediate goods with share γ going to labor and share $1 - \gamma$ going to intermediate goods, which we further assume are aggregated in the same way consumers aggregate their final goods.

Working through the derivations of Section 4.2.2 in this alternative framework yields the following system of equations that define the equilibrium:

$$W^{\gamma(\sigma-1)} w_i^\sigma L_i = \sum_j \tau_{ij}^{1-\sigma} A_i^{\sigma-1} \left(u_i^{1-\gamma} \right)^{1-\sigma} u_j^{\sigma-1} w_j^\sigma L_j \quad (59)$$

$$W^{\gamma(\sigma-1)} w_i^{1-\sigma} = \sum_j \tau_{ji}^{1-\sigma} A_j^{\sigma-1} u_i^{\sigma-1} u_j^{(1-\gamma)^{1-\sigma}} w_j^{1-\sigma}. \quad (60)$$

A comparison of equations (59) and (60) to the equilibrium conditions (31) and (32) of the baseline model shows two differences. First, the elasticity of welfare to the endogenous scalar changes due to the presence of roundabout production. This change is to be expected, as the presence of intermediate inputs magnifies the welfare gains from trade (see for example Arkolakis, Costinot, and Rodríguez-Clare (2012); Ossa (2015)). The second difference is a bit more surprising: now the amenities in both locations affect both equilibrium conditions. Intuitively, this is because round-about production, combined with welfare equalization, introduces a new way through which amenities elsewhere affect the price index in a location. Hence, a formal isomorphism to the baseline model only occurs in the special case where amenities are identical in all locations.

6 A “universal” quantitative economic geography framework

We now present a “universal” quantitative economic geography framework, special cases of which include the models presented above.

6.1 Setup

As in Section 4, we consider a world comprising a compact set \mathcal{N} of locations and inhabited by workers who consume and produce in the location in which they reside.

There are two components of the framework. The first component comprises the spatial linkages between locations. We assume that locations are linked together through the flow of goods. Consistent with the evidence presented in Section 2.2 and 2.3, we assume that the value of trade flow from $i \in \mathcal{N}$ to $j \in \mathcal{N}$ follow a gravity equation:

$$X_{ij} = T_{ij} \times \frac{Y_i}{MA_i^{out}} \times \frac{E_j}{MA_j^{in}}, \quad (61)$$

where $T_{ij} \leq 1$ is an (exogenous) trade friction, Y_i is the aggregate value of economic output in $i \in \mathcal{N}$, E_j is the aggregate value of economic consumption in $j \in \mathcal{N}$, MA_i^{out} is the outward market access in $i \in \mathcal{N}$ and MA_j^{in} is the inward market access in $j \in \mathcal{N}$.

The second component comprises the labor markets in each location. We assume that in each location $i \in \mathcal{N}$, the (inverse) labor supply curve can be written as the following log-linear function:

$$\ln w_i = \varepsilon_{local}^S \ln L_i - \varepsilon_{global}^S \ln MA_i^{in} - \ln C_i^S - \ln \phi^S, \quad (62)$$

where ε_{local}^S is the elasticity of wages to the labor supplied, ε_{global}^S is the elasticity of wages to the inward market access, C_i^S is an (exogenous) labor supply factor, and ϕ^S is an endogenous scalar common to all locations to ensure an aggregate labor market condition (or aggregate welfare normalization) is satisfied. Note that our sign convention is such that each parameter would typically be positive, e.g. the supply curve is normally upward sloping and better inward market access or higher C_i^S would normally imply that labor would be willing to accept a lower wage to work in a location.

Similarly, we assume that in each location $i \in \mathcal{N}$, the (inverse) labor demand curve can be written as:

$$\ln w_i = -\varepsilon_{local}^D \ln L_i + \varepsilon_{global}^D \ln MA_i^{out} + \ln C_i^D + \ln \phi^D, \quad (63)$$

where ε_{local}^D is the elasticity of wages to labor demanded, ε_{global}^D is the elasticity of wages to the outward market access, C_i^D is an (exogenous) labor demand factor, and ϕ^D is an endogenous scalar common to all locations to ensure the choice of numeraire is satisfied. Again, we maintain a sign convention that each parameter should typically be positive.

It is immediately evident that the Rosen (1979)-Roback (1982) model of location choice presented in Section (3.1), the Krugman (1991) core-periphery model presented in Section (3.2) and the Allen and Arkolakis (2014) quantitative economic geography model presented in Section 4 are all special cases of this “universal” quantitative economic geography framework. Table 1 shows how the parameters in each of these models (and others) map to the local and global supply and demand elasticities in the universal framework.

A couple of notes are necessary. First, note that when the global elasticities are absent, i.e. $\varepsilon_{global}^S = \varepsilon_{local}^D = 0$ (as in Rosen (1979)-Roback (1982) model), the equilibrium is trivial

Table 1: Mappings of models to the universal quantitative economic geography framework

Paper	ϵ_{local}^S	ϵ_{global}^S	ϵ_{local}^D	ϵ_{global}^D	Model parameters	Notes
Rosen (1979) and Roback (1982)	$-\beta$	0	α	0	β : Externality of population on utility α : Capital share in production	Based on the version developed in Section 3.1.
Krugman (1991)	0	$\frac{1}{\sigma-1}$	0	$\frac{1}{\sigma}$	σ : Elasticity of substitution	Based on the extended version developed in Section 3.2, assuming no agricultural sector.
Helpman et al. (1995)	$\frac{1-\mu}{\mu}$	$\frac{1}{\sigma-1}$	0	$\frac{1}{\sigma}$	σ : Elasticity of substitution μ : Expenditure share of manufacturing goods	Based on the multi-regional version developed by Redding and Sturm (2008)
Eaton and Kortum (2002)	0	$\frac{1}{\theta}$	$\frac{1}{1+\theta}$	$\frac{1}{1+\theta}$	θ : Fréchet shape parameter	We differ from Eaton and Kortum (2002) by assuming (1) no round-about production, (2) no non-manufacturing sector, and (3) assuming perfect labor mobility across regions.
Allen and Arkolakis (2014)	$-\beta$	$\frac{1}{\sigma-1}$	$\frac{1-\alpha(\sigma-1)}{\sigma}$	$\frac{1}{\sigma}$	σ : Elasticity of substitution α : Externality of population on productivity β : Externality of population on utility	
Donaldson and Hornbeck (2016)	0	$\frac{1}{\theta}$	$\frac{1+\alpha\theta}{1+(\alpha+\gamma)\theta}$	$\frac{1}{1+(\alpha+\gamma)\theta}$	θ : Fréchet shape parameter α : Expenditure share of land γ : Expenditure share of labor	
Redding (2016) (with constant return to scale production)	$\frac{1+(1-\alpha)\epsilon}{\alpha\epsilon}$	$\frac{1}{\theta}$	$\frac{1}{1+\theta}$	$\frac{1}{1+\theta}$	θ : Fréchet shape parameter of productivity draw α : Expenditure share of manufacturing goods ϵ : Fréchet shape parameter of preference draw	Redding (2016) develops two models, the first with Eaton-Kortum style constant return to scale technology, and the second with Krugman-style increasing return to scale technology.
Redding (2016) (with increasing return to scale production)	$\frac{1+(1-\alpha)\epsilon}{\alpha\epsilon}$	$\frac{1}{\sigma-1}$	0	$\frac{1}{\sigma}$	σ : Elasticity of substitution α : Expenditure share of manufacturing goods ϵ : Fréchet shape parameter of preference draw	
Idiosyncratic productivity a la Bryan and Morten (2019)	$\frac{1-\beta\theta}{\theta-1}$	0	$\frac{1-\alpha(\sigma-1)}{\theta}$	0	θ : $\frac{\theta}{1-\rho}$, $\tilde{\theta}$ the shape parameter and ρ the correlation parameter of the multi-variate Fréchet distribution σ : Elasticity of substitution α : Externality of population on productivity β : Externality of population on utility	We differ from Bryan and Morten (2019) by assuming there is no migration cost.

Notes: This table shows how the universal gravity framework comprises a number of notable economic geography models.

to solve, so in what follows we focus on the more interesting case where they are non zero. Second, note that unlike the simple model from Section 4, here we do not need to distinguish between the iceberg trade costs τ_{ij} and the trade elasticity $\sigma - 1$: for the purposes of determining how the equilibrium distribution of economic activity is affected by a change in the underlying geography, only their composite $T_{ij} = \tau_{ij}^{1-\sigma}$ is what matters. This is not to say, however, that the trade elasticity does not matter for counterfactual analysis. Indeed, it matters in two ways: first, it affects the model elasticities ($\varepsilon_{local}^S, \varepsilon_{global}^S, \varepsilon_{local}^D, \varepsilon_{global}^D$); second, it affects the mapping from the endogenous scalars to agent welfare.

6.2 Equilibrium

Analogously to Section 4, we define the *geography* of the world as the set $\left\{ \{T_{ij}\}_{i,j \in \mathcal{N}}, \{C_i^S, C_i^D\}_{i \in \mathcal{N}} \right\}$, the *model elasticities* as the set $\left\{ \varepsilon_{local}^S, \varepsilon_{global}^S, \varepsilon_{local}^D, \varepsilon_{global}^D \right\}$, and the *distribution of economic activity* as the set $\left\{ \{L_i, w_i, MA_i^{in}, MA_i^{out}\}_{i \in \mathcal{N}}, \phi^S, \phi^D \right\}$.

Given any geography and model parameters, an equilibrium is a distribution of economic activity such that the following four conditions hold:

1. Goods markets clear, i.e. $Y_i = \sum_{j \in \mathcal{N}} X_{ij}$ and $E_i = \sum_{j \in \mathcal{N}} X_{ji}$ for all $i \in \mathcal{N}$.
2. Labor markets clear, i.e. labor supply (defined by equation (62)) and labor demand (defined by equation (63)) are equalized for all $i \in \mathcal{N}$.
3. Output and expenditure is equal to the total payments to labor, i.e. $w_i L_i = Y_i = E_i$ for all $i \in \mathcal{N}$.
4. The endogenous scalars ϕ^S and ϕ^D are determined by a choice of numeraire and a labor market clearing condition, i.e. $\sum_{i \in \mathcal{N}} L_i = \bar{L}$.

Combining the equilibrium condition 1 with the gravity equation (61) for bilateral trade flows yields the following system of equations relating inward and outward market access to economic activity in all locations:

$$MA_i^{out} = \sum_{j \in \mathcal{N}} T_{ij} \times \frac{E_j}{MA_j^{in}} \quad (64)$$

$$MA_i^{in} = \sum_{j \in \mathcal{N}} T_{ji} \times \frac{Y_j}{MA_j^{out}}, \quad (65)$$

which is the “universal” form of equations (43) and (44) in Section 4.6.

In the following three subsections, we summarize the state of knowledge about the equilibrium properties of this framework. If you would prefer to skip them, the punchline is as follows: (1) under very mild conditions, an equilibrium will exist; (2) as long as agglomeration

forces are not “too” strong, the equilibrium will be unique; and (3) these results hold for any number of locations and any geography (although we still do not know if more precise conditions can be given for uniqueness that depend on the particular geography being considered).

