Populations in Spatial Equilibrium*

Matthew Easton and Patrick W. Farrell

This Version: May 2024

Main Text Word Count: 11,300

Main Text and Appendix Text Word Count: 21,200

Abstract

The appearance of power law-like distributions for city populations is a distinctive, recurring feature of human geography. We propose an explanation for this phenomenon that reflects both variation in geography and trade between locations. The equilibrium population in many spatial models consists of a term reflecting a location's own productivity and amenity value, its "fundamentals," and a term reflecting a location's ability to benefit from trade, its "market access." Using random variation in geographic attributes within a location and trade costs across locations to realistically model these two terms results in lognormal population distributions which appear to follow a power law for the most populous locations (i.e., cities). Simulations mirror several empirical results in the literature. *JEL Codes:* R12, R13, F10, C31, C6.

*Easton email: me2713@columbia.edu. Farrell email: pwf2108@columbia.edu. The authors thank Arslan Ali, Nadia Ali, Costas Arkolakis, Pierre-Philippe Combes, Donald Davis, Richard A. Davis, Jonathan Dingel, Thibault Fally, Madeline Hansen, Raphael Lafrogne-Joussier, Serena Ng, Eshaan Patel, Stephen Redding, Andrés Rodríguez-Clare, Esteban Rossi-Hansberg, Bernard Salanié, Lin Tian, Conor Walsh, David Weinstein, Natalie Yang, and seminar participants at Columbia University for helpful comments and discussions. Easton acknowledges the Program for Economic Research at Columbia University and the Alliance Doctoral Mobility Grant from Columbia University and Sciences Po for support.

1 Introduction

The distribution of city populations in most countries appears to follow a power law. This distinctive empirical regularity is remarkable given the substantially different contexts in which it has been observed, having been documented extensively across countries,[1] varying definitions of cities,[2] and periods of human history spanning millennia.[3] That is, a power law-like distribution appears even when the meaning of "city," the forces shaping cities, levels of economic development, and the degree of integration of cities into national or global networks all vary greatly. An example of this characteristic distribution, often called Zipf's law, can be seen in Figure I which shows the appearance of a power law for the top 250 U.S. metropolitan statistical areas (MSAs) in 2020.

We explain the appearance of power law-like city size distributions based on variation in geography and trade costs between locations. The characteristics of a location matter for whether or not people want to live there, which we refer to as the importance of "place." A location's position relative to other populated locations also matters for whether or not people want to live there, which we refer to as the importance of "space." Prior theoretical explanations for the city size distribution have incorporated at most one of place or space, but often lack both. Our paper is the first to show how these forces together result in the distinctive distribution of human populations. In doing so our work bridges the literatures on population distributions and Zipf's law, on models of the spatial economy, and on the influence of physical geography for human settlement patterns.

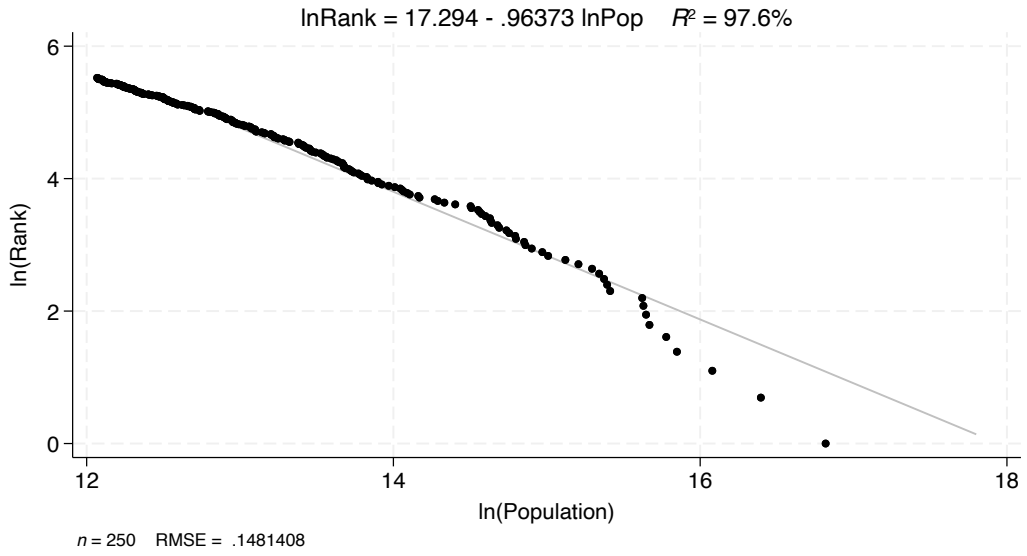
We demonstrate our result within the quantitative spatial equilibrium model developed in Allen and Arkolakis (2014). In this framework, as in many spatial models, the population in each location can be expressed as the product of a trade-weighted summation across

[1] A power law-like city distribution was first documented in Germany by Auerbach (1913) and the early comparative literature began with Zipf (1949). A recent and comprehensive comparative investigation is Soo (2005), which looks at 73 countries.

[2] Several papers have used nightlights data to define cities rather than administrative borders, such as Jiang, Yin, and Liu (2014) and Dingel, Miscio, and Davis (2021).

[3] Davis and Weinstein (2002) demonstrate a power-law relationship existed in pre-modern Japan, while Barjamovic et al. (2019) show evidence of this pattern in Bronze Age Anatolia.

Figure I:
An Apparent Power Law for U.S. Cities in 2020



Note: The appearance of a power law for top 250 U.S. metropolitan statistical areas (MSAs) in 2020.
Data Source: U.S. Census

all locations, reflecting its "market access" and the role of trade across space, and a term capturing the geographic advantages of a location, reflecting its "fundamentals" and the role of the characteristics of a place. As all elements of the market access summation are positive random variables, a useful lemma from Marlow (1967) shows that under limited conditions the market access terms will be lognormally distributed. Further, following Lee and Li (2013), we model locational fundamentals as resulting from multiplicative shocks due to random variation in the geographic attributes of a place. As multiplicative processes are additive in logs, the fundamentals will also be lognormally distributed.

Our key innovation is to show that these two different types of random variation—that across locations modelled as an additive process and that within locations modelled as a multiplicative process—both produce lognormally distributed terms within the equilibrium population equation. The lognormal distribution of both the market access term reflecting the role of space and the fundamentals term reflecting the role of place are preserved by the algebraic operations within the equilibrium population equation. The population will

be lognormally distributed because products of lognormal random variables also follow a lognormal distribution under limited conditions. As the tail of the lognormal distribution closely resembles a Pareto distribution, the distribution of highly populous locations (i.e., cities) will appear to follow a power law.

We demonstrate the success of the model at generating both lognormal full population distributions and power law-like city size distributions through simulation. The result is robust to changes in model parameters and allows us to explain the persistent appearance of, and variation in, the city size distribution in different contexts. We explore how changes to local productivity spillovers, intra-city congestion externalities, and inter-city transportation costs influence the observed city size distribution and show that the changes to the distribution are consistent with the empirical evidence. The model is also consistent with other commonly observed features of the city size distribution such as the similarly striking empirical observation that city growth rates often appear unrelated to city population, referred to in the literature as Gibrat's law.[4]

This paper touches on several topics within the spatial economics literature. First, it relates to work on the power-law like city size distribution and Zipf's law. Many theoretical explanations of this phenomenon are based on Gibrat's law and use a random growth process for cities over time as the basis for the appearance of a power law-like distribution (Gabaix, 1999a; Gabaix, 1999b; Blank and Solomon, 2000; Eeckhout, 2004; Rossi-Hansberg and Wright, 2007; Córdoba, 2008). However, the assumption of random growth is inconsistent with the empirical evidence on the distribution of cities in significant ways. Cities tend to recover rapidly following major shocks rather than grow randomly,[5] and the growth of cities does not appear random during transitions to new spatial equilibria.[6] Further, ran-

[4]The "law" is an application of the central limit theorem to the log of the product of independent shocks, and was originally formulated to describe the growth of firms (Gibrat, 1931).

[5]Notable instances of recovery from shocks are documented in Davis and Weinstein (2002), Brakman, Garretsen, and Schramm (2004) and Davis and Weinstein (2008) following bombings, and in Johnson, Jedwab, and Koyama (2019) following pandemics.

[6]Desmet and Rappaport (2017) document the absence of the random growth phenomenon for cities during the settlement of the American West.

dom growth explanations fail to capture the influence of the observable characteristics of a place on the attractiveness of producing or residing there. Random growth models imply that the large populations of New York City, Tokyo, and London are unrelated to their advantageous geographies. These theories are also aspatial and do not allow for interactions between different locations across space to shape settlement patterns, failing to capture the contribution of trade to the scale of the aforementioned global cities. That is, the random growth explanations lack both place and space. Papers that do account for the importance of place in determining the city size distribution, such as Lee and Li (2013) and Behrens and Robert-Nicoud (2015), have not included a role for space, and those which account for the importance of space, such as Brakman, Garretsen, Van Marrewijk, et al. (1999) and Rante, Trionfetti, and Verma (2024), have not included a role for differences in place and lack differentiated exogenous geographies.

Branches within this literature differ in their interpretation of the population distribution, and we provide support for prior work that has argued for a lognormal distribution as best matching the population distribution within countries. Some work has argued that the city size distribution follows a Pareto distribution, such as Gabaix (1999a) and Gabaix (1999b), ignoring smaller settlements and interpreting the truncated city size distribution as reflecting a true power law. Other work, such as Eeckhout (2004), has argued that a lognormal distribution better matches the full population distribution, appears similar to a Pareto distribution for realizations in the right tail like cities, and naturally captures deviations from a true power law in the data. We argue in favor of the lognormal interpretation, and show how lognormality emerges within a broad class of spatial models. Our explanation of the lognormal distribution's emergence is also consistent with the appearance of an approximate power law at different levels of spatial aggregation (Holmes, 2010; Rozenfeld et al., 2011; Mori, Smith, and Hsu, 2020).