6.3 [Technical] Re-defining the equilibrium system

We can then solve for the inward market access in the labor supply equation (62) and the outward market access in the labor demand equation (63), impose equilibrium condition (3), and substitute into the market access equations (64) and (65) to yield:

$$\tilde{\lambda} w_i \frac{1}{\varepsilon_{global}^D} L_i^{\frac{\varepsilon_{local}^D}{\varepsilon_{global}^D}} = \sum_{j \in \mathcal{N}} K_{ij} w_j \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} L_j^{\frac{\varepsilon_{global}^S - \varepsilon_{local}^S}{\varepsilon_{global}^S}} \quad (66)$$

$$-\frac{1}{\varepsilon_{global}^S} L_i^{\frac{\varepsilon_{local}^S}{\varepsilon_{global}^S}} = \sum_{j \in \mathcal{N}} K_{ji} w_j \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} L_j^{\frac{\varepsilon_{global}^D - \varepsilon_{local}^D}{\varepsilon_{global}^D}} \quad (67)$$

where $\tilde{\lambda} \equiv (\phi^D)^{-1/\varepsilon_{global}^D} (\phi^S)^{-1/\varepsilon_{global}^S}$ is an endogenous scalar and $K_{ij} \equiv T_{ij} (C_i^D)^{1/\varepsilon_{global}^D} (C_j^S)^{1/\varepsilon_{global}^S}$ is a function of the geography and the model parameters. It is not surprising that equations (66) and (67) bear a strong resemblance to the equilibrium system of equations (31) and (32) for the simple quantitative economic geography framework presented in Section 4, as that simple framework is a special case of this more general framework. Both systems have a similar mathematical structure, where a particular log linear combination of wages and population in one location are equal to a weighted sum of a different log linear combination of wages and population in all other location. In both systems, the weights of the sum depend on the geography of the world, whereas the exponents depend on the labor supply and demand elasticities. In the simple quantitative economic geography framework presented in Section 4, these supply and demand elasticities were constructed from particular assumptions regarding consumer preferences and production and amenity externalities. In contrast, the “universal” framework considers the supply and demand elasticities directly.

As in Section (4.6), it turns out to be helpful to rewrite the equilibrium system in terms of shares rather than levels. Define $y_i \equiv w_i L_i / Y^W$ as the share of world income in location $i \in \mathcal{N}$ and recall that $l_i \equiv l_i / \bar{L}$ is the population share. We can then rewrite equations (66) and (67) as

follows:

$$\lambda y_i \frac{1}{\varepsilon_{global}^D} l_i^{\frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D}} = \sum_{j \in \mathcal{N}} K_{ij} y_j \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} l_j^{-\frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S}} \quad (68)$$

$$-\frac{1}{\varepsilon_{global}^S} l_i^{\frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S}} = \sum_{j \in \mathcal{N}} K_{ji} y_j \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} l_j^{\frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D}}, \quad (69)$$

where $\lambda \equiv \tilde{\lambda} \frac{\frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S}}{\frac{1}{Y^W} \frac{1}{\varepsilon_{global}^S} - \frac{1}{\varepsilon_{global}^D}}$. This re-definition of the equilibrium in terms of output and population shares emphasizes that the aggregate scale of the economy (in either its nominal terms or in its aggregate labor endowment) has no effect on the equilibrium distribution of economic activity in the economy, as changes in either \bar{L} or Y^W simply change the endogenous scalar. In what follows, we view the endogenous scalar λ as a measure of the welfare of the equilibrium, although how exactly it maps into welfare will depend on the details of the micro-foundation being considered. For example, in the simple framework presented in Section 4, λ is log proportional to the welfare of the system, with the proportionality equal to the trade elasticity $(\sigma - 1)$.

6.4 [Technical] Existence

We first consider the question of whether or not there exists a solution to the system of equations defined by equations (68) and (69). Define a (new) function $F(\mathbf{y}, \mathbf{l}) : \Delta^N + \Delta^N \rightarrow \Delta^N + \Delta^N$ which maps from any combination of output shares and labor shares into a new set of output

shares and labor shares as follows. Let $F_1(\mathbf{y}, \mathbf{l})_i \equiv \sum_{j \in \mathcal{N}} K_{ij} y_j \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} l_i^{-\frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S}}$ and $F_2(\mathbf{y}, \mathbf{l})_i \equiv \sum_{j \in \mathcal{N}} K_{ji} y_j \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} l_j^{\frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D}}$ for all $i \in \mathcal{N}$. Then define:

$$F(\mathbf{y}, \mathbf{l})_i = \begin{cases} F_1(\mathbf{y}, \mathbf{l})_i^{b_{11}} F_2(\mathbf{y}, \mathbf{l})_i^{b_{12}} / \sum_{i \in \mathcal{N}} F_1(\mathbf{y}, \mathbf{l})_i^{b_{11}} F_2(\mathbf{y}, \mathbf{l})_i^{b_{12}} & \text{for } i \in \{1, \dots, N\} \\ F_1(\mathbf{y}, \mathbf{l})_i^{b_{21}} F_2(\mathbf{y}, \mathbf{l})_i^{b_{22}} / \sum_{i \in \mathcal{N}} F_1(\mathbf{y}, \mathbf{l})_i^{b_{21}} F_2(\mathbf{y}, \mathbf{l})_i^{b_{22}} & \text{for } i \in \{N + 1, \dots, 2N\} \end{cases}, \quad (70)$$

where b_{kl} is the $(k, l)^{th}$ element of the 2×2 matrix $\mathbf{B} \equiv \begin{pmatrix} 1/\varepsilon_{global}^D & \frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} \\ -\frac{1}{\varepsilon_{global}^S} & \frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S} \end{pmatrix}^{-1}$. Intuitively, what the function F does is take any set of output and labor shares as an input and return as an output the implied output and labor shares using equations (68) and (69). As a result, a

fixed point of $F(\mathbf{y}^*, \mathbf{l}^*) = F(\mathbf{y}^*, \mathbf{l}^*)$ is a set of equilibrium output and labor shares that solves equations (68) and (69). As in Section (4.6), as long as \mathbf{B} exists (which allows us to invert the left hand side of equations (68) and (69) to recover the implied output and labor shares), then because F is a continuous operator mapping a convex compact space to itself, Brouwer's fixed point theorem ensures that there exists an equilibrium. We refer the interested reader to Allen, Arkolakis, and Takahashi (2020) for details, including demonstrating that the equilibrium is an interior one.

6.5 [Technical] Uniqueness

We now provide sufficient conditions for the uniqueness of the equilibrium. We provide a sketch of the proof below and refer the interested reader to Allen, Arkolakis, and Li (2023) for

details. We begin with a change of variables. Define $x_i \equiv y_i^{1/\varepsilon_{global}^D} l_i^{\frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D}}$ and $z_i \equiv y_i^{-\frac{1}{\varepsilon_{global}^S}} l_i^{\frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S}}$ so that equations (68) and (69) can be written as:

$$\lambda x_i = \sum_{j \in \mathcal{N}} K_{ij} x_j^{a_{11}} z_j^{a_{12}} \quad (71)$$

$$\lambda z_i = \sum_{j \in \mathcal{N}} K_{ji} x_j^{a_{21}} z_j^{a_{22}}, \quad (72)$$

where $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \equiv \begin{pmatrix} \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} & -\frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} \\ \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} & \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} \end{pmatrix} \begin{pmatrix} \frac{1}{\varepsilon_{global}^D} & \frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} \\ -\frac{1}{\varepsilon_{global}^S} & \frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S} \end{pmatrix}^{-1}$ is a matrix of model elasticities. In Section 4.6, we showed that in a non-linear system of equations featuring a single equation for each location, a sufficient condition for uniqueness was that the exponent on the right hand side endogenous variable was less than or equal to one in magnitude. It turns out that a similar sufficient condition for uniqueness can be derived in this universal framework featuring two equations for each location. Let $|\mathbf{A}| \equiv \begin{pmatrix} |a_{11}| & |a_{12}| \\ |a_{21}| & |a_{22}| \end{pmatrix}$ be the absolute value of the matrix of the model elasticities. If the spectral radius of $|\mathbf{A}|$ (i.e. if the largest eigenvalue in terms of magnitude) is less than or equal to one, i.e. $\rho(|\mathbf{A}|) \leq 1$, then the equilibrium is unique.

To see this, we proceed by contradiction. Suppose that $\rho(|\mathbf{A}|) \leq 1$ and that there are distinct solutions $\{\lambda^A, \{x_i^A, z_i^A\}_{i \in \mathcal{N}}\}$ and $\{\lambda^B, \{x_i^B, z_i^B\}_{i \in \mathcal{N}}\}$ that both solve the system of equations (68) and (69). Let a hatted variable indicate the ratio of the two solutions, e.g. $\hat{x}_i \equiv x_i^B / x_i^A$, so

that we can write equations (68) and (69) as follows:

$$\hat{\lambda} \hat{x}_i = \sum_{j \in \mathcal{N}} C_{ij} \hat{x}_j^{a_{11}} \hat{z}_j^{a_{12}} \quad (73)$$

$$\hat{\lambda} \hat{z}_i = \sum_{j \in \mathcal{N}} D_{ij} \hat{x}_j^{a_{21}} \hat{z}_j^{a_{22}}, \quad (74)$$

where $C_{ij} \equiv \frac{K_{ij} (x_j^A)^{a_{11}} (z_j^A)^{a_{12}}}{\sum_{k \in \mathcal{N}} K_{ik} (x_k^A)^{a_{11}} (z_k^A)^{a_{12}}}$ and $D_{ij} \equiv \frac{K_{ji} (x_j^A)^{a_{21}} (z_j^A)^{a_{22}}}{\sum_{k \in \mathcal{N}} K_{ki} (x_k^A)^{a_{21}} (z_k^A)^{a_{22}}}$. Define $\mu_1 \equiv \frac{\max_{i \in \mathcal{N}} \hat{x}_i}{\min_{i \in \mathcal{N}} \hat{x}_i}$ and $\mu_2 \equiv \frac{\max_{i \in \mathcal{N}} \hat{z}_i}{\min_{i \in \mathcal{N}} \hat{z}_i}$. Because $\sum_{j \in \mathcal{N}} C_{ij} = \sum_{j \in \mathcal{N}} D_{ij} = 1$, we can use equations (73) and (74) to construct the following bounds:

$$\mu_1 \leq \mu_1^{|a_{11}|} \mu_1^{|a_{12}|} \quad (75)$$

$$\mu_2 \leq \mu_1^{|a_{21}|} \mu_2^{|a_{22}|}, \quad (76)$$

with the inequality strict for at least one of the two equations because the two solutions are assumed to be distinct. In matrix notation, this becomes:

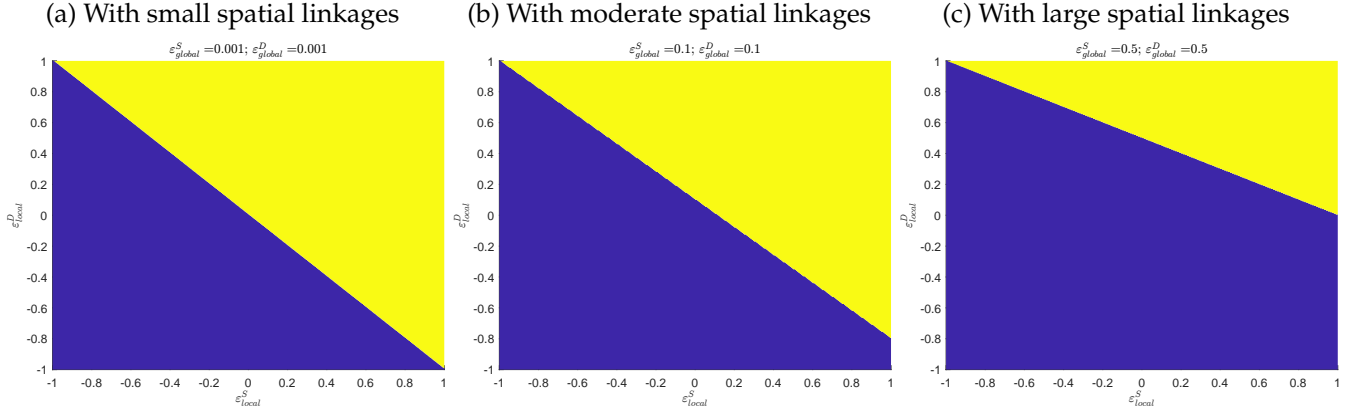
$$\begin{pmatrix} \ln \mu_1 \\ \ln \mu_2 \end{pmatrix} \leq |\mathbf{A}| \begin{pmatrix} \ln \mu_1 \\ \ln \mu_2 \end{pmatrix}, \quad (77)$$

with at least one of the two inequalities strict. Given their definitions, $\ln \mu_1 \geq 0$ and $\ln \mu_2 \geq 0$ with at least one inequality strict again because the two solutions are distinct. According to the Collatz-Wielandt Formula (i.e. that $\rho(\mathbf{A}) = \max_{\ln \mu \in \mathbb{R}_+^2, \ln \mu \neq 0} \min_{h \in \{1,2\}, \ln \mu_h \neq 0} \frac{(|\mathbf{A}| \ln \mu)_h}{\ln \mu_h}$), equation (77) then implies that $\rho(|\mathbf{A}|) > 1$, a contradiction. Hence there is a unique solution.

Note that the argument above generalizes readily to non-linear systems of equations with any number of equations in each location, as shown in Allen, Arkolakis, and Li (2023). Allen, Arkolakis, and Li (2023) also show that a similar argument can be applied to non log-linear systems of equations by considering a matrix of the bounds of the elasticities of the system. Note too that this result establishes the claim made in Section 4.6 that when trade costs are symmetric, the unique inward and outward market access terms are those that are equal up to scale.

What does the condition that $\rho(|\mathbf{A}|) \leq 1$ mean in economic terms? Panel (a) of Figure 9 depicts the uniqueness range in the simple case where the spatial linkages between locations are very small (e.g. ε_{global}^S and ε_{global}^D are close to zero). In this case, a sufficient condition for uniqueness is that $\varepsilon_{local}^S + \varepsilon_{local}^D \geq 0$. This condition holds if the supply curve is upward sloping ($\varepsilon_{local}^S > 0$) and the demand curve is downward sloping ($\varepsilon_{local}^D > 0$) but also holds if the supply curve is upward sloping and the demand curve is not “too” upward sloping (or vice versa). As the spatial linkages get stronger (as illustrated in Panels (b) and (c)), the range of acceptable

Figure 9: Uniqueness of the equilibrium



Notes: This figure displays the range of values for the local labor supply elasticity ε_{local}^S and demand elasticity ε_{local}^D for which a unique equilibrium is assured. Yellow indicates uniqueness.

supply and demand elasticities for which uniqueness is guaranteed shrinks.

It is important to emphasize that the condition $\rho(|\mathbf{A}|) \leq 1$ is a sufficient condition for uniqueness. Although Allen, Arkolakis, and Li (2023) show that there will always exist geographies for which there are multiple equilibria if the condition is not satisfied (i.e. the condition is “globally” necessary), given a particular geography, the equilibrium may still be unique even if $\rho(|\mathbf{A}|) > 1$. It remains an outstanding question how in these cases the particular geography interacts with the model elasticities to determine the number of possible equilibria.

6.6 Counterfactual analysis

We now turn to the question of how to use this quantitative economic geography framework to conduct counterfactual analyses.

In what follows, we will assume that the geography of the world was initially given by the $N \times N$ matrix $\mathbf{K}^A = [K_{ij}^A]$ and we are interested in how the equilibrium distribution of economic activity would change if the geography changed to a new $N \times N$ matrix $\mathbf{K}^B = [K_{ij}^B]$.

We proceed using the “exact hat” algebra introduced by Dekle, Eaton, and Kortum (2008) for its use in quantitative trade models and subsequently popularized by Costinot and Rodríguez-Clare (2014) (although similar techniques were present in Karlin and Nirenberg (1967)). We assume that the initial economy with geography \mathbf{K}^A is in equilibrium A with equilibrium distribution of economic activity $\{l_i^A, y_i^A\}_{i \in \mathcal{N}}$ and subsequently with geography \mathbf{K}^B would be in equilibrium B with equilibrium distribution of economic activity $\{l_i^B, y_i^B\}_{i \in \mathcal{N}}$. As in Section 6.5, we let a hatted variable indicate the ratio of the two equilibria, e.g. $\hat{x}_i \equiv x_i^B / x_i^A$. We can then

re-write the equilibrium system of equations (68) and (69) in changes as follows:

$$\hat{\lambda} \hat{y}_i \frac{1}{\varepsilon_{global}^D} \frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} = \sum_{j \in \mathcal{N}} C_{ij} \hat{K}_{ij} \hat{y}_j \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} - \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} \hat{l}_j \quad (78)$$

$$-\frac{1}{\varepsilon_{global}^S} \frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S} \hat{l}_i = \sum_{j \in \mathcal{N}} D_{ji} \hat{K}_{ji} \hat{y}_j \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} \hat{l}_j, \quad (79)$$

where we (now) define $C_{ij} \equiv \frac{K_{ij}^A (y_j^A) \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} (l_j^A) \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S}}{\sum_{k \in \mathcal{N}} K_{ik}^A (y_k^A) \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} (l_k^A) \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S}}$ and $D_{ji} \equiv \frac{K_{ji}^A (y_j^A) \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} (l_j^A) \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D}}{\sum_{k \in \mathcal{N}} K_{ki}^A (y_k^A) \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} (l_k^A) \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D}}$.

The key insight of Dekle, Eaton, and Kortum (2008) (and certainly not present in Karlin and Nirenberg (1967)) is that C_{ij} and D_{ji} are observable in the data. Indeed, it is straightforward to show that $C_{ij} = \frac{X_{ij}^A}{Y_i^A}$ are the initial export shares of i to j and $D_{ji} = \frac{X_{ji}^A}{E_i^A}$ are the initial import shares of i from j , allowing us to write equations (78) and (79) as follows:

$$\hat{\lambda} \hat{y}_i \frac{1}{\varepsilon_{global}^D} \frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} = \sum_{j \in \mathcal{N}} \left(\frac{X_{ij}^A}{Y_i^A} \right) \hat{K}_{ij} \hat{y}_j \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} - \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} \hat{l}_j \quad (80)$$

$$-\frac{1}{\varepsilon_{global}^S} \frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S} \hat{l}_i = \sum_{j \in \mathcal{N}} \left(\frac{X_{ji}^A}{E_i^A} \right) \hat{K}_{ji} \hat{y}_j \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} \hat{l}_j. \quad (81)$$

There are three important things to note about equations (80) and (81). First, they can be used to calculate how any potential change in geography $\{\hat{K}_{ij}\}_{i,j \in \mathcal{N}}$ affects the entire distribution of economic activity while requiring only knowledge of the initial equilibrium trade flows and model elasticities. Second, because the counterfactual equations (80) and (81) share an identical mathematical structure to the equilibrium equations (68) and (69), they inherit all the mathematical properties of that system. In particular, if the sufficient conditions for uniqueness are satisfied for the initial equilibrium, then any counterfactual being considered will also be unique. Third, we can also use equations (80) and (81) to perform normative analysis by examining how the endogenous scalar changes from equilibrium A to equilibrium B , i.e. $\hat{\lambda}$. To solve for $\hat{\lambda}$, we use the fact that output and population shares sum to one to derive the constraints $\sum_{i \in \mathcal{N}} y_i^A \hat{y}_i = \sum_{i \in \mathcal{N}} l_i^A \hat{l}_i = 1$.

In the next (technical) subsection, we turn to understanding the comparative statics of the quantitative economic geography framework. Since this subsection shows that it is possible to calculate how any change in the underlying geography affects the equilibrium spatial distribution of economic activity, it is not surprising that it is possible to calculate the impacts of

infinitesimal changes. However, considering small changes to the geography highlights the mechanisms in the model. In particular, we highlight how counterfactuals can be interpreted as shocks propagating through the trading network with the observed trade data reflecting the strength of network ties.

6.7 [Technical] Comparative statics

Section 6.6 showed how to calculate the change in the equilibrium distribution of economic activity for any change in the geography of the world. By focusing on small changes in geography, we can derive new insights into the economic mechanisms at work.

To begin, we follow Allen, Arkolakis, and Takahashi (2020), Kleinman, Liu, and Redding (2020), Adao, Arkolakis, and Esposito, 2019, and Kleinman, Liu, and Redding (2024) by log differentiating the equilibrium system of equations (68) and (69) for some small change $\{d \ln K_{ij}\}_{i,j \in \mathcal{N}}$ to yield:

$$d \ln \lambda + \frac{1}{\varepsilon_{global}^D} d \ln \mathbf{y}_i - \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} d \ln l_i = \sum_{j \in \mathcal{N}} \left(\frac{X_{ij}^A}{Y_i^A} \right) \left(d \ln K_{ij} + \frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} d \ln y_j - \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} d \ln l_j \right) \quad (82)$$

$$d \ln \lambda - \frac{1}{\varepsilon_{global}^S} d \ln \mathbf{y}_i + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} d \ln l_i = \sum_{j \in \mathcal{N}} \left(\frac{X_{ji}^A}{E_i^A} \right) \left(d \ln K_{ji} + \frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} d \ln y_j + \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} d \ln l_j \right), \quad (83)$$

where the fact that output and population shares sum to one additionally implies $\sum_{i \in \mathcal{N}} y_i^A d \ln y_i = 0$ and $\sum_{i \in \mathcal{N}} l_i^A d \ln l_i = 0$. We can re-write equations (82) and (83) in matrix notation as follows:

$$d \ln \lambda + \frac{1}{\varepsilon_{global}^D} d \ln \mathbf{y} - \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} d \ln \mathbf{l} = \text{diag} \left(\mathbf{T} d \ln \mathbf{K}^T \right) + \mathbf{T} \left(\frac{\varepsilon_{global}^S + 1}{\varepsilon_{global}^S} d \ln \mathbf{y} - \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} d \ln \mathbf{l} \right) \quad (84)$$

$$d \ln \lambda - \frac{1}{\varepsilon_{global}^S} d \ln \mathbf{y} + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} d \ln \mathbf{l} = \text{diag} \left(\mathbf{S}^T d \ln \mathbf{K} \right) + \mathbf{S}^T \left(\frac{\varepsilon_{global}^D - 1}{\varepsilon_{global}^D} d \ln \mathbf{y} - \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} d \ln \mathbf{l} \right), \quad (85)$$

with adding up constraints $\mathbf{y}^T d \ln \mathbf{y} = 0$ and $\mathbf{l}^T d \ln \mathbf{l} = 0$, where $d \ln \mathbf{y} = [y_i]$ and $d \ln \mathbf{l} = [l_i]$ are $N \times 1$ vectors of the respective changes in output and population shares, $d \ln \mathbf{K} = [d \ln K_{ij}]$ is an $N \times N$ matrix of the changes in geography, $\mathbf{T} = [X_{ij}^A / Y_i^A]$ is an $N \times N$ matrix of export shares in the initial equilibrium, and $\mathbf{S} \equiv [X_{ij}^A / E_j^A]$ is an $N \times N$ matrix of import shares in the initial equilibrium.