A second related literature is that on spatial models. These models highlight the roles of space, local spillovers, and the importance of trade and interactions between locations

in determining population distributions (Fujita, Krugman, and Venables, 1999; Allen and Arkolakis, 2014; Redding, 2016; Redding and Rossi-Hansberg, 2017). Indeed, the largest cities tend to be favorably located for trade with other locations.[7] Early spatial models such as Krugman (1991) primarily focused on the role of local population spillovers and trade, and assumed no differentiation in first-nature geography across locations.[8] Modern spatial equilibrium models as in Allen and Arkolakis (2014) also incorporate variation in exogenous fundamentals reflecting a differentiated geography. For certain parameter values, these models can be inverted to recover the fundamentals given any distribution of population. However, absent a theory for the distribution of fundamentals, this literature cannot explain why population tends to be distributed similarly in many different contexts. We demonstrate how both heterogeneity in the attributes of different places and trade across space, which are key features of these models, will generate lognormal population distributions in equilibrium as a result of random variation in geography.[9]

A third related branch of literature focuses on the role of favorable geography in explaining human settlement patterns. The largest cities around the world tend to be in locations that are good for production and offer quality-of-life benefits to residents. A literature on the intuitive importance of natural characteristics for explaining settlement patterns has found a large role for such "first-nature" geographic characteristics, as in Rappaport and Sachs (2003), Nordhaus (2006), Nunn and Puga (2012), Bosker and Buringh (2017), and Alix-Garcia and Sellars (2020). Especially relevant for our work is Henderson et al. (2018), who demonstrate that a granular dataset of first-nature geographic attributes explains roughly 47% of worldwide variation in economic activity as measured by nightlights. We use this

[7]New York City is located on one of the largest natural harbors on Earth and its much greater population relative to Lost Springs, Wyoming—the 2020 population ratio was 8,804,190 to 6—is likely related to New York's favorable geography and the benefits of its location for trade. Some attributes of landlocked Lost Springs include its low annual precipitation and a coal mine which last operated in the 1930s.

[8]In work similar to our own simulated exercises, Brakman, Garretsen, Van Marrewijk, et al. (1999) identify a power law-like population distribution in simulations of a model based on Krugman (1991) with trade but no differentiated geography.

[9]We also show that the exogenous locational fundamentals recovered in Allen and Arkolakis (2014) appear lognormal, supporting our modeling of fundamentals as lognormally distributed. We discuss this further in Section 3 and Online Appendix C.8.

dataset to investigate assumptions regarding the correlation within geographic attributes within a place and of attributes across space that we use to characterize the population distribution. We show that realistically modelling geographic variation within places and across space results in population distributions similar to those observed in the data within spatial models. Given the persistence of geography our framework can explain the recovery of cities from negative shocks.

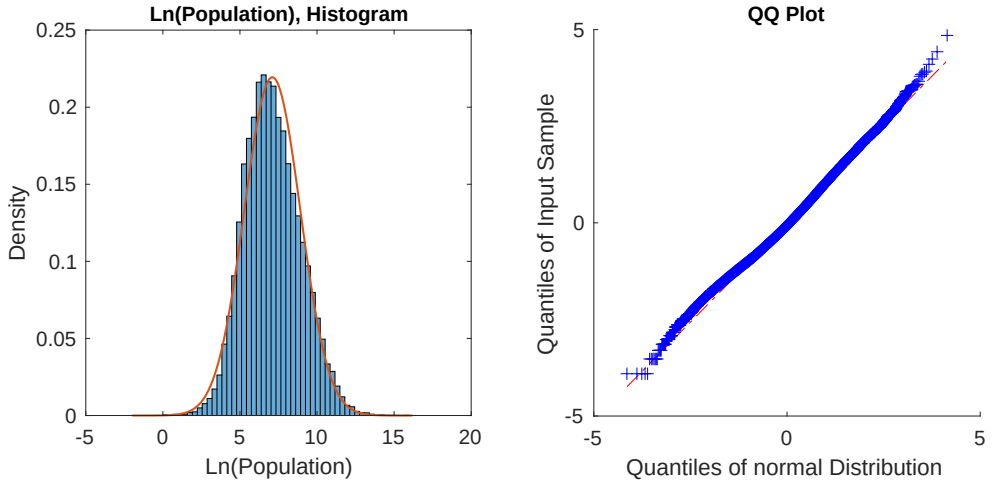
At the intersection of these literatures, our work is the first to generate realistic city size distributions based on heterogeneous geography within a broad class of spatial models. The paper proceeds as follows. Section 2 argues that populations are best described by a lognormal distribution and establishes the link between this distribution and the appearance of a power law for cities. Section 3 presents a standard quantitative spatial model and shows that, given the structure of the equilibrium condition, a lognormal population distribution results from both trade across space and random variation in geography within places. Section 4 demonstrates that the model captures several results in the empirical literature on city size distributions via simulation. Section 5 concludes.

2 Seeing a Power Law in Populations

The appearance of a power law-like distribution for city populations is a well-documented feature of human geography. As articulated in Gabaix (1999b), the regularity of its appearance across countries, the definition of a "city," and time means it can reasonably be held as a minimum criterion for a model of cities. This distribution is typically illustrated with a simple plot and accompanying regression. For some truncation of the population distribution to include only the most populous locations ("cities"), the plot of the log population rank of a city and the log population of the city often appears strikingly linear, and a regression given by:

$$\ln(city\,rank_i) = \theta_0 + \theta_1 \ln(city\,pop_i) + \epsilon_i \tag{1}$$

Figure II:
Log Population, U.S. Census Places in 2010



Notes: The left panel is a histogram of log population for U.S. incorporated places and census designated places in 2010, with an overlaid normal distribution matching the moments of the empirical log distribution. The left panel shows a QQ plot of the (normalized) empirical population distribution against a standard normal distribution. The fit of the log population to the normal distribution distribution means the population distribution appears lognormal. This right panel of this figure is an update to Figure 2 of Eeckhout (2004), which uses data from the 2000 Census.
Data Source: U.S. Census

for many countries delivers a high R^2 (over 0.95) and frequently an estimate for θ_1 near -1, as in Figure I for U.S. cities. This slope is characteristic of a specific power law referred to as Zipf's law, which can be stated as the largest city in a given country being n times the size of the n^{th}-largest city. Interpreting this regression as describing the true city size distribution would mean that city populations follow a Pareto distribution with shape parameter $\alpha_P = 1$ and minimum city size x_m, reflecting the choice of truncation point.[10]

Instead of being a true power law, the city size distribution may be result from cities being a subset of full population distribution which appears similar to a Pareto distribution for tail observations. Eeckhout (2004) demonstrates that the full population distribution for

[10]The estimate of $\theta_1 = -1$ means the power law is such that for size X, the probability that a city is larger than X is proportional to $\frac{1}{X}$. A Pareto distribution with shape parameter $\alpha_P = 1$ and minimum city size x_m gives the necessary $P(x > X) = \frac{x_m}{X}$, which is the Pareto counter-cumulative distribution function for this α_P. The link between the (log) rank-size plot and the Pareto distribution is established in more detail in Gabaix (2009).

the U.S. appears lognormal. We construct an update to one of the key figures of Eeckhout (2004) in Figure II, which shows that this continues to hold for the U.S. in 2010. The left panel shows a histogram of log population, which closely matches the overlaid normal distribution matching the mean and standard deviation of the empirical distribution, and the right panel shows close fit to the normal distribution in a quantile-quantile (QQ) plot. The tail of a lognormal distribution often appears similar to a Pareto distribution, which can be understood by considering the lognormal PDF:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \tag{2}$$

After some algebra (given in Online Appendix A.1), this can be rewritten as:

$$f(x) = \Gamma_{LN} x^{-\alpha(x)-1} \tag{3}$$

where $\Gamma_{LN} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)$ and $\alpha(x) = \frac{\ln(x)-2\mu}{2\sigma^2}$. Contrast this with the PDF of a Pareto distribution:

$$j(x) = \Gamma_P \, x^{-\alpha_P - 1} \tag{4}$$

where $\Gamma_P = \alpha_P x_m^{\alpha_P}$ and the minimum city population is denoted x_m. The lognormal PDF in Equation 3 is similar to the Pareto PDF in Equation 4, but with a scale-varying "shape parameter"-like term. As Malevergne, Pisarenko, and Sornette (2011) show, provided the σ parameter is large, the value $\alpha(x)$ takes in the right tail will be stable over much of the tail distribution as $\alpha(x)$ grows logarithmically in x.

The Pareto interpretation of the tail of the population distribution appears dominant in the literature despite its limitations and the strict assumptions it necessitates. First, the Pareto distribution is taken to apply to only a subset of large settlements and not the full population distribution. This requires truncating a data series with no obvious truncation point. Early studies were limited to only the largest cities or settlements because of the comparative ease of accessing population counts for the largest places.[11] With more complete

[11] This is true of early work, such as Auerbach (1913) (while Auerbach had data on many small settlements, a table in his paper includes just the 94 largest; see the recent translation in Auerbach and Ciccone (2023))

data on population distributions, the choice of a truncation point to support the Pareto interpretation becomes critical and there is no widely accepted method for determining such a cutoff. Many researchers rely on a visual test of the data to determine a cutoff (Gabaix, 2009). Second, beyond the need to truncate the data to fit a Pareto, models generating a Pareto population distribution must rely on unrealistic assumptions regarding city growth dynamics. For example, Gabaix (1999b) obtains a Pareto distribution by assuming that cities cannot fall below a certain minimum size, such that the otherwise random growth process is "reflected" at the lower bound. Third, the systematic deviation of the far right tail from the Pareto observed in many countries (evident in Figure I for the U.S.) are very large in magnitude, which is obscured by the log-log scale. In Online Appendix B, we show that the estimated Pareto exponent implies a cumulative absence of nearly 76 million people from the 250 U.S. MSAs in Figure I, roughly a quarter of the U.S. population.[12]

The lognormal interpretation's attractive properties stand in direct contrast to the shortcomings of the Pareto. The lognormal distribution appears to fit both the body of the population distribution as well as the right tail, obviating the need for arbitrary truncation. Further, the scale-varying "shape parameter"-like term of the lognormal (as shown in Equation 3) can explain commonly observed deviations in real-world city size distributions. The likelihood of very large cities is lower when the true distribution is lognormal than for a similar Pareto, because the scale-varying "shape parameter"-like term is increasing in x. This appears to match the global city distribution (Rossi-Hansberg and Wright, 2007), as the largest cities in most countries tend to fall below the slope of the illustrative power law regression line.[13] Other features of real-world population distributions, such as the sensitiv-

and Zipf (1949). Even more recent investigation of Zipf's law in Krugman (1996), for instance, included just the top 135 cities as the *Statistical Abstract of the United States* included only those cities (Eeckhout, 2004).

[12]Alternative estimates of the power law based on different truncation points imply as many as 500 million "missing" people in the largest U.S. MSAs, substantially more than the entire U.S. population, which we discuss in Online Appendix B.

[13]Proponents of the Pareto interpretation have attempted to accommodate this divergence by arguing that the forces acting on small cities are different from those acting on large cities, generating different power laws for different sizes of cities. A lognormal distribution naturally exhibits this deviation without the need to treat subsets of the distribution differently. We provide a further discussion of the scale variance of the lognormal distribution and its contrast with the Pareto distribution in Online Appendix B.

ity of the estimated slope to the choice of truncation point, are consistent with the lognormal distribution as well.[14]

The appearance of a power law-like city population distribution is likely the result of focusing on the tail of the true lognormal distribution of human populations. Such an interpretation requires fewer restrictive assumptions and appears to better fit the observed data, both in the body of the population distribution (which is necessarily ignored by the Pareto interpretation) and in the tail (which behaves more lognormal than Pareto). In the following sections, we demonstrate that the equilibrium population distribution within many spatial models is lognormal given a realistically modeled geography.

3 Lognormal Populations in Spatial Models

We now describe a canonical quantitative spatial model within which we demonstrate our key result. We use a discretized version of the model in Allen and Arkolakis (2014), which nests a broad class of spatial models, and show that a realistic modeling of geography and trade will lead to lognormal population distributions in equilibrium.[15]

3.1 A Quantitative Spatial Equilibrium Model

The world consists of locations indexed by $i \in \mathcal{N}$, where $\mathcal{N} = \{i \mid i \in \mathbb{N}, 1 \leq i \leq N\}$. The index i reflects each location's position in a one-dimensional geography arranged as a circle. An example of the geography is included in Figure III. We adopt a circular geography here to simplify the exposition but we show in the appendix that the result generalizes to a

[14]This property is discussed at length in Eeckhout (2004) and demonstrated in Online Appendix Figure A.II where we expand or reduce the number of cities relative to Figure I. The sensitivity to the truncation point and the lack of a reliable rule for truncating the distribution suggest that the frequently estimated -1 exponent is unlikely to be a meaningful feature of the data. For some truncation of tail observations drawn from many lognormal distributions, the log-rank log-population plot will appear to take a slope of -1 as the exponent in Equation 3 diverges smoothly.

[15]The Allen and Arkolakis (2014) model is based on the two location model presented in Helpman (1998), generalized to an arbitrary number of locations.

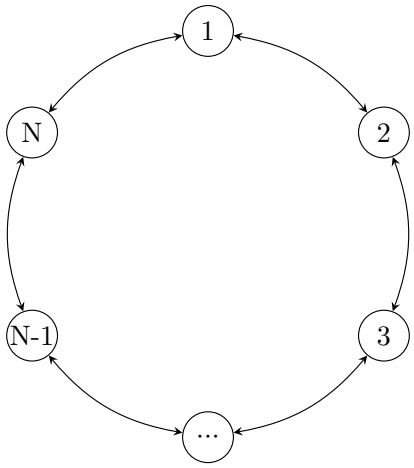


Figure III: An illustrative example of the circular geography we consider in the main text. Locations $i \in \mathcal{N}$ are arrayed around a circle and trade occurs along the perimeter of the circle.

two-dimensional Euclidean space.[16]

We assume that adjacent locations are evenly spaced at a distance of 1 around the circumference of the circle. The distance between any pair of locations $i, n \in \mathcal{N}$, $i \neq n$, denoted $\delta_{i,n}$, is the minimum distance around the circumference of the circle between the two locations in either the clockwise or counterclockwise direction. Without loss of generality, assume that $i \leq n$. Then we can write the minimum distance traveled $\delta_{i,n}$ as

$$\delta_{i,n} = \min(\underbrace{n - i}_{clockwise} , \underbrace{N - n + i}_{counterclockwise}). \tag{5}$$

The distance between locations is important for accurately reflecting correlations across space which often depend on distance. The similarities of locations nearby in space reflects the "first law of geography," formulated by Tobler (1970) as "everything is related to everything else, but near things are more related than distant things."

Trade between locations is costly, and trade costs are related to but distinct from distance. Travel between adjacent locations incurs a cost t_i, defined such that t_i for $i \in \{1, 2, ...N-1\}$ is the cost of traveling over the segment between i and $i+1$ and for $i = N$ is the cost of traveling over the segment between N and 1. All t_i are strictly positive and finite-valued random variables drawn from some common distribution, reflecting variation in geography

[16] In our simulated results we also use a two-dimensional geography.

whereby traversing some geographies is more difficult than traversing others. The realized trade costs $\tau_{i,n}$ between any pair of locations $i, n \in \mathcal{N}$, assuming without loss of generality that $i \leq n$, is the minimum trade cost when summing over the realized travel costs around the circle in the clockwise and counterclockwise directions between the two locations:

$$\tau_{i,n} = 1 + \min\left(\sum_{j \in A} t_j \, , \sum_{k \in B} t_k\right), \quad A = \{j | j \in \mathcal{N}, j \in [i, n)\}, B = \{k | k \in \mathcal{N}, k \notin A\}. \quad (6)$$

This structure imposes symmetric iceberg trade costs such that $\tau_{i,i} = 1$ for all $i \in \mathcal{N}$, $\tau_{s,t} = \tau_{t,s}$, $\tau_{s,t} > 1$ for $s \neq t$, and ensures that trade costs between any pair of locations $s, t \in \mathcal{N}$ are bounded above by a positive number. The circle geography ensures all locations face the same expected average trade costs. By construction, expected trade costs between locations are increasing in distance but reflect a notion of "effective" distance relevant for trade between locations rather than measured geographic distance relevant for considering the similarity of nearby locations.

Each location has an exogenous productivity fundamental A_i and an exogenous amenity fundamental U_i, both of which are strictly positive, real-valued random variables drawn from some probability distribution. We discuss the fundamentals further later in this section, where we argue that these should reflect variation in the geographic attributes of a place and the spatial correlation patterns of attributes across space. A location's effective productivity and amenity value may also be affected by negative or positive externalities due to the local population L_i. We define the "composite fundamentals" as:

$$\tilde{U}_i = U_i L_i^{\beta} \quad (7)$$

$$\tilde{A}_i = A_i L_i^{\alpha} \quad (8)$$

where the typical case will consist of $\beta < 0$ and $\alpha > 0$, reflecting the negative impact of overcrowding on amenities and positive productivity spillovers from agglomeration. Geography within this model is represented by the set of functions defining the locational fundamentals, $\tilde{A}_i$ and $\tilde{U}_i$, along with the trade costs function τ defining the spatial relationship between

locations in the model.

We assume Armington-style production of a differentiated good in each location. There is a population of homogeneous workers $\bar{L} \in \mathbb{R}_{++}$ who can freely move to any location. Workers have common constant elasticity of substitution preferences over goods in their welfare function given by:

$$W_i = \left(\sum_{n \in \mathcal{N}} q_{n,i}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tilde{U}_i \tag{9}$$

where $\tilde{U}_i$ is the composite amenity fundamental of location i and $q_{n,i}$ denotes the consumption in i of the good produced in n and $\sigma > 1$ governs the elasticity of substitution.

Production is perfectly competitive.[17] A worker in location i can produce $\tilde{A}_i$ units of the local differentiated good, where $\tilde{A}_i$ is the composite productivity fundamental of location i. The number of workers and wages in a location are given by the functions $L : N \to \mathbb{R}_{++}$ and $w : N \to \mathbb{R}_{++}$.[18]

Based on the CES assumption, we can write the amount of each good produced in any location i consumed in location n as:

$$q_{i,n} = Q_n \left(\frac{p_{i,n}}{P_n} \right)^{-\sigma} \tag{10}$$

where Q_n is aggregate consumption in n and P_n is the price index in location n, given by:

$$P_n = \left(\sum_{i \in \mathcal{N}} p_{i,n}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{11}$$

Given the assumption of perfect competition, $p_{i,n}$ can be expressed as:

$$p_{i,n} = \frac{\tau_{i,n} w_i}{\tilde{A}_i} \tag{12}$$

Combining the quantity (Equation 10) and price (Equation 12) expressions, we can write

[17] This model nests cases of monopolistic competition, as demonstrated in the appendix to Allen and Arkolakis (2014).

[18] No location will be unpopulated or offer zero wages in equilibrium given the range of parameters we consider.

the value of the good produced in i consumed by n as:

$$X_{i,n} = \left(\frac{\tau_{i,n} w_i}{\tilde{A}_i P_n}\right)^{1-\sigma} w_n L_n \tag{13}$$

By the CES assumption we can express welfare in each location as:

$$W_i = \frac{w_i}{P_i} \tilde{U}_i \tag{14}$$

The value of income in a location must be equal to the value of production:

$$w_i L_i = \sum_{n \in \mathcal{N}} X_{i,n} \tag{15}$$

The labor market clears:

$$\sum_{n \in \mathcal{N}} L_n = \bar{L} \tag{16}$$

We can then combine the welfare expression (Equation 14), value of consumption expression (Equation 13), and income expression (Equation 15) to get:

$$L_i w_i^{\sigma} = \sum_{n \in \mathcal{N}} W_n^{1-\sigma} \tau_{i,n}^{1-\sigma} \tilde{A}_i^{\sigma-1} \tilde{U}_n^{\sigma-1} L_n w_n^{\sigma} \tag{17}$$

The welfare expression combined with the price index yields:

$$w_i^{1-\sigma} = \sum_{n \in \mathcal{N}} W_i^{1-\sigma} \tau_{n,i}^{1-\sigma} \tilde{A}_n^{\sigma-1} \tilde{U}_i^{\sigma-1} w_n^{1-\sigma} \tag{18}$$

We now focus on the case of the model with spillovers and externalities and impose our assumption of symmetric trade costs.[19] Given the form of the externalities in Equations 7 and 8, free movement between locations which ensures welfare is equal in all locations ($W_i = \bar{W}$ for all i), and symmetric trade costs we can combine Equations 17 and 18 into a single equation given by:

$$\bar{W}^{\sigma-1} L_i^{\tilde{\sigma}\gamma_1} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in \mathcal{N}} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \left(L_n^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \tag{19}$$

[19] Our results also hold in the model without spillovers ($\alpha, \beta = 0$).

where:

$$\tilde{\sigma} = \frac{\sigma - 1}{2\sigma - 1}, \quad \gamma_1 = 1 - \alpha(\sigma - 1) - \beta\sigma, \quad \gamma_2 = 1 + \alpha\sigma + (\sigma - 1)\beta$$

The existence and uniqueness of the equilibrium, and a mechanism for finding it, are established in Allen and Arkolakis (2014) when $\frac{\gamma_2}{\gamma_1} \in (-1, 1]$ (the discrete case is considered in their online appendix). We focus on this part of the parameter space, which occurs when $\alpha + \beta \leq 0$. For any realization of fundamentals and trade costs we thus can recover a unique vector of populations.

In the following subsections we demonstrate that the equilibrium population within this model will be lognormally distributed given a realistic modeling of the fundamentals and trade costs based on variation in geography. We begin by rewriting equation 19 in terms of the population in each location i in simplifying notation as:

$$L_i = \Omega_i \left(S_i\right)^{\frac{1}{\tilde{\sigma}\gamma_1}} \tag{20}$$

where $\Omega_i = \left(\bar{W}^{1-\sigma} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma}\right)^{\frac{1}{\tilde{\sigma}\gamma_1}}$ and $S_i = \sum_{n \in \mathcal{N}} s_{i,n}$, where $s_{i,n} = \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} L_n^{\tilde{\sigma}\gamma_2}$. We refer to Ω_i as the "own-fundamental" term of location i, as it consists only of location i's fundamentals and the common positive constant $\bar{W}$ and will reflect the role of "place" in determining the population of each location. We refer to S_i as the "market access" term of location i, as it is a sum over all locations $n \in \mathcal{N}$, and reflects the role of "space."