To proceed, let us consider a change of variables equivalent to the one used when proving uniqueness in Section 6.5. Define:

$$d \ln \mathbf{x} \equiv \frac{1}{\varepsilon_{global}^D} d \ln \mathbf{y} - \frac{1 - \varepsilon_{local}^D}{\varepsilon_{global}^D} d \ln \mathbf{l} \quad (86)$$

$$d \ln \mathbf{z} \equiv -\frac{1}{\varepsilon_{global}^S} d \ln \mathbf{y} + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{global}^S} d \ln \mathbf{l} \quad (87)$$

We can then re-write equations (84) and (85) as the following $2N$ system of equations:

$$\begin{pmatrix} d \ln \mathbf{x} \\ d \ln \mathbf{z} \end{pmatrix} + d \ln \lambda = \begin{pmatrix} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) \\ \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) \end{pmatrix} + \begin{pmatrix} a_{11} \mathbf{T} & a_{12} \mathbf{T} \\ a_{21} \mathbf{S}^T & a_{22} \mathbf{S}^T \end{pmatrix} \begin{pmatrix} d \ln \mathbf{x} \\ d \ln \mathbf{z} \end{pmatrix}, \quad (88)$$

where the 2×2 matrix \mathbf{A} was defined in Section 6.5. Re-writing the adding up constraints in terms of the new variables yields:

$$\mathbf{y}^T (b_{11} d \ln \mathbf{x} + b_{12} d \ln \mathbf{z}) = 0 \quad (89)$$

$$\mathbf{l}^T (b_{21} d \ln \mathbf{x} + b_{22} d \ln \mathbf{z}) = 0, \quad (90)$$

where $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{\varepsilon_{global}^D} & \frac{\varepsilon_{local}^D - 1}{\varepsilon_{global}^D} \\ -\frac{1}{\varepsilon_{global}^S} & \frac{\varepsilon_{local}^S + 1}{\varepsilon_{global}^S} \end{pmatrix}^{-1}$. Applying either constraint to equation (88)

allows us to solve for the change in the endogenous scalar as follows:

$$d \ln \lambda = \frac{\mathbf{y}^T (b_{11} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) + b_{11} \mathbf{T} (a_{11} d \ln \mathbf{x} + a_{22} d \ln \mathbf{z}) + b_{12} \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) + b_{12} \mathbf{S}^T (a_{21} d \ln \mathbf{x} + a_{22} d \ln \mathbf{z}))}{b_{11} + b_{12}} \quad (91)$$

$$d \ln \lambda = \frac{\mathbf{l}^T (b_{21} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) + b_{21} \mathbf{T} (a_{11} d \ln \mathbf{x} + a_{22} d \ln \mathbf{z}) + b_{22} \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) + b_{22} \mathbf{S}^T (a_{21} d \ln \mathbf{x} + a_{22} d \ln \mathbf{z}))}{b_{21} + b_{22}}, \quad (92)$$

which in turn results (finally) in the following linear system of equations how the distribution of economic activity responds to any change in the underlying geography:

$$\begin{pmatrix} d \ln \mathbf{x} \\ d \ln \mathbf{z} \end{pmatrix} = \left(\mathbf{I} - (\mathbf{I} - \mathbf{Q}) \begin{pmatrix} a_{11} \mathbf{T} & a_{12} \mathbf{T} \\ a_{21} \mathbf{S}^T & a_{22} \mathbf{S}^T \end{pmatrix} \right)^{-1} (\mathbf{I} - \mathbf{Q}) \begin{pmatrix} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) \\ \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) \end{pmatrix}, \quad (93)$$

where $\mathbf{Q} \equiv \begin{pmatrix} \frac{b_{11}}{b_{11} + b_{12}} \mathbf{y}^T & \frac{b_{12}}{b_{11} + b_{12}} \mathbf{y}^T \\ \frac{b_{21}}{b_{21} + b_{22}} \mathbf{l}^T & \frac{b_{22}}{b_{21} + b_{22}} \mathbf{l}^T \end{pmatrix}$ and \mathbf{y}^T and \mathbf{l}^T are $N \times N$ matrices with the output and population shares $1 \times N$ vectors, respectively, stacked across N rows. The \mathbf{Q} matrix ensures the underlying output and population shares sum to one. Given $d \ln \mathbf{x}$ and $d \ln \mathbf{z}$, the distribution of economic activity can be immediately recovered by inverting equations (86) and (87).

There are three things to note about equation (93). First, with the system of equations equa-

tions (80) and (81) governing how arbitrary changes in geography affect the equilibrium spatial distribution of economic activity, equation (93) shows that for any small change in the geography, the resulting change in the distribution of economic activity can be calculating knowing only the model elasticities, the observed trade flows (which allow one to calculate the income shares \mathbf{T} and expenditure shares \mathbf{S}), and the observed distribution of economic activity $\{y_i, l_i\}_{i \in \mathcal{N}}$.

Second, equation (93) provides some intuition about where the sufficient condition for uniqueness $\rho(|\mathbf{A}|) \leq 1$ comes from. Note that if $\rho(|\mathbf{A}|) \leq 1$, we can follow Allen, Arkolakis, and Takahashi (2020) and Kleinman, Liu, and Redding (2020) and consider a Neumann series expansion of equation (93):

$$\begin{pmatrix} d \ln x \\ d \ln z \end{pmatrix} = \sum_{k \geq 0} \left((\mathbf{I} - \mathbf{Q}) \begin{pmatrix} a_{11} \mathbf{T} & a_{12} \mathbf{T} \\ a_{21} \mathbf{S}^T & a_{22} \mathbf{S}^T \end{pmatrix} \right)^k (\mathbf{I} - \mathbf{Q}) \begin{pmatrix} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) \\ \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) \end{pmatrix}. \quad (94)$$

We can consider each element of this infinite sum as the degree of its effect. For example, the zeroth degree (direct) effect (corresponding to $k = 0$) is $(\mathbf{I} - \mathbf{Q}) \begin{pmatrix} \text{diag}(\mathbf{T} d \ln \mathbf{K}^T) \\ \text{diag}(\mathbf{S}^T d \ln \mathbf{K}) \end{pmatrix}$, i.e. how a change in geography affects each location directly through export shares and import shares, holding constant the economic activity in all other locations. The first degree effect (corresponding to $k = 1$) then captures how the direct effect of the geography shock indirectly affects a location through its trading partners via the matrix $(\mathbf{I} - \mathbf{Q}) \begin{pmatrix} a_{11} \mathbf{T} & a_{12} \mathbf{T} \\ a_{21} \mathbf{S}^T & a_{22} \mathbf{S}^T \end{pmatrix}$. And in general, the k^{th} degree effect captures how a location is affected by the $k - 1^{\text{th}}$ degree effect on its trading partners. As long as $\rho(|\mathbf{A}|) \leq 1$, the indirect effects will get smaller and smaller as we consider higher and higher degree effects. But if $\rho(|\mathbf{A}|) > 1$, then it is possible that small shocks to the geography can become increasingly amplified as they percolate throughout the economy. As we mentioned in Section (6.5), an unresolved question is how the underlying geography affects whether or not there are multiple equilibria when $\rho(|\mathbf{A}|) > 1$. It is possible that further analysis of the properties of equation (93) will help shed light on this question.

The third thing to note about equation (93) is that, when combined with equations (91) and/or (92), it allows us to efficiently calculate the first order welfare changes from any small changes in the underlying geography. This can assist in the search for globally optimal spatial policies, which is an active and exciting area of inquiry, see e.g. the recent work of Fajgelbaum and Gaubert (2020) and Donald, FUKU, and Miyauchi (2023). We refer the interested reader to Chapter XXX Cecile & Pablo XXX for more details on the efficiency properties of spatial models.

7 Bringing the model to the data

One of the major advantages of the quantitative economic geography framework is its tight link with empirical data, a point we now discuss in detail.

Suppose you are interested in assessing how a particular policy changes the geography of the world and, in particular, how it affects the spatial distribution of economic activity and the equilibrium welfare of agents. This policy may have already occurred, in which case your analysis would be conducted *ex post*, or it may be a policy that may occur in the future, in which case your analysis would be conducted *ex ante*. In either case, the quantitative economic geography framework is a powerful tool for such an analysis.

Recall from Section 6.6 that the quantitative economic geography framework can be used to assess how any change to the underlying geography affects the equilibrium spatial distribution of economic activity and the associated welfare using equations (80) and (81). These equations also highlight the three necessary inputs for such analysis: (1) observed spatial data, including trade flows $\{X_{ij}\}_{i,j \in \mathcal{N}'}$, income $\{Y_i\}_{i \in \mathcal{N}'}$, expenditure $\{E_i\}_{i \in \mathcal{N}'}$, and the distribution of population $\{L_i\}_{i \in \mathcal{N}'}$; (2) knowledge of how the policy or event being studied affects the underlying geography $\{\hat{K}_{ij}\}_{i,j \in \mathcal{N}'}$; and (3) the model parameters $\{\varepsilon_{local}^S, \varepsilon_{global}^S, \varepsilon_{local}^D, \varepsilon_{global}^D\}$. As highlighted in Section 2, these spatial data are readily available for many countries, so in what follows, we will focus on how to determine how a policy affects the underlying geography (#2) and the estimation of the model parameters (#3).

7.1 Estimating the effect of spatial policies on the geography

Recall from Section 6 that $K_{ij} \equiv T_{ij} (C_i^D)^{1/\varepsilon_{global}^D} (C_j^S)^{1/\varepsilon_{global}^S}$, i.e. there are three components of geography that a potential policy may affect: the bilateral trade frictions $\{T_{ij}\}_{i,j \in \mathcal{N}'}$, an exogenous factor that affects the demand for labor in a location $\{C_i^D\}_{i \in \mathcal{N}'}$, and an exogenous factor that affects the supply for a labor in a location $\{C_j^S\}_{j \in \mathcal{N}'}$. There are many examples of policies that affect bilateral trade frictions, e.g. infrastructure improvements to highways (as in Allen and Arkolakis (2014) and Jaworski and Kitchens (2019)), railroads (as in Donaldson (2018) and Donaldson and Hornbeck (2016)), or subways (as in Severen (2023)); we refer the interested reader to Chapter XXX Dave's Chapter XXX for a much more thorough analysis. Here, we begin by positing that bilateral trade frictions are some function f —parameterized by a vector β —of a vector of observables \mathbf{z} , i.e. $T_{ij} = f_{ij}(\mathbf{z}; \beta)$, so that we can write the gravity equation (61) as follows:

$$X_{ij} = f_{ij}(\mathbf{z}; \beta) \times \frac{Y_i}{MA_i^{out}} \times \frac{E_j}{MA_j^{in}}. \quad (95)$$

Following Silva and Tenreyro (2006), equation (95) can be estimated by assuming a particular functional for $f(\cdot)$ (oftentimes log-linear) and then using a Poisson psuedo-maximum-

likelihood (PPML) estimator to estimate:

$$X_{ij} = \exp(\log f_{ij}(\mathbf{z}; \boldsymbol{\beta}) + \gamma_i + \delta_j) \epsilon_{ij}, \quad (96)$$

where the origin fixed effect $\gamma_i \equiv \ln \frac{Y_i}{MA_i^{out}}$ and destination fixed effect $\delta_j \equiv \ln \frac{E_j}{MA_j^{in}}$ and ϵ_{ij} is an error term. Because Y_i and E_i are observed, an advantage of this approach is that we recover the inward and outward market access terms from the estimated origin and destination fixed effects. Moreover, as Fally (2015) shows, the estimated inward and outward market access terms from the PPML estimation will satisfy the system of equations (64) and (65) for the observed incomes and expenditures and estimated bilateral trade frictions. As a result, this procedure yields estimated trade frictions $\tilde{T}_{ij} \equiv f_{ij}(\mathbf{z}; \tilde{\boldsymbol{\beta}})$ and market access measures $\left\{ \widetilde{MA}_i^{out}, \widetilde{MA}_i^{in} \right\}_{i \in \mathcal{N}}$, where we use a tilde to indicate a variable has been estimated.

An alternative way of recovering the market access given estimated trade frictions is to directly rely on the system of equations (64) and (65). Given income and expenditure, the results from Section 6.5 imply that for any set of trade frictions, there exists unique (to-scale) inward and outward market accesses consistent with equations (64) and (65) (as it is straightforward to show that the spectral radius of the absolute value of the elasticities of the system is one). This alternative procedure is helpful in situations where bilateral trade flows are not observed for all $i, j \in \mathcal{N}$ pairs. For example, this may occur when locations are U.S. counties, but trade frictions are estimated based on state-to-state trade flows.

7.1.1 Algorithms for calculating changes in travel times

Once the function $f(\mathbf{z}; \tilde{\boldsymbol{\beta}})$ is estimated, it is in principle straightforward to determine how a policy affects trade frictions: simply determine how the observables \mathbf{z} change in response to the policy and calculate the resulting change in the trade frictions. In practice, however, this process can be a bit more involved. A common assumption is that trade frictions are a (log linear) function of the travel time (or cost) along the fastest (or least costly) route from an origin to a destination. Under this assumption, \mathbf{z} is a high dimensional set of observables including the time (or cost) of travel on every part of the entire infrastructure network. A policy that, say, builds a new link on the infrastructure network could affect the trade frictions between many origins and destinations simultaneously. While this seems like it would be quite difficult to calculate, there fortunately exist efficient algorithms for determining the fastest travel time (or lowest costs). Loosely speaking, the Dijkstra algorithm (Dijkstra, 1959)—and its continuous space generalization, the Fast Marching Method (Tsitsiklis, 1995; Sethian, 1996)—search outward from a given origin, keeping track of the least cost route so far, so that a single iteration of the algorithm calculates the least cost route to all possible destinations. Such procedures can be employed to calculate how large scale infrastructure improvements affect trade frictions

between all origins and destinations. For example, Donaldson (2018) applies the Dijkstra algorithm to study the impact of the Indian railroad expansion, and Allen and Arkolakis (2014) apply the Fast Marching Method to study the impact of the U.S. interstate highway system.