We now demonstrate that, given a realistic modeling of variation in geography across space and within a place, both terms will be lognormally distributed. This will allow us to show that the equilibrium population also follows a lognormal distribution.

3.2 Variation over Space and the Distribution of Market Access

We first show that the distribution of the market access term, which is a summation over positive random variables, converges in distribution to a lognormal. This result follows from the application of a central limit theorem that allows for a particular type of dependence structure across elements of the sequence in the summation that is likely to hold in spatial

contexts and a useful lemma due to Marlow (1967) for sums of positive random variables.

The market access term for each location i given by $S_i = \sum_{n \in \mathcal{N}} s_{i,n}$, consists of a summation over the random variables $s_{i,n}$, each a function of the random variables $\tau_{i,n}$, A_n, U_n, and L_n, for all $n \in \mathcal{N}$.[20] Many central limit theorems require that attributes are independent and identically distributed in order to establish that the distribution of a sum converges to a normal distribution. These assumptions are likely violated in this context and the sequences $\{s_{i,n}\}$ are unlikely to consist of independent random variables. The fundamentals of a location are intended to reflect its geographic advantages, as we argue in more detail in the following section, and as geography is similar for nearby locations the productivity and amenity fundamentals A_i and U_i are unlikely to be independent for nearby locations. Further, a location near a highly populous city will (provided trade costs are not prohibitive) benefit from access to the market of its neighbor such that the populations L_i of nearby locations are also unlikely to be independent. Lastly, trade costs from any i to any two other adjacent locations $j, j+1$, $\tau_{i,j}$ and $\tau_{i,j+1}$, will tend to be similar by construction as they differ by at most t_j, the trade cost of traversing the additional interval between j and $j+1$.

Despite the potential for correlations among nearby locations, we argue that over long distances the fundamentals, populations, and trade costs should approach independence. Nearby locations often have broadly similar geographic attributes, like the climactic similarities of New York City and northern New Jersey, but this spatial correlation declines with distance such that New York City is quite dissimilar from both Monterrey, Mexico, and Nuuk, Greenland.[21] In Online Appendix C.4 we demonstrate the similarity of locations across many geographic attributes, measured in terms of the correlation of attributes, declines as the distance between locations increases. Trade costs from any i to any pair of locations $j, k \neq i$, $\tau_{i,j}, \tau_{i,k}$, may also differ substantially when the distance between k and j is large as trade costs are not a pure function of distance but rather reflect random variation

[20] In Online Appendix A.4, we motivate the treatment of $s_{i,n}$ and L_n as random variables, because each L_n is an element of a random sequence corresponding to the eigenvector of a random matrix.

[21] Nuuk and Monterrey are roughly equidistant from New York City, both at a distance of 1,850 miles.

in the difficulty of travel. For example, it is possible that while $\delta_{i,j} < \delta_{i,k}$ such that i is closer geographically to j than to k, travel from i to k occurs clockwise around the circular geography and is, due to random variation in trade costs, substantially easier than travel in the counterclockwise direction from i to j such that $\tau_{i,j} \gg \tau_{i,k}$. Given independence in fundamentals and trade costs over long distances, the distribution of population can also be reasonably assumed to approach independence as distance increases. As such, for each i the realizations of $s_{i,j}$ and $s_{i,k}$, for locations $j, k \in \mathcal{N}$ are likely to near independence as distance between j and k increases.

Many modern central limit theorems allow for precisely this type of asymptotic independence. The formal assumption we must impose in order to apply a central limit theorem is that the sequences must be α-mixing for each $i \in \mathcal{N}$.[22] A formal definition of α-mixing for sequences is given in Online Appendix A.3. The concept requires that all events defined on arbitrary subsets of an α-mixing sequence approach independence as the "distance" between the subsets increases, where distance is reflected in the index of the sets. This concept is often used in the analysis of time series, where the index reflects the timing of the observation and imposes a natural concept of the distance between elements of the sequence.

In our setting, we also have a natural concept of distance, $d_{i,n}$. We can re-index the elements of the sequence $\{s_{i,n}\}$ in terms of distance by defining for each i an alternative sequence ordered by distance from i, $\{s^d_{i,j}\}$, where $s^d_{i,1} = s_{i,i}$ (at a distance of 0) is followed by $s^d_{i,2} = s_{i,i+1}$, $s^d_{i,3} = s_{i,i-1}$ (both at a distance of 1), and so on.[23] Note that as the sequences $\{s^d_{i,j}\}$ are simply a re-indexing of the sequences $\{s_{i,n}\}$, $S_i = \sum_{n \in \mathcal{N}} s_{i,n} = \sum_{j \in \mathcal{N}} s^d_{i,j}$ for all $i \in \mathcal{N}$. We assume that the sequences $\{s^d_{i,j}\}$ are α-mixing for all $i \in \mathcal{N}$ to reflect asymptotic independence with respect to distance.

Lemma 1 is a central limit theorem for α-mixing sequences due to Herrndorf, 1984, introduced to the spatial economics literature by Lee and Li (2013). It allows us to demonstrate

[22] This concept is also referred to as "strong mixing" or "weak dependence."

[23] Given that N is adjacent to 1 in the indexing, when considering $i = 1$ the value for $i - 1$ will be equal to N.

convergence of a sum over a sequence of non-i.i.d random variables to a normal distribution, provided the assumption of asymptotic independence (α-mixing) and other moment conditions for the sequence.

***Lemma 1 (Herrndorf (1984)**: Let $\{x_i\}$ be an α-mixing sequence of random variables satisfying the following conditions:*

i. $\mathbb{E}[x_i] = 0, \forall i$

ii. $\lim_{n\to\infty} \frac{\mathbb{E}\left[(\sum_{i=1}^{n} x_i)^2\right]}{n} = \bar{\sigma}^2, 0 < \bar{\sigma}^2 < \infty$

iii. $\sup_{i\in\mathcal{N}} \mathbb{E}[x_i^b] < \infty$, *for some* $b > 2$

iv. $\sum_{s=1}^{\infty} (\alpha_s)^{1-\frac{2}{b}} < \infty$

Let $X_n = \sum_{i=1}^{n} x_i$. *Then as* $n \to \infty$, $\frac{1}{\sqrt{n}\bar{\sigma}} X_n$ *converges in distribution to the standard normal distribution.*

The proof of Lemma 1 is given in Herrndorf (1984), and discussion of the conditions is given in Lee and Li (2013). Lemma 1 allows us to apply a central limit theorem to the sums S_i for all $i \in \mathcal{N}$, provided the sequences $\{s_{i,n}^d\}$ fulfill the requirements of the lemma.

An additional property of the summations S_i allows us to move from a central limit theorem in levels to one in logs, which will allow us to characterize the distribution of S_i as lognormal. Each $s_{i,n}$ must be strictly positive, as A_n, U_n, $\tau_{i,n}$, and L_n are strictly positive and so $s_{i,n} = \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} L_n^{\tilde{\sigma}\gamma_2} > 0$. This allows us to apply a useful lemma due to Marlow (1967), which provides conditions under which a lognormal distribution may appear from a summation of positive random variables:

***Lemma 2 (Marlow, 1967)**: Let $\{X_n\}$ be a sequence of positive random variables. Suppose there exist sequences of positive real numbers $\{a_n\}$ and $\{b_n\}$, and a distribution F such that*

i. *At each point of continuity of* F, $\lim_{n\to\infty} P\left\{\frac{X_n - a_n}{b_n} \leq x\right\} = F(x)$

ii. $\lim_{n\to\infty}\left(\frac{b_n}{a_n}\right)=0$

Then at each point of continuity of F, $\lim_{n\to\infty} P\left\{\left(\frac{a_n}{b_n}\right)\ln\left(\frac{X_n}{a_n}\right)\leq x\right\}=F(x)$

The proof of Lemma 2 is given in Marlow, 1967. Condition (i) can reflect convergence under a central limit theorem (such as Lemma 1), where $F(x)$ is the standard normal distribution and the sequences a_n and b_n are the mean and standard deviation of some X_n resulting from a sum of random variables. Condition (ii) then necessitates that the coefficient of variation (the ratio of the standard deviation to the mean) of X_n is zero in the limit as n grows large. Many sums of positive random variables fulfill this requirement and examples are given in Online Appendix A.2.

For a sum that satisfies the conditions for a central limit theorem and condition (ii), Lemma 2 states that the given normalization of the sum will converge in distribution to a lognormal random variable.[24] Lemma 2 is crucial for understanding the population distribution within many spatial equilibrium models, as these models often incorporate a notion of "market access" via a trade cost-weighted sum over all locations. All of the elements of these sums must be strictly positive, and provided these fulfill the conditions necessary for applying a central limit theorem and Lemma 2 the distribution of these "market access" terms will approach a lognormal distribution as the number of locations grows large. We state this result in Proposition 1.

Proposition 1: *For each* $i\in\mathcal{N}$, *define the sequences* $\{s_{i,1}^d, s_{i,2}^d, ...s_{i,N}^d\}$ *and the demeaned sequences* $\{\hat{s}_{i,n}^d\}$ *such that* $\mathbb{E}[\hat{s}_{i,n}^d]=0$ *for all* $n\in\mathcal{N}$. *Define* $S_i^{(N)}=\sum_{n=1}^N s_{i,n}^d$, $M_i^{(N)}=\sum_{n=1}^N \mathbb{E}[s_{i,n}^d]$, *and* $\sigma\left(S_i^{(N)}\right)^2=Var[S_i^{(N)}]$. *If, for all* i,

i. *The sequences* $\{\hat{s}_{i,1}^d, \hat{s}_{i,2}^d, ...\hat{s}_{i,N}^d\}$ *are* α*-mixing and fulfill the conditions in Lemma 1*

ii. *The coefficients of variation associated with the sequences* $\{S_i^{(1)}, S_i^{(2)}, ..., S_i^{(N)}\}$ *fulfill*

[24] We discuss Lemma 2 further in Online Appendix A.2. Beyond the context in which we apply the Marlow (1967) lemma, it appears to have broad usefulness within economics. For example, a CES aggregator over positive random variables fulfilling the conditions of the lemma should approach lognormality as the lognormal distribution is preserved over exponentiation.

condition (ii) of Lemma 2 as $N \to \infty$

then the distribution of $\frac{M_i^{(N)}}{\sqrt{N}\sigma\left(S_i^{(N)}\right)} \ln\left(\frac{S_i^{(N)}}{M_i^{(N)}}\right)$ *converges in distribution to a lognormal for all* i *as* $N \to \infty$.

The proof follows directly from Lemma 1 and Lemma 2. Lemma 1 allow us to apply an α-mixing central limit theorem to the sum S_i, and Lemma 2 allows us to move to a central limit theorem in logs as all elements of this sum are positive.[25] If the necessary assumptions on $s_{i,n}$ hold, the market access summation S_i converges in distribution to a lognormal as N grows large. We assume S_i is lognormally distributed for some sufficiently large N and as a lognormal raised to a power results in another lognormal distribution each $(S_i)^{\frac{1}{\tilde{\sigma}\gamma_1}}$ will also be lognormally distributed.

3.3 Variation within Place and the Distribution of Fundamentals

We now turn to modelling the distribution of fundamentals, and begin by noting that these are intended to reflect differences in the suitability of a place for habitation or settlement. There is clear evidence for observable geographic attributes, alone and in combination, playing a role in shaping human settlement patterns (Henderson et al., 2018). The substantial differences between areas of high population in terms of many geographic attributes suggests that no one particular observable attribute alone is a sufficient proxy for what makes a location good for human habitation and that many attributes should contribute to the quality of a place.

We model fundamentals as resulting from random shocks via the geographic attributes of a location, building on the approach of Lee and Li (2013) who similarly model a location's quality as resulting from many random "factors." Using random variation in geography to model locational fundamentals is the cross-sectional analog of random growth models based

[25] Interestingly, an implication of Proposition 1 is that the distribution of S_i should appear normal in both levels and logs. We discuss this further in Section 4 and show that this does indeed hold in Online Appendix D.2.

on Gibrat's law over time, like those of Gabaix (1999a) and Eeckhout (2004), where rather than productivity shocks occurring over time each attribute within a location contributes a productivity or amenity shock to the respective fundamental.

Each location i has many geographic attributes a_{ig}, which are indexed by $g \in \mathcal{G}$, where $\mathcal{G} = \{g \mid g \in \mathbb{N}, 1 \leq g \leq G\}$. Attributes for a productive location could be fertile soil, regular and mild weather patterns, and favorable topography, among many others. We assume all attributes are strictly positive in value and for any g, higher values of a_{ig} reflect *better* realizations of that attribute.[26] We also assume each individual attribute is drawn from a common distribution in all locations, while different attributes may differ in their respective distribution.

For brevity we focus our discussion on productivity fundamentals before returning to consider the similarly constructed amenity fundamentals. The locational productivity fundamental for a location i, denoted A_i, should be a function of its many attributes such that $A_i = F_A(a_{i1}, a_{i2}, ..., a_{iG})$. This function should be increasing in each a_{ig}, to reflect that better attributes increase productivity, such that $\frac{\partial F_A}{\partial a_{ig}} > 0$ for all $g \in \mathcal{G}$. Further, the aggregating function should exhibit complementarities between each of the attributes—the benefit of having reliable rainfall for production is increased when there is better arable land in a location, for instance. This means the aggregating function also needs a positive cross-partial for all arbitrary combinations of attributes, such that $\frac{\partial^2 F_A(\cdot)}{\partial a_{ij} \partial a_{ig}} > 0$, for $j, g \in \mathcal{G}, j \neq g$.

Consistent with these assumptions, we can view the contribution of attributes to the fundamental as representing multiplicative shocks. We assume a Cobb-Douglas form for the aggregating function F_A, consistent with the requirements outlined above. The varying importance of different attributes can be reflected by the exponents $\xi_g > 0$ associated with

[26] These should not be thought of as being measured in the familiar units for each attribute. Rainfall in inches has a nonlinear relationship with agricultural output, for instance. We instead want a measure reflecting how positive the "shock" from a given attribute is.

each g:[27]

$$A_i = \prod_{g \in \mathcal{G}} {a_{ig}}^{\xi_g} \tag{21}$$

Taking the natural log yields the following expression:

$$\ln(A_i) = \sum_{g \in \mathcal{G}} \xi_g \ln a_{ig} \tag{22}$$

where we express the logged fundamental in location i as a sum of random variables $\xi_g \ln a_{ig}$.

It is possible that some attributes a_{ig} are not independent within locations. For example, high July temperatures and the number of growing days may be correlated within places such that knowing the realization of one is informative about the likely values of the other. However, over the large number of attributes of a place there do appear to exist pairs of attributes which appear nearly independent within locations (e.g., topography and rainfall), as we demonstrate in Online Appendix C.4 using data on geographic attributes from Henderson et al. (2018).

As in Lee and Li (2013) we assume that attributes are α-mixing within locations and define the ordering of the sequence of attributes $\{a_{i1}, a_{i2}, ..., a_{iG}\}$ such similar attributes are close together (July temperatures and growing days have indicies near each other), while independent attributes differ greatly in their index values (rainfall and topography have indices set far apart). Together with further restrictions, given in Lemma 1, on the moments of the random variables $\xi_g \ln a_{ig}$ and the rate of α-mixing, which determines how rapidly the elements of the sequence approach independence, we apply the central limit theorem in Lemma 1 to characterize the distribution of each $\ln A_i$.

Proposition 2: *Define* $\widehat{\xi_g \ln a_{ig}} = \xi_g \ln a_{ig} - \mathbb{E}[\xi_g \ln a_{ig}]$ *for all* $g \in \mathcal{G}$ *and* $i \in \mathcal{N}$, *and define* $\widehat{\ln A_i} = \sum_{g \in \mathcal{G}} \widehat{\xi_g \ln a_{ig}}$ *and* $\sigma\left(\widehat{\ln A_i}\right)^2 = Var[\widehat{\ln A_i}]$ *for all* $i \in \mathcal{N}$. *If the sequences* $\{\widehat{\xi_g \ln a_{ig}}\}$ *are* α*-mixing and fulfill the conditions of Lemma 1 for all* $i \in \mathcal{N}$, *then as* $G \to \infty$,

[27] We could also include an index for the time t, to allow for attributes a_{igt} to vary over time and to vary in their importance over time ξ_{gt}, which could capture structural transformation of the economy or changing production technologies at time t. In this case, $A_{it} = \prod_{g \in \mathcal{G}} {a_{igt}}^{\xi_{gt}}$ where the fundamentals can vary with t.

$\frac{1}{\sqrt{G}\sigma\left(\widehat{\ln A_i}\right)}\widehat{\ln A_i}$ *converges in distribution to the standard normal distribution for all* $i \in \mathcal{N}$.

Proposition 1 follows from Lemma 1. By Proposition 1, as the number of attributes grows large, the log productivity fundamental $\ln A_i$ will converge in distribution to a normal distribution and so A_i will converge in distribution to a lognormal distribution. We assume that, for a large number of attributes, A_i will be lognormally distributed based on this asymptotic argument.

The amenity fundamental is defined similarly, but we allow for different weights as the attributes most relevant for determining quality of life may differ from those influencing productivity. The log of the amenity fundamental, which has weights given by $\iota_g > 0$, is:

$$\ln(U_i) = \sum_{g \in \mathcal{G}} \iota_g \ln a_{ig} \tag{23}$$

and, given the same conditions as on the productivity fundamental, will also converge in distribution to a lognormal as the number of attributes grows large.

We provide support for the appearance a lognormal distribution for fundamentals in two ways. First, the fundamentals recovered by inverting the model in Allen and Arkolakis (2014), plotted in Online Appendix C.6, appear lognormal for U.S. counties. Second, we use the Henderson et al. (2018) data on attributes and, following our modeling assumptions, calculate a "naive" fundamental by applying our aggregating function.[28] The plots are reported in Online Appendix C.6, and show that aggregating the eleven attributes results in a distribution of fundamentals that appears lognormal.[29] While only suggestive, both the strategy of recovering fundamentals within a structural model based on true populations and the construction of a "naive" fundamental from attributes result in strikingly lognormal distributions for fundamentals.

[28] We say the fundamental is "naive" in the sense that we do not know the appropriate weights or scaling of the attributes. The construction of the fundamental is discussed in Online Appendix C.5. A similar exercise was earlier done by Behrens and Robert-Nicoud (2015) for U.S. MSAs.

[29] Notably, the limited number of attributes we have exhibit a reasonably high degree of correlation within place but still generate a lognormal distribution when aggregated as described in this section. Behrens and Robert-Nicoud (2015) showed that as few as six attributes can be aggregated to produce a fundamental that appears lognormal.

Given lognormally distributed productivity and amenity fundamentals A_i and U_i, we can show that the "own-fundamental" term Ω_i for each location i will be lognormally distributed.

***Proposition 3**: If A_i and U_i have a bivariate normal distribution in logs, Ω_i will be lognormally distributed.*

Proposition 3 follows immediately from the properties of lognormals, as either raising a lognormal distribution to a power or multiplying by a positive constant result in another lognormally distributed random variable and multiplying two lognormal random variables which have a bivariate normal distribution in logs results in another lognormally distributed random variables.

We here formalize the discussion in the prior section distribution of the exogenous productivity and amenity fundamentals across space by assuming that the sequences $\{A_i\}$ and $\{U_i\}$ are α-mixing with respect to distance.[30] This captures the potential geographic similarity of nearby locations and how this similarity vanishes with increasing distance, as discussed in the prior section. The spatial correlation patterns of geographic attributes and the "naive" fundamentals we constructed also indicate declining correlation towards 0 over distance as shown in in Online Appendix C.6. We take this as support for our decision to model locational fundamentals as lognormal and α-mixing with respect to distance.

3.4 The Distribution of Population

We have now shown that Ω_i and $(S_i)^{\frac{1}{\bar{\sigma}\gamma_1}}$ are lognormally distributed when realistically modeled to reflect variation in geography and trade costs over space. Using this, we can then show that the population will be lognormally distributed as population in each location i is given by $L_i = \Omega_i(S_i)^{\frac{1}{\bar{\sigma}\gamma_1}}$. This result is given in Theorem 1.

[30] We extend the concept of α-mixing to include two-dimensional random fields in Online Appendix A.3. The complication involves the need to introduce a concept of distance when this information is not directly encoded in the series index. All of our proofs generalize to the two-dimensional case.

Theorem 1: *If Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ have a bivariate normal distribution in logs for all $i \in \mathcal{N}$, then L_i follows a lognormal distribution for all $i \in \mathcal{N}$.*

The proof follows directly from the lognormality of Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$, and the property that products of lognormal distributions which have a bivariate normal distribution in logs also follow a lognormal distribution.

For the realized population distribution to appear lognormal we further require that populations asymptotically independent with respect to distance. While we have proven that the distribution of the population in each i approaches a lognormal it is possible that the equilibrium population is highly correlated across locations such that the full probability distribution is not realized. We need to demonstrate that the population distribution is also α-mixing, which will ensure that due to the asymptotic independence of the population over space the realized population distribution will reflect the lognormality of the underlying distribution for large N.[31]

Proposition 4: *If $\{A_i\}$ and $\{U_i\}$ are α-mixing and independent sequences, then $\{\Omega_i\}$ is an α-mixing sequence.*

The proof follows directly from Theorem 5.2 of Bradley (2005), as α-mixing is maintained over measurable transformations and over combinations of independent α-mixing random sequences.[32] We must make the stronger assumption that A_i and U_i are independent, but given α-mixing of these sequences independence establishes that $\{\Omega_i\}$ will also be α-mixing.[33]

We can then show that the population sequence $\{L_i\}$ will be α-mixing as well, given an additional assumption that the market access terms S_i are also α-mixing. This will ensure

[31]The stronger concept of α-mixing implies ergodicity (Frigg, Berkovitz, and Kronz, 2020). An ergodic process will visit all parts of the probability space associated with the process given large N.