7.1.2 An analytical solution for changes in travel times

One limitation with the Dijkstra and Fast Marching Method algorithms is that their black box nature limits the ability to analytically assess how improvements to the transportation infrastructure network affect the spatial distribution of economic activity. An alternative method proposed by Allen and Arkolakis (2022) explicitly derives the the least cost route between any two locations along the transportation network. The basic idea is as follows. Suppose that there is a unit mass of perfectly competitive transportation firms $\nu \in [0, 1]$ trying to ship goods from origin $i \in \mathcal{N}$ to destination $j \in \mathcal{N}$ over a transportation network summarized by the $N \times N$ matrix $\mathbf{t} \equiv [t_{kl}]$, where t_{kl} is the (iceberg) trade cost incurred from moving a good directly along a link between k and l . Each transportation firm ν will then choose a route—i.e. a sequence of links beginning in i and ending in j —in order to minimize the total transportation costs incurred from i to j :

$$r_{ij}(\nu) = \min_{r \in \mathcal{R}_{ij}^K, K \geq 1} \left(\prod_{k=1}^K t_{k,k+1} \right) \times \varepsilon_{ij,r}(\nu), \quad (97)$$

where \mathcal{R}_{ij}^K is the set of all possible routes from i to j of length K and $\varepsilon_{ij,r}(\nu)$ is an idiosyncratic preference of route r for the transportation firm ν , which we assume is Frechet distributed with shape parameter θ . Note that as $\theta \rightarrow \infty$, all transportation firms will choose the lowest cost route, so the solution to equation (97) will coincide exactly with the output of the Dijkstra algorithm.

Let us define $\tau_{ij} \equiv \mathbb{E} [r_{ij}(\nu)]$ as the expected total iceberg trade cost from i to j across all transportation firms. The Frechet distributional assumption on the idiosyncratic term allows us to derive an analytical expression for τ_{ij} by explicitly enumerating all possible routes from i to j :

$$\tau_{ij}^{-\theta} = \left(\sum_{K=0}^{\infty} \sum_{k_1=1}^N \sum_{k_2=1}^N \dots \sum_{k_{K-1}=1}^N \sum_{k_K=1}^N (t_{i,k_1} \times t_{k_1,k_2} \times \dots \times t_{k_{K-1},k_K} \times t_{k_K,j})^{-\theta} \right), \quad (98)$$

where we note that pairs of locations that are not directly connected in the transportation network will have infinitely high direct iceberg costs, thereby not affecting the sum in equation (98). We can write the expression for the expected transportation cost more succinctly in matrix notation as a geometric sum of the $N \times N$ adjacency matrix $\mathbf{A} \equiv [t_{ij}^{-\theta}]$:

$$\tau_{ij}^{-\theta} = \sum_{K=0}^{\infty} \mathbf{A}_{ij}^K. \quad (99)$$

As long as the spectral radius of \mathbf{A} is less than one, this geometric sum will be equal to its Leontief inverse $\mathbf{B} = [b_{ij}] \equiv (\mathbf{I} - \mathbf{A})^{-1}$, so that:

$$\tau_{ij}^{-\theta} = b_{ij}. \quad (100)$$

Finally, if we assume that the shape parameter governing the idiosyncratic component of the transportation firms is equal to the trade elasticity (justified, for example, by the setup presented in Allen and Arkolakis (2022) where the producers receive productivity draws over product-route pairs), then the trade friction T_{ij} from equation (61) is equal to the Leontief inverse, i.e. $T_{ij} = b_{ij}$. This, in turn, allows us to write equilibrium equations (64) and (65) for market access directly as functions of the infrastructure network directly as follows:

$$MA_i^{out} - \sum_{j \in \mathcal{N}} t_{ij}^{-\theta} MA_j^{out} = \frac{E_i}{MA_i^{in}} \quad (101)$$

$$MA_i^{in} - \sum_{j \in \mathcal{N}} t_{ji}^{-\theta} MA_j^{in} = \frac{Y_i}{MA_i^{out}}. \quad (102)$$

Equations (101) and (102) highlight that market access of a location can be written recursively as only dependent on the market accesses of locations to which it is directly linked. When combined with the labor supply and demand equations (62) and (63), one can then directly calculate how a spatial policy which changes the transport costs along certain segments of the transportation network affects the equilibrium market access in all locations without the need to calculate the change in bilateral trade frictions. Moreover, despite the difference in its mathematical structure, Allen and Arkolakis (2022) shows that one can use similar mathematical arguments to those sketched above to establish conditions for the existence and uniqueness of an equilibrium where the transportation network (rather than the bilateral trade frictions) comprise the underlying geography, even in the presence of traffic congestion externalities (in which case the equilibrium distribution of economic activity will depend on the aggregate labor endowment). It also turns out that incorporating optimal routing into a quantitative economic geography framework is quite flexible; for example, recent work by Fan, Lu, and Luo (2023) and Fuchs and Wong (2022) has extended this optimal routing setup to incorporate multiple modes of transit and flexible transshipment costs.

7.2 Estimation of model elasticities

We now turn to the estimation of the market elasticities. We offer two alternative estimation procedures.

7.2.1 Estimation of the labor supply and demand

The basic idea is straightforward: we simultaneously estimate the labor supply and demand equations (62) and (63), which we write here in terms of the observed output Y_i and population L_i :

$$\ln Y_i = \left(1 + \varepsilon_{local}^S\right) \ln L_i - \varepsilon_{global}^S \ln MA_i^{in} - \ln C_i^S - \ln \phi^S \quad (103)$$

$$\ln Y_i = \left(1 - \varepsilon_{local}^D\right) \ln L_i + \varepsilon_{global}^D \ln MA_i^{out} + \ln C_i^D + \ln \phi^D. \quad (104)$$

Equations (103) and (104) look like regressions you could run (and indeed, that is the direction we are going), but with several major issues. The first issue is that the inward and outward market access terms $\ln MA_i^{in}$ and $\ln MA_i^{out}$ are not directly observed in the data. Recall from Section 7.1 that we can estimate inward and outward market access from either a gravity regression or from the inversion of equations (64) and (65), yielding estimates $\ln \widehat{MA}_i^{in}$ and $\ln \widehat{MA}_i^{out}$, which can be used as proxies for the true market accesses.

The second issue is a one of simultaneity: both the inward and outward market accesses and the equilibrium population will be correlated with the supply and demand shifters, so it would be inappropriate to use ordinary least squares to estimate equations (103) and (104). Instead, we need to construct instruments for the market access and population distribution in each regression. Valid instruments would have to be correlated with the population and market access but uncorrelated with unobserved supply shifters in equation (103) and unobserved demand shifters in equation (104).

What are potentially valid instruments? The standard strategy employed in the estimation of supply and demand curves is to estimate the labor supply equation (103) by using an observable that shifts labor demand but that is uncorrelated with labor supply as an instrument. Similarly, one could estimate the labor demand equation (104) by using an observable that shifts labor supply but that is uncorrelated with labor demand as an instrument. The difficulty, of course, is identifying such observables. For example, Allen and Donaldson (2018), following Glaeser and Gottlieb (2009), propose using climate variation as an instrument for estimating labor demand equation (104), arguing that the advent of air conditioning shifted labor supply in locations with hot climates by improving the amenities there but did not directly affect labor demand. They also propose using crop suitability as an instrument for estimating labor supply equation (103), arguing that changes in international demand for certain crops were unlikely to be correlated with local labor supply shocks.

One difficulty with this approach is that estimation of either equation (103) or (104) requires two instruments: one for the population of a location and one for its inward or outward market access. An option for constructing two instruments is as follows. Suppose you are estimating the labor supply equation (103) and have an instrument $\ln z_i$ that is uncorrelated with the

supply shifter $\ln C_i^S$. If the instrument is also uncorrelated with the supply shifter in all other locations, then you can construct a second instrument as a trade-friction weighted sum of the instrument in all other locations, i.e. $\ln MA_i^{in,Z} \equiv \ln \sum_{j \in \mathcal{N}} \tilde{T}_{ji} z_j$. Intuitively, the local variation in the instrument will be used to identify the local labor supply elasticity ε_{local}^S whereas the trade-friction weighted variation in the instrument elsewhere will be used to identify the global labor supply elasticity ε_{global}^S .

Once the model elasticities have been estimated, one can recover the supply and demand shifters $\{C_i^S, C_i^D\}$ (to-scale) directly from equations (103) and (104). Intuitively, the supply and demand shifters act as residuals so that the equilibrium distribution of economic activity in the theory matches exactly the observed distribution of economic activity.

7.2.2 The market access approach

A second approach is to use an existing policy change as an instrument. Suppose that you are conducting ex-post evaluation of a policy which affects trade frictions but does not affect the supply or demand shifters. For example, Ahlfeldt, Redding, Sturm, and Wolf (2015) analyze the impact of the construction and destruction of the Berlin Wall on the spatial distribution of economic activity in Berlin and identify model elasticities under the assumption that the distance of a location to the wall is uncorrelated with changes in its amenities or productivities. By equating equations (103) and (104), we can solve for L_i and Y_i as functions of inward and outward market accesses and the supply and demand shifters, as in the “market access” approach of Donaldson and Hornbeck (2016). Writing these expressions in first differences to compare variables before and after the policy was implemented yields:

$$\Delta \ln L_i = \frac{\varepsilon_{global}^S}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln MA_i^{in} + \frac{\varepsilon_{global}^D}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln MA_i^{out} + \Delta \ln \varepsilon_i^L \quad (105)$$

$$\Delta \ln Y_i = \frac{(1 - \varepsilon_{local}^D) \varepsilon_{global}^S}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln MA_i^{in} + \frac{(1 + \varepsilon_{local}^S) \varepsilon_{global}^D}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln MA_i^{out} + \Delta \ln \varepsilon_i^Y, \quad (106)$$

where $\Delta \ln \varepsilon_i^L \equiv \frac{1}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln C_i^S + \frac{1}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln C_i^D + \frac{1}{\varepsilon_{local}^S + \varepsilon_{local}^D} \ln \phi^D + \frac{1}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln \phi^S$ and $\Delta \ln \varepsilon_i^Y \equiv \frac{1 - \varepsilon_{local}^D}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln C_i^S + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln C_i^D + \frac{1 - \varepsilon_{local}^D}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln \phi^S + \frac{1 + \varepsilon_{local}^S}{\varepsilon_{local}^S + \varepsilon_{local}^D} \Delta \ln \phi^D$ are residuals that depend on the changes in both the supply and demand shifters. One can instrument for the changes in inward and outward market access with exposure to the policy in equations (105) and (106), which are valid instruments as long as the policy exposure is not correlated with the supply and demand shifters. The four resulting coefficient estimates can then be used to recover the four model elasticities. (For example, the local demand elasticity ε_{local}^D is one minus the ratio of the inward market access coefficients in the output and labor regressions). How should one construct the policy exposure instruments? The simplest way would be to use

equations (64) and (65) to calculate the market access terms before and after the policy shock, holding fixed income and expenditure before the policy shock to ensure that the changes in market access captured by the instruments is only arising from the effect of the policy shock on trade frictions.

There is one important (but subtle) caveat to the market access approach: when trade frictions are symmetric, the inward and outward market access terms will be equal up to scale (see Section 4.6). As a result, this approach will not be able to recover the full set of model elasticities. Nevertheless, one can still determine the resulting change in the equilibrium distribution of economic activity, a point which we show in the next (technical) subsection.