[32]In Online Appendix A.3 and A.5, we extend Theorem 5.2 of Bradley (2005) from random sequences to random fields in two-dimensional space.

[33]This assumption is necessary for the application of Theorem 5.2 of Bradley (2005) in the proof of Proposition 3, but we will relax this assumption in our simulations and show that it does not appear necessary for the result. Allen and Arkolakis (2014) find a low, but non-zero, correlation of 0.12 between A_i and U_i.

that the full probability space of the population distribution will be realized for some large N.

Theorem 2: *If $\{\Omega_i\}$ and $\{S_i\}$ are independent α-mixing sequences, then the population sequence $\{L_i\}$ is α-mixing.*

The proof also directly from Theorem 5.2 of Bradley (2005). In Online Appendix A.5 we show that these α-mixing properties for sequences can be generalized to two dimensional fields and also hold for more realistic geographies.

4 Results

We now demonstrate that the model successfully generates lognormal population distributions and power law-like city size distributions via simulation. We provide comparative statics based on varying parameter values across simulations to document how changes in congestion, spillovers, and trade costs influence the observed power law. We identify changes to the power law in the directions implied by the empirical literature. Finally, we show that Gibrat's law holds within the model when the aggregate population increases, showing that size-invariant growth is a feature of a lognormal equilibrium population distribution based on variation in geography and trade, in contrast to earlier literature that took random growth to be the basis for lognormal populations.

4.1 Simulation of the Population Distribution

We first simulate the model to demonstrate that the resulting populations are indeed lognormally distributed and that the city size distribution appears to follow a power law. We simulate the results within a two dimensional geography. Each location in the model is a place that can host a settlement.[34] Note that the definition of a "location" is not specified

[34]This interpretation matches that in Redding and Rossi-Hansberg (2017), which frames locations as regions which can potentially hold a single settlement.

by the model, beyond being a place within which local spillovers occur, and these could either be large regions or small locales.[35] We define the most populous 5% of locations as "cities" within the model, to demonstrate that the tail behavior of the resulting population distribution mirrors the appearance of a power law in empirical city size distributions.

To ensure dispersion in trade costs while remaining consistent with the triangle inequality and maintaining symmetry, we model settlements as occurring randomly over a large surface and take the Euclidean distance between all settlements.[36] We simulate a large geography and take the central locations as the geography of interest to limit the impact of border effects on the population distribution. We are left with a central geography consisting of approximately 20,000 settlements.[37]

We take model parameters from the literature and from Allen and Arkolakis (2014) where possible. We first create randomly-generated draws of exogenous productivity and amenity fundamentals with declining spatial correlation. Fundamentals are drawn from lognormal distributions with parameters $\sigma_{LN} = 1$ and $\mu_{LN} = 0$, and we induce spatial correlation in the fundamentals using a Choleski decomposition.[38] We allow the productivity and amenity fundamentals in a location to be correlated and set the correlation between A_i and U_i within each location i to $\rho_{AU} = 0.12$ to match the correlation between the recovered productivity and amenity fundamentals in Allen and Arkolakis (2014).

The magnitude of local productivity spillovers is given by $\alpha = 0.03$, in line with the

[35]That the model is ambiguous on the level of aggregation of population means it is consistent with the observation of the characteristic population distribution at varying levels of aggregation, as in Holmes (2010), Rozenfeld et al. (2011), and Mori, Smith, and Hsu (2020). Spillovers are unlikely to be purely local at any level of aggregation, and so differing levels of aggregation may require different parameters governing "local" spillovers that are a useful abstraction from more complex patterns of spillovers.

[36]An example of a portion of this "dartboard" geography can be seen in Online Appendix D.1. We could alternatively model trade costs as having an idiosyncratic component to ensure dispersion, and not respect the triangle inequality or symmetry. We choose the more restrictive setting without an idiosyncratic component for our main results to demonstrate that only a limited degree of dispersion is necessary. We simulate with additional idiosyncratic shocks to trade costs in Online Appendix D.4.

[37]We uniformly distribute 30,000 settlements across a 1200-by-1200 grid and discard those within 100 cells of a border. This leaves an expected number of settlements of $\frac{100}{144} * 30000 = 20,833.\bar{3}$. We draw a new distribution of randomly drawn distributions of fundamentals each simulation.

[38]We assume the degree of spatial correlation of the log-scale fundamental declines exponentially, consistent with the empirical attribute correlations we show in Online Appendix C.6, so that $\rho_{ij} = e^{-\delta_\rho d_{ij}}$. For $j = i$, this gives $\rho_{ii} = 1$ as $d_{ii} = 0$. We set $\delta_\rho = 0.5$.

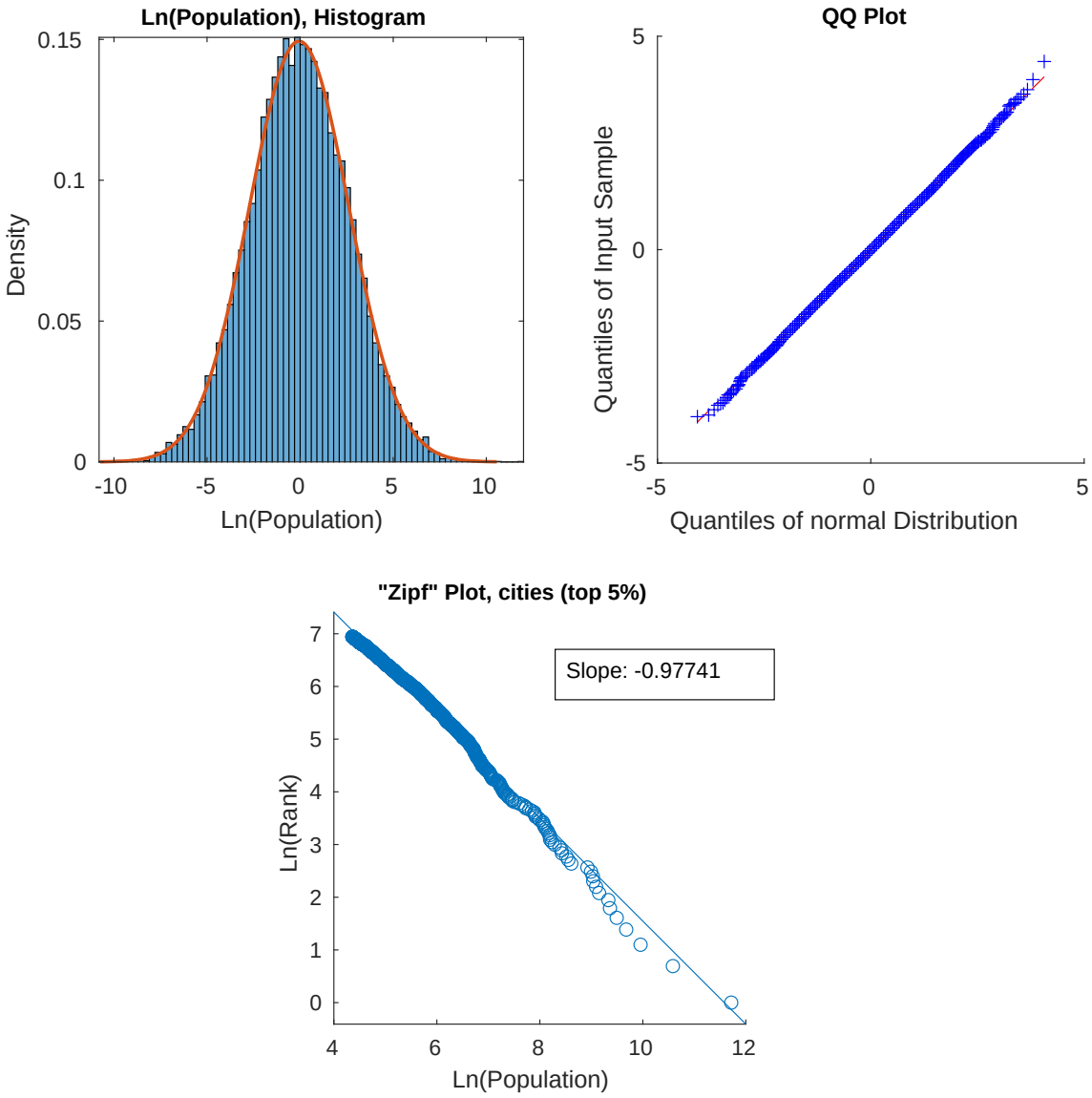
estimates in Combes, Duranton, and Gobillon (2008) and those surveyed in Rosenthal and Strange (2004) and Combes and Gobillon (2015).The model contains an isomorphism which we use to parameterize congestion costs. As discussed in Allen and Arkolakis (2014), the model is isomorphic to one with a fixed quantity of housing where spending on housing is δ and $\beta = -\frac{\delta}{1-\delta}$. Congestion costs are parameterized to match a level of spending on housing of 25% of income, which gives a congestion parameter of $\beta = -\frac{1}{3}$, consistent with the estimates in Combes, Duranton, and Gobillon (2019) and Davis and Ortalo-Magné (2011). We model trade costs as an exponential function of distance, $\tau_{ij} = e^{\delta_{TC} d_{ij}}$ and set $\delta_{TC} = 0.001$.[39] We vary these parameters in Online Appendix D.3. We set the elasticity of substitution $\sigma = 5$, as in Redding and Rossi-Hansberg (2017) and consistent with the estimates in Simonovska and Waugh (2014).

Figure IV shows the equilibrium population distribution associated with a random draw of productivity and amenity fundamentals and trade costs.[40] The log of the population distribution very closely matches the overlaid normal distribution, demonstrating lognormality. The upper-right panel shows a quantile-quantile (QQ) plot of the log population and a normal distribution, demonstrating very close fit throughout the full distribution. The output in the upper panels is very similar to Figure II for the full U.S. population distribution. Given the lognormality of the population distributions, the most populated locations in our model will appear to follow a power law distribution as demonstrated in Section 2. Concentrating only on the most populated 5% of locations, in the bottom panel of Figure IV we find the model generates power law-like city size population distributions like those commonly identified in the data. Indeed, this panel is similar to Figure I for U.S. cities and the log rank-size regression on this simulated data gives a slope of -0.977, close to the classic Zipf's

[39] Allen and Arkolakis (2014) find that the cost of road travel is .56 when the width of the continental U.S. is normalized to 1, given an elasticity parameter of $\sigma = 9$. The geography we simulate has a width of 1000, which would imply a scaled parameter of 0.00056, near the value we choose. Note that we selected an elasticity parameter of $\sigma = 5$, roughly half the value used by Allen and Arkolakis (2014) (who adopt $\sigma = 9$) in their estimation of trade costs.

[40] Whenever we present a single random draw, we present the first draw using our seed value. Our seed value is the four-digit catalog number of Columbia Economics Professor Serena Ng's graduate macroeconometrics course.

Figure IV:
Example of the Equilibrium Population Distribution



Notes: The top left panel shows the model's log population appears to follow a normal distribution. The top right panel contains a QQ plot of the model's log population distribution, indicating that it very closely matches a normal distribution. The power-law plot at the bottom shows a strong resemblance to the typical log rank-size plot along with the characteristic divergence of the largest locations below the trendline.

law result of a slope of -1 for this particular random draw of fundamentals.

In Online Appendix D.2, we also check that the summation term S_i behaves as predicted

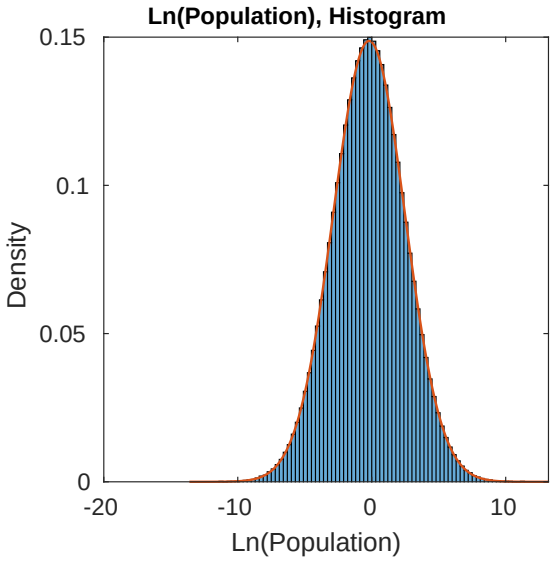
by Proposition 1 and the conditions in Theorem 2 which imply that this term should appear normal in both levels and logs and be independent from the own-lognormal term for each location. We find evidence that both of these hold predictions within our simulations. When we induce further idiosyncratic variation in trade costs in Online Appendix D.4 the fit of S_i to the normal distribution in both levels and in logs improves, consistent with the importance of dispersion in trade costs for the distribution of market access.

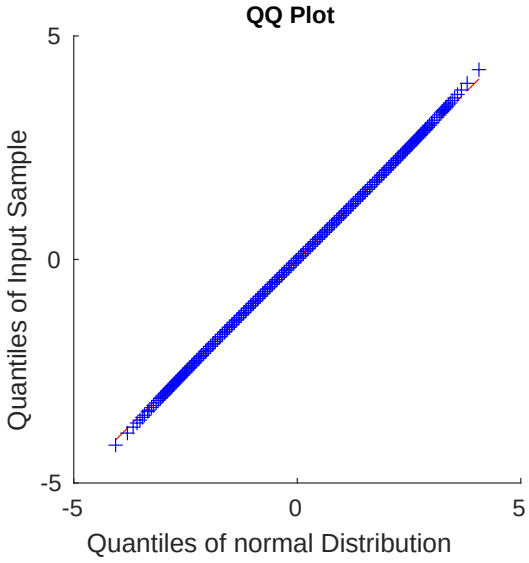
We next demonstrate the robustness of the lognormal population distribution by performing 1000 simulations, each time drawing a new randomly generated distribution of fundamentals. Figure V displays smoothed results over 1000 simulations of the model.[41] The QQ plot also demonstrates lognormality of the expected log population over these simulations. We verify the robustness of the power law coefficient estimate across the simulations. The average estimated power law coefficient from the (log) rank-size regression across the 1000 simulations is -0.95, with a standard deviation of 0.03. 90% of estimated coefficients are between -0.900 and -1.004. Performing the (log) rank-size regression on the smoothed distribution delivers a slope of -0.95. The parameter values used here are consistent with the literature and estimates are near the Zipf's law of -1 for all simulated values. While the model using standard parameters from the literature closely approximates the Zipf's law coefficient of -1, we maintain our argument that the -1 coefficient is not a meaningful feature of the data. Changes in scale and the truncation point can influence the estimate, as discussed in Section 2 and Online Appendix B. Nonetheless, it is interesting to note that the estimated power law exponent appears consistent with Zipf's law for typical parameter values in the literature.

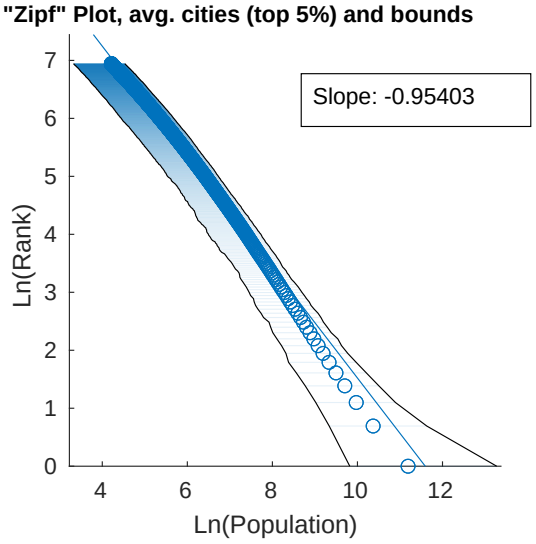
We test each of the 1000 simulated population distributions against the null hypothesis that the logged population distribution is normally distributed using the Kolmogorov-Smirnov, Lilliefors, and Jarque-Bera tests. The results of these tests are given in Table I. None of the tests reliably reject the normal distribution for the logged population. Rejec-

[41] The log of population is averaged at each rank of the distribution over the 1000 simulations. Results are similar when averaging the population and taking the log.

Figure V:
Smoothed output over 1000 simulations







1000-draw MC	Value
Coef. of Avg.	-.954
Avg. over Coefs	-.952
Std. Dev.	(.033)
Min	-1.052
5pct	-1.007
25pct	-.976
50pct	-.952
75pct	-.931
95pct	-.895
Max	-.855

Notes: The smoothed population distribution resulting from 1000 simulations of model is in the upper left panel, and the resulting QQ plot is on the upper right. Both show that the equilibrium population distribution appears lognormal. The city size distributions from the 1000 simulations are in the lower left, and statistics over model simulations the lower right. The slope on the lower left represents the slope taken over the average of log(pop) at each rank over 1000 simulations, and the bounds contain 95% of the log populations at each rank of the distribution. The table displays statistics over the 1000 estimated power law coefficients from the simulations.

Table I: Normality Tests

	Kolmogorov-Smirnov	Lilliefors	Jarque-Bera
Rejected at 1%	0.000	0.022	0.171
Rejected at 5%	0.001	0.103	0.353

Notes: Table shows the share of tests for a normal distribution rejected for the log equilibrium population of 1000 simulations.

Table II: Comparative Statics

	$\frac{\partial \theta_1}{\partial \alpha}$	$\frac{\partial \theta_1}{\partial \beta}$	$\frac{\partial \theta_1}{\partial \theta}$
Sign of change	+	+	+

Notes: Direction of the change in slope coefficient in the (log) rank-size regression for changes in α, β, and δ_{TC}, holding other parameters constant. "+" means the estimated slope (which is negative by construction) has become flatter. This implies that the largest large cities are relatively bigger. Note that $\beta < 0$, so an increase in β means a decrease in congestion. The signs are from a regression of the parameters on the estimated coefficients (discussed in Online Appendix D.3)

tions of normality occur most often under the Jarque-Bera test, which tests for skewness and kurtosis.[42] A degree of kurtosis is evident in the QQ plot as both tails appear slightly heavier than a normal distribution, which may be attributable to the finite grid.

4.2 Comparative Statics Across Simulations

We next perform comparative statics on the estimated power law coefficient for the city size distribution, testing the sensitivity to changes in model parameters. Changing these parameters alters the estimated coefficient of the log rank-size regression, which we denote θ_1 as in Equation 1. We perform 100 simulations for each of 150 combination of parameters.[43] A summary of the signs of changes (estimated by a regression given in Online Appendix D.3) is provided in Table II.

[42]The higher rejections under the Jarque-Bera test may also be attributable to the inappropriateness of this test for spatial data, similar to its inappropriateness for time series data documented in Bai and Ng (2005).

[43]We simulate for 5 values of α, 6 values of β, and 5 values of δ_{TC} given in Online Appendix D.3. The geography has the same dimensions as our baseline simulations but we change the number of locations in the full geography to 10000 to reduce computation time.

The comparative statics of our model demonstrate changes in the estimated power law coefficient in line with the empirical evidence. Increasing the benefits of agglomeration by raising $\alpha > 0$ results in a flatter slope (greater dispersion, or larger biggest cities) and increasing local congestion costs by reducing $\beta < 0$ results in a steeper slope (less dispersion, or smaller biggest cities). Increasing trade costs by increasing the rate at which these costs grow with distance, $\delta_{TC} > 0$, likewise results in a flatter slope. The flatter slopes for developing countries (documented in Duben and Krause (2021)) may be attributable to high domestic transportation costs, which are often substantially higher in developing than in developed countries. Atkin and Donaldson (2015) estimate that domestic trade costs are roughly four to five times higher in Nigeria and Ethiopia than in the U.S., in line with the empirical evidence documented in Teravaninthorn and Gaël (2009). These high domestic trade costs could contribute to the phenomenon of large metropolises ("primate cities") within the developing world. Additionally, the flattening slope in the U.S. in recent decades (documented in Gabaix and Ioannides (2004)) could be a result of increased agglomeration benefits in the modern services economy.

4.3 Gibrat's Law

Allen and Arkolakis (2014) demonstrate that the population vector is scaled by changes in the aggregate population, but the relative populations over locations are not changed by changes in the total population. We now demonstrate that based on this property the equilibrium population distribution demonstrates proportional growth and satisfies Gibrat's law in response to increases in the aggregate population $\bar{L}$.

We can also write Equation 31, the matrix form representation of Equation 19 as:

$$\tilde{\mathbf{h}} = \mathbf{J}[\tilde{\mathbf{h}}]^{\frac{\gamma_2}{\gamma_1}}$$

where $\tilde{h}_{i'} = h_{i'}\bar{W}^{\frac{\sigma-1}{1-\frac{\gamma_2}{\gamma_1}}}$, the notation $[\cdot]^a$ indicates raising each element of the vector to the power a, and the matrix $\mathbf{J}$ is given in Online Appendix A.4. As this relationship must hold

for any level of $\bar{L}$, changing $\bar{L}$ does not impact the resulting population distribution even as it impacts welfare ($\bar{W}$, which is the same across all locations). That is, a percentage increase in overall population will result in each location experiencing population growth of the same percentage and so population growth rates will be unrelated to initial population and Gibrat's law will hold within the equilibrium of this model.