7.2.3 [Technical] The market access approach with symmetric trade costs

Recall that when trade costs are symmetric, equations (64) and (65) become:

$$\kappa MA_i^{in} = \sum_{j \in \mathcal{N}} T_{ij} \times \frac{E_j}{MA_j^{in}}, \quad (107)$$

where $MA_i^{out} = \kappa MA_i^{in}$ for some scalar κ , which we can write in changes as follows:

$$\widehat{\kappa MA_i^{in}} = \sum_{j \in \mathcal{N}} \left(\frac{X_{ij}^A}{Y_i^A} \right) \hat{T}_{ij} \frac{\hat{E}_j}{\widehat{MA_j^{in}}}. \quad (108)$$

Equation the supply and demand equations (103) and (104) and solving for the change in output holding constant the supply and demand shifters yields:

$$\hat{Y}_i \propto \widehat{MA_i^{in}} \frac{(1 - \varepsilon_{local}^D)^{\varepsilon_{global}^S} + (1 + \varepsilon_{local}^S)^{\varepsilon_{global}^D}}{\varepsilon_{local}^S + \varepsilon_{local}^D}, \quad (109)$$

which, when combined with equation (108) and imposing that $E_i = Y_i$ yields:

$$\nu \widehat{MA_i^{in}} = \sum_{j \in \mathcal{N}} \left(\frac{X_{ij}^A}{Y_i^A} \right) \hat{T}_{ij} \left(\widehat{MA_j^{in}} \right)^{1 - \frac{(1 - \varepsilon_{local}^D)^{\varepsilon_{global}^S} + (1 + \varepsilon_{local}^S)^{\varepsilon_{global}^D}}{\varepsilon_{local}^S + \varepsilon_{local}^D}}, \quad (110)$$

for an endogenous scalar ν that depends on κ and the endogenous scalars ϕ^D and ϕ^S . Equation (110) shares the same mathematical structure as equation (47) in Section 4.6 and hence its properties follow from the discussion there. In particular, as long as $\left| 1 - \frac{(1 - \varepsilon_{local}^D)^{\varepsilon_{global}^S} + (1 + \varepsilon_{local}^S)^{\varepsilon_{global}^D}}{\varepsilon_{local}^S + \varepsilon_{local}^D} \right| < 1$, there exists a unique (to-scale) set of inward market access changes that solve equation (110). Given these $\left\{ \widehat{MA_i^{in}} \right\}_{i \in \mathcal{N}}$, the equilibrium changes in output $\{\hat{Y}_i\}_{i \in \mathcal{N}}$ can be calculated from equation (109) (and the choice of numeraire requiring that $\sum_{j \in \mathcal{N}} Y_i^A \hat{Y}_i = 1$). An equivalent

derivation as in equation (109) can be used to solve for the equilibrium change in population, yielding $\hat{L}_i \propto \widehat{MA}_i^{in} \frac{\varepsilon_{global}^S + \varepsilon_{global}^D}{\varepsilon_{local}^S + \varepsilon_{local}^D}$ where the scale is determined by $\sum_{j \in \mathcal{N}} (L_j^A / \bar{L}) \hat{L}_j = 1$.

7.3 Testing the Models

So far we have discussed how to generate predictions from the quantitative geography framework and how to fit it to the data. Because fitting the model to the data involves exactly matching the observed distribution of economic activity, this leaves little room for testing the model. How can we then test whether the framework yields reliable predictions?

Testing a multi-region model is a significant challenge: a precursor of quantitative spatial models has been the large literature of Computable General Equilibrium (CGE) models that have been used to predict what happens in the economy when tariffs in certain sectors decrease (Brown and Stern (1989); Brown, Deardorff, and Stern (1991)); within and across countries. Despite nearly five decades of development CGE models have faced criticism for failing to accurately predict what happened to output and sectoral trade after tariff changes (see e.g. Kehoe (2005) and Kehoe, Pujolas, and Roszbach (2017)). We believe that the quantitative geography model hold more promise, for two reasons: first, the frameworks, as we showed in Sections 4-6, they are grounded in modern economic theory. Second, as we showed in Sections 7.1 and 7.2, their model parameters are estimated using modern techniques (unlike CGE models where parameters are typically externally calibrated).

We now discuss a few ways to test these models. The first way to test quantitative economic geography framework is to look at aspects of its cross-sectional predictions that are not related to the data we aim to fit. This is difficult, as the model is calibrated to exactly match observed incomes and populations, but one way to do so is to see if the calibrated supply and demand recovered “make sense.” For example, Desmet and Rossi-Hansberg (2013) show that the amenities recovered to calibrate the model indeed correlate well with observed characteristics that likely correlate strongly with amenities such as climate and life-quality measures.

A second way to test the quantitative economic geography framework is to see how well it is able to predict how the spatial distribution of economic activity actually responds to an observed shock. Typically, this is done by holding the calibrated supply and demand shifters fixed at their pre-shock levels, simulating the effect of the spatial shock, and comparing the predicted changes to the observed changes. For example, Donaldson (2018) shows that the observed change in real rental rates in response to railroad construction in Colonial India closely matches the model predicted changes. Similarly, Adao, Arkolakis, and Esposito (2019) and Allen, Fuchs, Ganapati, Graziano, Madera, and Montoriol-Garriga (2020) use linearized versions of spatial models to look at whether changes in cross-sectional income are predicted by the ones in the model, when taking into account indirect general equilibrium effects discussed

in Section 6.7. The two papers simulate an arguably exogenous shock, the China shock (Autor, Dorn, and Hanson (2013)) and tourism shock in Barcelona, respectively, after controlling for variation orthogonal to the shock. In fact, Allen, Fuchs, Ganapati, Graziano, Madera, and Montoriol-Garriga (2020) show that adding higher order effects, as in equation (94), offers a successively better approximation to the full equilibrium effects of tourism shock in Barcelona. One issue with these tests is that it is not clear what constitutes a “good” model fit.

Recently, Adão, Costinot, and Donaldson (2023) offer a formal econometric test for quantitative models based on a similar approach. In their test, one calibrates the supply and demand shifters to match the observed distribution of economic activity both before and after the shock. If the quantitative framework correctly predicts the causal changes in the distribution of economic activity from some exogenous shock, then the changes in the supply and demand shifters should themselves be uncorrelated with that shock, which offers moment restrictions that can be tested. They illustrate their approach by testing the setup of Fajgelbaum, Goldberg, Kennedy, and Khandelwal (2020) that is used to evaluate the impact of Trump tariffs across regions in the United States.

8 Exciting extensions

In this section, we briefly describe two recent advances that have extended the universal quantitative economic geography framework presented in the previous section in fruitful directions.

8.1 Multiple sectors

The framework presented above considers a single sector. Extending the framework to incorporate multiple sectors is a main motivation for quantitative work: since the pioneering work of Ricardo in 1817 there is a long tradition in international economics to consider multi-sector models and a literature in quantitative macroeconomics that considers multi-sector models as important for addressing the observed growth patterns (Long Jr and Plosser (1983)). More recently Anderson and Yotov (2010), Costinot, Donaldson, and Komunjer (2012), Ossa (2015), Caliendo and Parro (2015), Levchenko and Zhang (2016), among others, have discussed an extension of a quantitative international trade model with multiple sectors and Caliendo, Parro, Rossi-Hansberg, and Sarte (2018), Faber and Gaubert (2019), Arkolakis and Walsh (2023) have considered such an extension in a quantitative economic geography framework. Incorporating multiple sectors in the framework of Section 6 is straightforward. We now sketch out the procedure.

Suppose that there are $s \in \{1, \dots, S\} \equiv \mathcal{S}$ sectors and that the utility of a representative

consumer residing in location $j \in \mathcal{N}$ is given by:

$$W_j = \left(\sum_{s \in \mathcal{S}} \beta_j^s \left(\sum_{i \in \mathcal{N}} c_{ij}^{\frac{\sigma^s - 1}{\sigma^s}} \right)^{\frac{\sigma^s}{\sigma^s - 1} / \frac{\sigma - 1}{\sigma}} \right)^{\frac{\sigma}{\sigma - 1}} u_j, \quad (111)$$

where β_j^s , $\sum_s \beta_j^s = 1$, are consumption weights that consumer in location j assigns to the consumption of sector $s \in \mathcal{S}$. The elasticity of substitution σ_s is now across locations at the sector level and captures the varying degrees of substitutability that each goods in every sector may experience, whereas the elasticity of substitution σ is now across sectors. For example, estimates for petroleum indicate that this commodity is very substitutable, while advanced semiconductors are clearly less.

Maximizing the consumer's demand from equation (111) yields the following sectoral gravity equation:

$$X_{ij}^s = \frac{(p_{ij}^s)^{1 - \sigma^s}}{(P_j^s)^{1 - \sigma^s}} e_j^s w_j L_j, \quad (112)$$

where $e_j^s \equiv \frac{(p_j^s)^{1 - \sigma^s}}{P_j^{1 - \sigma^s}}$ is the consumption share of sector s in location j , $(P_j^s)^{1 - \sigma^s} \equiv \sum_{i \in \mathcal{N}} (p_{ij}^s)^{1 - \sigma^s}$ is a sector-specific price index, and $P_j^{1 - \sigma} \equiv \sum_{s \in \mathcal{S}} (\beta_j^s)^\sigma (P_j^s)^{1 - \sigma}$ is the composite price index.

The rest of the multi-sector framework bear a strong resemblance to the single sector baseline framework. Welfare of agents residing in location j can still be written in indirect utility form as $W_j = w_j u_j / P_j$. Goods market clearing now simply requires summing over sectors as well, i.e. $Y_i = \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{N}} X_{ij}^s$ and $E_i = \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{N}} X_{ji}^s$. Labor market clearing is unchanged, i.e. $w_i L_i = E_i = Y_i$.

For example, consider the simple case where labor is the only factor of production (so that $p_{ij}^s = \tau_{ij}^s w_i / A_i^s$) and the productivity and amenity externalities follow equations (22) and (23). Then following the derivations in Section 4.3, we arrive at the following (implicit) labor demand curve:

$$w_i L_i = \sum_{s \in \mathcal{S}} L_i^{\alpha(\sigma^s - 1)} w_i^{1 - \sigma^s} \sum_{j \in \mathcal{N}} \left(\tau_{ij}^s / \bar{A}_i^s \right)^{1 - \sigma^s} P_j^{\sigma^u - 1} w_j L_j, \quad (113)$$

while the labor supply curve is identical to equation (36) (just the new definition of the consumer price index).

The addition of multiple sectors can be further extended to include round-about production (as in Krugman and Venables (1995) and Eaton and Kortum (2002) and discussed in Section 5.8) and input-output linkages between sectors, as in Caliendo and Parro (2015) and Caliendo, Dvorkin, and Parro (2019).

8.2 Migration costs

The framework presented in Section 6 abstracts from any costs associated with movement across space. There has been substantial recent progress extending the quantitative economic geography framework to incorporate such migration costs, including Tombe and Zhu (2019), Allen, de Castro Dobbin, and Morten (2018), Desmet, Nagy, and Rossi-Hansberg (2018), Caliendo, Dvorkin, and Parro (2019), Allen and Donaldson (2018) and Kleinman, Liu, and Redding (2024). Here we sketch out a simple static quantitative economic geography model with migration costs, and we refer the interested reader to Chapter XXX Klaus and Fernando’s Chapter XXX in his handbook for more details of recent dynamic advances.

Suppose that each location $i \in \mathcal{N}$ is endowed with an initial allocation of agents \bar{L}_i . Each agent ω initially allocated to $i \in \mathcal{N}$ then choose in which location to live in order to solve:

$$\max_{j \in \mathcal{N}} \frac{W_j}{\mu_{ij}} \times \varepsilon_{ij}(\omega), \quad (114)$$

where W_j is the welfare (real wage and amenity value) of residing in location $j \in \mathcal{N}$, $\mu_{ij} \geq 1$ is a bilateral migration friction, and $\varepsilon_{ij}(\omega)$ is an idiosyncratic preference term, which we assume is Frechet distributed, i.e. $\varepsilon_{ij}(\omega) \sim e^{-z^{-\theta}}$. As in Sections 5.1 and 5.5, maximization of (114) leads to a probabilistic decision where the fraction of workers initially allocated to $i \in \mathcal{N}$ that migrate to $j \in \mathcal{N}$ is:

$$L_{ij}/\bar{L}_i = (W_j/\mu_{ij})^\theta / \sum_{k \in \mathcal{N}} (W_k/\mu_{ik})^\theta. \quad (115)$$

Summing across all possible origin locations then allows us to construct the following modified labor supply curve:

$$\ln w_j = \frac{1}{\theta} \ln L_j + \frac{1}{1-\sigma} \ln P_j - \ln u_i - \frac{1}{\sigma} \ln \sum_{i \in \mathcal{N}} \frac{\mu_{ij}^{-\theta} \bar{L}_i}{\sum_{k \in \mathcal{N}} (W_k/\mu_{ik})^\theta}. \quad (116)$$

The first three terms of Equation (116) are identical to the standard labor supply curve in equation (62), but the fourth term is new. This fourth term is a *migration* market access term that summarizes how costly it is for migrants to reach location j relative to all other locations. Intuitively, the labor supply curve will shift outward for locations to which it is easier to migrate.

9 The next generation

In the previous sections, we have discussed the major advances made in turning the theory of earlier generations “quantitative” so that it can be combined with detailed spatial data to perform real world analyses. We have also highlighted current research extending and expanding

the scope of quantitative economic geography frameworks. Here, we conclude by offering our best guess of what the next generation of quantitative models may bring. To do so, we highlight four possible paths forward for future research.

9.1 Taking inequality and agent heterogeneity more seriously

In the quantitative economic geography framework developed above, all agents were identical (except, perhaps, up to an idiosyncratic component). But, of course, in reality people differ along a wide variety of dimensions, including their education, skills, race, sex, gender, age, etc. These differences affect agent's preferences of where they live, where they work, and how they interact with others. And an important concern when making public policy is how such policies differentially impact different groups of people.

Extending the quantitative economic geography framework to incorporate such a rich heterogeneity is necessary to tackle issues related to inequality head on, and we think this is a fruitful direction for the next generation of research. Indeed, there are a number of recent papers that have made important strides in this direct. For example, Tsivanidis (2022) incorporates multiple types of workers endowed with varying skills and non-homothetic preferences to evaluate the differential impacts of rapid bus transit in Bogota. Similarly, Zárata (2022) examines the differential impacts of subway expansions in Mexico City using a model where agents are ex-ante identical but ex-post sort into different sectors and locations using a Roy (1951)-like framework.

One important policy-relevant topic that requires a richer treatment of agent heterogeneity is segregation. Here too there are several recent exciting advances. Hoelzlein (2019) introduces agents who differ in their skill endowments and non-homothetic preferences to study the sorting of firms and residents and the impact of Empowerment Zones on segregation in Los Angeles. Recent papers by Bagagli (2023) and Weiwu (2023) specifically study racial segregation, incorporating homophily into agents' preferences and used such a framework to understand how highway construction has amplified segregation in cities. Allen, Arkolakis, and Li (2023) characterize the theoretical properties of quantitative spatial frameworks with many types of agents and arbitrary cross-group preferences.

Another dimension on which agents can differ is in their family structure. Recent work has emphasized how family ties affects preferences through remittances (see Albert and Monras (2022)), but there are number of interesting questions relating to the spatial implications of "family economics" that have yet to be addressed. How do families with and without children differ in their preferences and choices of location? How should couples jointly choose their location? How do people's preferences over locations and different types of amenities change over their life cycle?

There are also interesting and important issues related to inequality across generations.

How do parents' choices of the children's human capital acquisition affect the intergenerational transfer of wealth, long-run migration patterns, and ultimately spatial inequality? To date, most dynamic quantitative economic geography models have assumed an infinitely lived agent (see XXX Klaus and Fernando's Chapter XXX). While there has been some early work on the topic (see e.g. Eckert and Kleineberg (2024) for an application to U.S. school funding and Pellegrina and Sotelo (2021) for the implications for the evolution of comparative advantage), extending the dynamic economic geography frameworks to more seriously model inter-generational transfers of wealth and capital can better guide policies meant to ameliorate systemic inequalities and help us better understand the long-term effects of spatial policies.

9.2 Quantum quantitative economic geography

One of the central achievements in quantitative economic geography is its ability to incorporate a rich geography: locations vary in their innate productivities, amenities, and costs in interacting with other locations. Yet while the geography being considered is incredibly rich in these dimensions, it is also surprisingly poor across others. Here are two dimensions in which we think future work could make substantial improvements.

First, in the workhorse framework presented above, all locations produce differentiated varieties that enter consumer demand in the same way. While it is straightforward to extend this framework to multiple sectors (see Section 8.1), currently absent is the idea that locations may play fundamentally different roles. For example, the "Central Place Theory" of Christaller (1933) proposed a hierarchy of locations, where large central cities are surrounded by smaller towns, which are in turn surrounded by agricultural areas. The central cities are able to sustain more specialized businesses (e.g. sports stadiums) that surrounding residents also enjoy, whereas firms in smaller surrounding towns are only those necessary for more every day needs (e.g. grocery stores). Such concepts of spatial specialization are absent from current quantitative frameworks. However, there have been some recent advances in this direction: for example, Nagy (2023) develops a model featuring cities surrounded by agricultural hinterlands.

A necessary step toward allowing different locations to host distinctly different sets of firms is to "quantize" quantitative economic geography by considering individual firms rather than a continuum of firms. Determining how a discrete number of firms optimally choose their locations is a difficult problem, but there have been exciting steps forward in recent work by Oberfield, Rossi-Hansberg, Sarte, and Trachter (2024). Technical advances in combinatorial discrete choice problems by Arkolakis, Eckert, and Shi (2017) and applications of clustering techniques (see e.g. Allen (2023)) may offer other potential paths forward.

9.3 Optimal policy design

One of the most exciting components of quantitative economic geography is its ability to offer normative prescriptions to real world spatial policies. Existing research has mostly used the framework to evaluate the welfare effects of economic policies *ex post*. But we think there is an enormous potential to use the same frameworks to design better spatial policies *ex ante*.

As summarized in XXX Cecile and Pablo's Chapter XXX, there has been substantial recent progress in characterizing the welfare implications of this class of quantitative economic geography models, and we refer the interested reader to that chapter for a much more thorough treatment of the related research. Yet while the work by Fajgelbaum and Gaubert (2020), Fajgelbaum and Schaal (2020), and Donald, FUKU, and Miyauchi (2023) have clarified the normative properties of this broad class of models and recent work has started to assess political economy considerations in spatial policies (see e.g. Bordeu (2023) and Fajgelbaum, Gaubert, Gorton, Morales, and Schaal (2023)), there remains the opportunity to apply these models to policy design. Pressing questions in this vein include: How should we best design congestion tolls on city streets? How should a government best allocate its infrastructure budget across its road network? What is the best spatial transfer scheme to address both efficiency and equity considerations?

9.4 Market power

A fourth exciting direction for future research would be to explore the implications of incorporating firm market power. The existing workhorse models described above typically assume perfect or monopolistic competition in both product and labor markets, but in reality firms likely exhibit substantial market power in both. While there has been substantial number of papers examining the role of market power in international trade (see e.g. Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019) for the study of imperfect competition in product markets and Felix (2021) for the study of imperfect competition in labor markets), much less has been done studying the role of market power in the spatial distribution of economic activity. Notable exceptions include Gutiérrez (2022), who develops a model features imperfect competition in both product and labor markets to understand how market power affects the welfare gains from trade and Hong (2024), who develops a model with monopsony power of firms to consider the two-sided matching of heterogeneous firms and workers and its implications for the equilibrium distribution of economic activity across space.

Another potential source of market power is in the transportation sector. Typically, quantitative economic models assume that trade frictions are exogenous or, equivalently, that the transportation sector is perfectly competitive. But in reality, the transportation sector is often-times quite concentrated, and the imperfect competition in the sector can affect the equilibrium spatial distribution of economic activity. There have been several recent studies examining

how imperfect competition in the transportation sector affects trade costs, including Braccacio, Kalouptsi, and Papageorgiou (2020) (bulk shipping), Wong (2022) (container shipping), and Allen, Atkin, Cleves, and Hernandez (2024) (trucking), but it remains an unresolved question about how such endogenous trade costs affects the equilibrium distribution of economic activity.

9.5 Conclusion

Over the past decade, there has been a remarkable advance in the understanding of how geography shapes the spatial distribution of the economy. In this handbook chapter, we have shown how a new quantitative economic geography framework has been able to unify the economic mechanisms from two seminal models of the previous generation in such a way that they can be combined with detailed spatial economic data to better understand real world economic policy. There are many exciting aspects of the framework. It is tractable: it is mathematically well-behaved and easy to solve. It is also powerful, allowing one to evaluate the spatial impacts of nearly any change in the underlying geography. And it is flexible: it has many different micro-foundations and it can be extended in many directions. Most excitingly, however, is that while some of those extensions have been explored, most have not, suggesting that the future of economic geography is bright.

References

- ADAO, R., C. ARKOLAKIS, AND F. ESPOSITO (2019): “General equilibrium effects in space: Theory and measurement,” Discussion paper, National Bureau of Economic Research. 43, 53
- ADÃO, R., A. COSTINOT, AND D. DONALDSON (2023): “Putting quantitative models to the test: An application to Trump’s trade war,” Discussion paper, National Bureau of Economic Research. 54
- 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. 31, 51
- ALBERT, C., AND J. MONRAS (2022): “Immigration and spatial equilibrium: the role of expenditures in the country of origin,” *American Economic Review*, 112(11), 3763–3802. 57
- ALLEN, T. (2023): “The topography of nations,” Discussion paper, National Bureau of Economic Research. 58
- ALLEN, T., AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 129(3), 1085–1140. 16, 26, 29, 34, 35, 46, 48
- (2018): “Modern spatial economics: a Primer,” *World Trade Evolution*, p. 435. 25
- (2022): “The welfare effects of transportation infrastructure improvements,” *The Review of Economic Studies*, 89(6), 2911–2957. 48, 49
- ALLEN, T., C. ARKOLAKIS, AND X. LI (2023): “On the equilibrium properties of spatial models,” *Am. Econ. Rev.: Insights (Forthcoming)*. 29, 39, 40, 41, 57
- ALLEN, T., C. ARKOLAKIS, AND Y. TAKAHASHI (2020): “Universal gravity,” *Journal of Political Economy*, 128(2), 393–433. 26, 39, 43, 45
- ALLEN, T., D. ATKIN, S. C. CLEVES, AND C. E. HERNANDEZ (2024): “Trucks,” Discussion paper, Working paper. 60
- ALLEN, T., C. DE CASTRO DOBBIN, AND M. MORTEN (2018): “Border walls,” Discussion paper, National Bureau of Economic Research. 56
- ALLEN, T., AND D. DONALDSON (2018): “The geography of path dependence,” *Unpublished manuscript*. 50, 56
- ALLEN, T., S. FUCHS, S. GANAPATI, A. GRAZIANO, R. MADERA, AND J. MONTORIOL-GARRIGA (2020): “Is tourism good for locals? Evidence from Barcelona,” *Dartmouth College, Mimeograph*. 53, 54
- ANDERSON, J. E. (1979): “A theoretical foundation for the gravity equation,” *The American economic review*, 69(1), 106–116. 5, 16
- ANDERSON, J. E., AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzle,” *American Economic Review*, 93(1), 170–192. 5, 21, 26
- ANDERSON, J. E., AND Y. V. YOTOV (2010): “The changing incidence of geography,” *American Economic Review*, 100(5), 2157–2186. 54
- ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2019): “The elusive pro-competitive effects of trade,” *The Review of Economic Studies*, 86(1), 46–80. 59

- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New trade models, same old gains?," *American Economic Review*, 102(1), 94–130. 33
- ARKOLAKIS, C., F. ECKERT, AND R. SHI (2017): "Combinatorial discrete choice: A Quantitative Model of Multinational Location," *Available at SSRN 3455353*. 58
- ARKOLAKIS, C., AND C. WALSH (2023): "Clean Growth," Discussion paper, National Bureau of Economic Research. 54
- ARMINGTON, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *International Monetary Fund Staff Papers*, 16, 159–178. 16
- 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), 2121–2168. 54
- BAGAGLI, S. (2023): "The (Express) Way to Segregation: Evidence from Chicago," . 57
- BORDEU, O. (2023): "Commuting infrastructure in fragmented cities," *Job Market Paper, University of Chicago Booth School of Business*, 6. 59
- BRANCACCIO, G., M. KALOUPTSIDI, AND T. PAPAGEORGIU (2020): "Geography, transportation, and endogenous trade costs," *Econometrica*, 88(2), 657–691. 60
- BROWN, D., A. DEARDORFF, AND R. STERN (1991): "A North American free trade agreement: Analytical issues and a computational assessment," *World Scientific Studies in International Economics*, p. 393. 53
- BROWN, D. K., AND R. M. STERN (1989): "US-Canada bilateral tariff elimination: The role of product differentiation and market structure," in *Trade policies for international competitiveness*, pp. 217–254. University of Chicago Press. 53
- BRYAN, G., AND M. MORTEN (2019): "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 127(5), 2229–2268. 32, 35
- 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. 55, 56
- CALIENDO, L., AND F. PARRO (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82(1), 1–44. 54, 55
- 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. 54
- CHRISTALLER, W. (1933): *Die zentralen Orte in SÄCdeutschland*. 58
- COSTINOT, A., D. DONALDSON, AND I. KOMUNJER (2012): "What goods do countries trade? A quantitative exploration of Ricardo's ideas," *The Review of economic studies*, 79(2), 581–608. 54
- 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. 41
- DEKLE, R., J. EATON, AND S. KORTUM (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55(3), 511–540. 20, 41, 42
- DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2018): "The Geography of Development," *Journal of Political Economy*, 126(3), 903–983. 56

- DESMET, K., AND E. ROSSI-HANSBERG (2013): "Urban accounting and welfare," *American Economic Review*, 103(6), 2296–2327. 53
- DIJKSTRA, E. W. (1959): "A note on two problems in connexion with graphs," *Numerische Mathematik*, 1(1), 269–271. 47
- DONALD, E., M. FUKU, AND Y. MIYAUCHI (2023): "Unpacking Aggregate Welfare in a Spatial Economy," Discussion paper, Tech. rep. 45, 59
- DONALDSON, D. (2018): "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 108(4-5), 899–934. 46, 48, 53
- DONALDSON, D., AND R. HORNBECK (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 131(2), 799–858. 26, 31, 35, 46, 51
- EATON, J., AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70(5), 1741–1779. 5, 30, 31, 33, 35, 55
- ECKERT, F., AND T. KLEINEBERG (2024): "The geography of opportunity: Education, work, and inter-generational mobility across us counties," Discussion paper, Working Paper. 58
- ENRICO, M. (2011): "Local labor markets," in *Handbook of labor economics*, vol. 4, pp. 1237–1313. Elsevier. 7, 15
- FABER, B., AND C. GAUBERT (2019): "Tourism and economic development: Evidence from Mexico's coastline," *American Economic Review*, 109(6), 2245–2293. 54
- FAJGELBAUM, P. D., AND C. GAUBERT (2020): "Optimal spatial policies, geography, and sorting," *The Quarterly Journal of Economics*, 135(2), 959–1036. 45, 59
- FAJGELBAUM, P. D., C. GAUBERT, N. GORTON, E. MORALES, AND E. SCHAAL (2023): "Political Preferences and the Spatial Distribution of Infrastructure: Evidence from California's High-Speed Rail," Discussion paper, National Bureau of Economic Research. 59
- FAJGELBAUM, P. D., P. K. GOLDBERG, P. J. KENNEDY, AND A. K. KHANDELWAL (2020): "The return to protectionism," *The Quarterly Journal of Economics*, 135(1), 1–55. 54
- FAJGELBAUM, P. D., AND E. SCHAAL (2020): "Optimal transport networks in spatial equilibrium," *Econometrica*, 88(4), 1411–1452. 59
- FALLY, T. (2015): "Structural gravity and fixed effects," *Journal of international economics*, 97(1), 76–85. 47
- FAN, J., Y. LU, AND W. LUO (2023): "Valuing domestic transport infrastructure: A view from the route choice of exporters," *Review of Economics and Statistics*, 105(6), 1562–1579. 49
- FELIX, M. (2021): "Trade, labor market concentration, and wages," *Job Market Paper*. 59
- FOGEL, R. W. (1964): *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press. 24
- FROBENIUS, G., F. G. FROBENIUS, F. G. FROBENIUS, F. G. FROBENIUS, AND G. MATHEMATICIAN (1912): "Über Matrizen aus nicht negativen Elementen," . 23
- FUCHS, S., AND W. F. WONG (2022): "Multimodal transport networks," Discussion paper, Working Paper. 49

- FUJITA, M., P. KRUGMAN, AND A. J. VENABLES (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press. 10, 13, 14, 29
- 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. 7, 50
- GUTIÉRREZ, A. (2022): "Labor market power and the pro-competitive gains from trade," *Unpublished Working Paper*. 59
- HEAD, K., AND T. MAYER (2014): "Gravity equations: Workhorse, toolkit, and cookbook," in *Handbook of international economics*, vol. 4, pp. 131–195. Elsevier. 5
- HELPMAN, E., ET AL. (1995): *The size of regions*. Foerder Institute for Economic Research. 30, 35
- HOELZLEIN, M. (2019): "Two-sided sorting and spatial inequality in cities," *Unpublished manuscript*. 57
- HONG, G. (2024): "Two-sided sorting of workers and firms: Implications for spatial inequality and welfare," Discussion paper, Canadian Labour Economics Forum (CLEF), University of Waterloo. 59
- HULTEN, C. R. (1978): "Growth accounting with intermediate inputs," *The Review of Economic Studies*, 45(3), 511–518. 24
- JACKSON, M. O., ET AL. (2008): *Social and economic networks*, vol. 3. Princeton university press Princeton. 24
- 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. 46
- KARLIN, S., AND L. NIRENBERG (1967): "On a theorem of P. Nowosad," *Journal of Mathematical Analysis and Applications*, 17(1), 61–67. 28, 41, 42
- KEHOE, T. J. (2005): "An Evaluation of the Performance of Applied General Equilibrium Models on the Impact of NAFTA," *This page intentionally left blank*, 341. 53
- KEHOE, T. J., P. S. PUJOLAS, AND J. ROSSBACH (2017): "Quantitative trade models: Developments and challenges," *Annual Review of Economics*, 9(1), 295–325. 53
- KLEINMAN, B., E. LIU, AND S. J. REDDING (2020): "International friends and enemies," Discussion paper, National Bureau of Economic Research. 43, 45
- (2024): "The linear algebra of economic geography models," in *AEA Papers and Proceedings*, vol. 114, pp. 328–333. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. 43, 56
- KRUGMAN, P. (1991): "Increasing returns and economic geography," *Journal of political economy*, 99(3), 483–499. 7, 10, 12, 15, 32, 34, 35
- KRUGMAN, P., AND A. J. VENABLES (1995): "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, 110(4), 857–880. 33, 55
- KUCHERYAVYY, K., G. LYN, AND A. RODRÍGUEZ-CLARE (2021): "Spatial equilibria: The case of two regions," Discussion paper, National Bureau of Economic Research. 29
- (2023): "Grounded by gravity: A well-behaved trade model with industry-level economies of scale," *American Economic Journal: Macroeconomics*, 15(2), 372–412. 15

- LEVCHENKO, A. A., AND J. ZHANG (2016): "The evolution of comparative advantage: Measurement and welfare implications," *Journal of Monetary Economics*, 78, 96–111. 54
- LONG JR, J. B., AND C. I. PLOSSER (1983): "Real business cycles," *Journal of political Economy*, 91(1), 39–69. 54
- M. GROSSMAN, G., AND E. ROSSI-HANSBERG (2010): "External economies and international trade redux," *The Quarterly journal of economics*, 125(2), 829–858. 15
- MARSHALL, A. (1890): *Principles of Economics*. 1
- MISES, R., AND H. POLLACZEK-GEIRINGER (1929): "Praktische Verfahren der Gleichungsauflösung.," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 9(1), 58–77. 23
- NAGY, D. K. (2023): "Hinterlands, city formation and growth: Evidence from the US westward expansion," *Review of Economic Studies*, 90(6), 3238–3281. 58
- OBERFIELD, E., E. ROSSI-HANSBERG, P.-D. SARTE, AND N. TRACHTER (2024): "Plants in space," *Journal of Political Economy*, 132(3), 867–909. 58
- OSSA, R. (2015): "Why trade matters after all," *Journal of International Economics*, 97(2), 266–277. 33, 54
- PELLEGRINA, H. S., AND S. SOTELO (2021): "Migration, Specialization, and Trade: Evidence from Brazil's March to the West," Discussion paper, National Bureau of Economic Research. 58
- PERRON, O. (1907): "Zur theorie der matrices," *Mathematische Annalen*, 64(2), 248–263. 23
- REDDING, S., AND A. J. VENABLES (2004): "Economic geography and international inequality," *Journal of international Economics*, 62(1), 53–82. 10
- REDDING, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101, 148–167. 31, 35
- REDDING, S. J., AND D. M. STURM (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766–1797. 35
- ROBACK, J. (1982): "Wages, rents, and the quality of life," *The Journal of Political Economy*, pp. 1257–1278. 7, 8, 9, 10, 15, 34, 35
- ROSEN, S. (1979): "Wage-Based Indexes of Urban Quality of Life," in *Current Issues in Urban Economics*, ed. by P. Mieszkowski, and M. Straszheim, pp. 74–104. Johns Hopkins University Press. 7, 8, 9, 10, 15, 34, 35
- ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3(2), 135–146. 57
- SAMUELSON, P. A. (1954): "The transfer problem and transport costs, II: Analysis of effects of trade impediments," *The Economic Journal*, 64(254), 264–289. 16
- SETHIAN, J. (1996): "A Fast Marching Level Set Method for Monotonically Advancing Fronts," *Proceedings of the National Academy of Sciences*, 93(4), 1591–1595. 47
- SEVEREN, C. (2023): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *Review of Economics and Statistics*, 105(5), 1073–1091. 46

- SILVA, J. S., AND S. TENREYRO (2006): "The log of gravity," *The Review of Economics and statistics*, 88(4), 641–658. 46
- SMITH, A. (1776): *The Wealth of Nations*. 1
- TOMBE, T., AND X. ZHU (2019): "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 109(5), 1843–1872. 56
- TSITSIKLIS, J. (1995): "Efficient Algorithms for Globally Optimal Trajectories," *Automatic Control, IEEE Transactions on*, 40(9), 1528–1538. 47
- TSIVANIDIS, N. (2022): "Evaluating the impact of urban transit infrastructure: Evidence from Bogota's transmilenio," *Unpublished manuscript*, 18. 57
- VAHRENKAMP, R. (1976): "Derivatives of the dominant root," *Applied Mathematics and Computation*, 2(1), 29–39. 24
- VON THUNEN, J. H. (1826): *The Isolated State*. 1
- WEIWU, L. (2023): "Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities," Discussion paper, Technical report, MIT. 57
- WONG, W. F. (2022): "The round trip effect: Endogenous transport costs and international trade," *American Economic Journal: Applied Economics*, 14(4), 127–166. 60
- ZÁRATE, R. D. (2022): "Spatial misallocation, informality, and transit improvements: Evidence from mexico city," . 57