This is a key difference between our explanation for observed population distributions based on locational fundamentals and trade and the prior literature on random growth models. Rather than being the force creating the equilibrium distribution, random growth is a feature of an equilibrium based on underlying characteristics of place and trade. This view that is supported by the absence of Gibrat's law in systems that are in transition or have suffered dis-equilibrating shocks (Desmet and Rappaport, 2017; Davis and Weinstein, 2002; Davis and Weinstein, 2008).

5 Conclusion

The power law-like distribution of city populations is a striking empirical regularity that holds across countries and millennia. In this paper, we demonstrate that a broad class of economic geography models generate these characteristic population distributions when modeled with a realistic geography. We integrate insights from economic geography theory regarding the importance of both place and space into the extensive literature on power law-like population distributions and Zipf's law. Viewing population distributions as arising naturally in response to favorable geography and trade access provides a simple explanation for the emergence of the distinctive city size distribution. This explanation is consistent with the persistence of human settlements, the recovery of cities from disasters, and the random growth of cities in equilibrium.

References

Alix-Garcia, Jennifer and Emily A. Sellars (2020). "Locational fundamentals, trade, and the changing urban landscape of Mexico". *Journal of Urban Economics* 116, p. 103213.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". *Quarterly Journal of Economics* 129.3, pp. 1085–1140.

Anderson, G.W., Alice Guionnet, and O. Zeitouni (2010). *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Atkin, David and Dave Donaldson (2015). "Who is Getting Globalized? The Size and Implications of Intranational Trade Costs". Working paper.

Auerbach, Felix (1913). "Das Gesetz der Bevölkerungskonzentration". *Petermann's Geographische Mitteilungen*.

Auerbach, Felix and Antonio Ciccone (2023). "The Law of Population Concentration". *Envmt. and Planning B: Urban Analytics and City Sc.* 50.2, pp. 290–298.

Bai, Jushan and Serena Ng (2005). "Tests for Skewness, Kurtosis, and Normality for Time Series Data". *Journal of Business and Economic Statistics* 23.1, pp. 49–60.

Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortaçsu (2019). "Trade, Merchants, and the Lost Cities of the Bronze Age". *The Quarterly Journal of Economics*, pp. 1455–1503.

Behrens, Kristian and Frederic Robert-Nicoud (2015). "Agglomeration Theory with Heterogenous Agents". *Handbook in Regional and Urban Economics* Ch. 5.

Ben Arous, Gerard and Alice Guionnet (2010). "Wigner matrices". English (US). *Handbook on random matrices*. Ed. by G Ackemann, Jinho Baik, and P Di Francesco. Oxford University Press, pp. 433–451.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.

Blank, Aharon and Sorin Solomon (2000). "Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components)". *Physica A: Statistical Mechanics and its Applications* 287.1, pp. 279–288.

Bosker, Maarten and Eltjo Buringh (2017). "City seeds: Geography and the origins of the European city system". *Journal of Urban Economics* 98, pp. 139–157.

Bradley, Richard C. (1993). "Equivalent Mixing Conditions for Random Fields". *The Annals of Probability* 21.4, pp. 1921–1926.

— (Jan. 2005). "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions". *Probability Surveys* 2.N.A.

Brakman, Steven, Harry Garretsen, and Marc Schramm (Apr. 2004). "The strategic bombing of German cities during World War II and its impact on city growth". *Journal of Economic Geography* 4.2, pp. 201–218.

Brakman, Steven, Harry Garretsen, Charles Van Marrewijk, and Marianne Van Den Berg (1999). "The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution". *Journal of Regional Science* 39.1, pp. 183–213.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008). "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics* 63.2, pp. 723–742.

— (2019). "The Costs of Agglomeration: House and Land Prices in French Cities". *Review of Economic Studies*, pp. 1556–1589.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "Chapter 5 - The Empirics of Agglomeration Economies". *Handbook of Regional and Urban Economics*. Vol. 5. Elsevier, pp. 247–348.

Córdoba, Juan-Carlos (2008). "On the distribution of city sizes". *Journal of Urban Economics* 63.1, pp. 177–197.

Davis, Donald R. and David E. Weinstein (Dec. 2002). "Bones, Bombs, and Break Points: The Geography of Economic Activity". *American Economic Review* 92.5, pp. 1269–1289.

Davis, Donald R. and David E. Weinstein (2008). "A Search for Multiple Equilibria in Urban Industrial Structure". *Journal of Regional Science* 48.1, pp. 29–65.

Davis, Morris A. and François Ortalo-Magné (2011). "Household expenditures, wages, rents". *Review of Economic Dynamics* 14.2, pp. 248–261.

Desmet, Klaus and Jordan Rappaport (2017). "The settlement of the United States, 1800–2000: The long transition towards Gibrat's law". *Journal of Urban Economics* 98, pp. 50–68.

Dingel, Jonathan I., Antonio Miscio, and Donald R. Davis (2021). "Cities, lights, and skills in developing economies". *Journal of Urban Economics* 125, p. 103174.

Doukhan, P. (1994). *Mixing: Properties and Examples*. Lecture notes in statistics. 3Island Press.

Duben, Christian and Melanie Krause (2021). "Population, light, and the size distribution of cities". *Journal of Regional Science*, pp. 189–211.

Eeckhout, Jan (2004). "Gibrat's Law for (All) Cities". *American Economic Review*, pp. 1429–1451.

— (Sept. 2009). "Gibrat's Law for (All) Cities: Reply". *American Economic Review* 99.4, pp. 1676–83.

Frigg, Roman, Joseph Berkovitz, and Fred Kronz (2020). "The Ergodic Hierarchy". *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables (June 1999). *The Spatial Economy: Cities, Regions, and International Trade*. The MIT Press.

Gabaix, Xavier (1999a). "Zipf's Law and the Growth of Cities". *The American Economic Review* 89.2, pp. 129–132.

— (1999b). "Zipf's Law for Cities: An Explanation". *The Quarterly Journal of Economics*.

— (2009). "Power Laws in Economics and Finance". *Annual Review of Economics* 1.1, pp. 255–294.

Gabaix, Xavier and Yannis M. Ioannides (2004). "The Evolution of City Size Distributions". *In: Handbook of Regional and Urban Economics*, pp. 2341–2378.

Gibrat, Robert (1931). "Les inégalités économiques".

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

Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil (Sept. 2018). "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade". *The Quarterly Journal of Economics* 133.1, pp. 357–406.

Herrndorf, Norbert (1984). "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables". *The Annals of Probability* 12.1, pp. 141–153.

Holmes Thomas J. andLee, Sanghoon (2010). "Cities as Six-by-Six-Mile Squares: Zipf's Law?" *Agglomeration Economics*. NBER.

Jiang, Bin, Junjun Yin, and Qingling Liu (2014). "Zipf's Law for all the natural cities in the world". *Working Paper*.

Johnson, Noel, Remi Jedwab, and Mark Koyama (2019). "Pandemics, places, and populations: Evidence from the Black Death". *Working Paper*.

Krugman, Paul (1991). "Increasing Returns and Economic Geography". *Journal of Political Economy* 99.3, pp. 483–499.

— (1996). *The Self-Organizing Economy*. Blackwell.

Lee, Sanghoon and Qiang Li (2013). "Uneven landscapes and city size distributions". *Journal of Urban Economics* 78, pp. 19–29.

Malevergne, Y., V. Pisarenko, and D. Sornette (2011). "Gibrat's Law for Cities: Uniformly Most Powerful Unbiased Test of the Pareto Against the Lognormal". *Physical Review*.

Marlow, N.A. (1967). "A Normal Limit Theorem for Power Sums of Independent Random Variables". *Bell System Technical Journal* 46.9, pp. 2081–2089.

Mazmanyan, Lilit, Victor Ohanyan, and Dan Treitsch (2008). "The Lognormal Central Limit Theorem for Positive Random Variables". *Working Paper*.

Mori, Tomoya, Tony E. Smith, and Wen-Tai Hsu (2020). "Common power laws for cities and spatial fractal structures". *Proceedings of the National Academy of Sciences* 117.12, pp. 6469–6475.

Nordhaus, William D. (2006). "Geography and macroeconomics: New data and new findings". *Proceedings of the Natl Academy of Sciences* 103.10, pp. 3510–3517.

Nunn, Nathan and Diego Puga (2012). "Ruggedness: The Blessing of Bad Geography in Africa". *Review of Economics and Statistics* 94.1, pp. 20–36.

O'Rourke, Sean, Van Vu, and Ke Wang (2016). *Eigenvectors of random matrices: A survey*.

Rante, Rocco, Federico Trionfetti, and Priyam Verma (Mar. 2024). "The Size Distribution of Cities: Evidence from the Lab". working paper or preprint.

Rappaport, Jordan and Jeffrey D. Sachs (2003). "The United States as a Coastal Nation". *Journal of Economic Growth* 8.1, pp. 5–46.

Redding, Stephen J. (2016). "Goods trade, factor mobility and welfare". *Journal of International Economics*.

Redding, Stephen J. and Esteban Rossi-Hansberg (2017). "Quantitative Spatial Economics". *Annual Review of Economics*.

Rosenthal, Stuart S. and William C. Strange (2004). "Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies". *Cities and Geography*. Ed. by J. Vernon Henderson and Jacques-François Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, pp. 2119–2171.

Rossi-Hansberg, Esteban and Mark Wright (2007). "Urban Structure and Growth". *The Review of Economic Studies*, pp. 597–624.

Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse (Aug. 2011). "The Area and Population of Cities: New Insights from a Different Perspective on Cities". *American Economic Review* 101.5, pp. 2205–25.

Simonovska, Ina and Michael E. Waugh (2014). "The elasticity of trade: Estimates and evidence". *Journal of International Economics* 92.1, pp. 34–50.

Soo, Kwok Tong (2005). "Zipf's Law for Cities: A Cross-Country Investigation". *Regional Science and Urban Economics*, pp. 239–263.

Teravaninthorn, Supee and Raballand Gaël (2009). "Transport Prices and Costs in Africa: A Review of the Main International Corridors". *World Bank Report*.

Tobler, Waldo (1970). "A computer movie simulating urban growth in the Detroit region". *Economic Geography* 46.

Zipf, George Kingsley (1949). "Human Behavior and the Principle of Least Effort". *Addison-Wesley Press*.

ONLINE APPENDIX

A Discussion of Main Text Results and Proofs

A.1 Algebra for "Pareto form" of Lognormal PDF

The density function of a lognormal distribution is given by:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

Expanding the square and grouping the $ln(x)$ terms yields:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\ln\left(x^{\left(\frac{-\ln(x)+2\mu}{2\sigma^2}\right)}\right) - \frac{\mu^2}{2\sigma^2}\right)$$

Applying $e^{\ln(a^b)} = a^b$ and combining with x^{-1}:

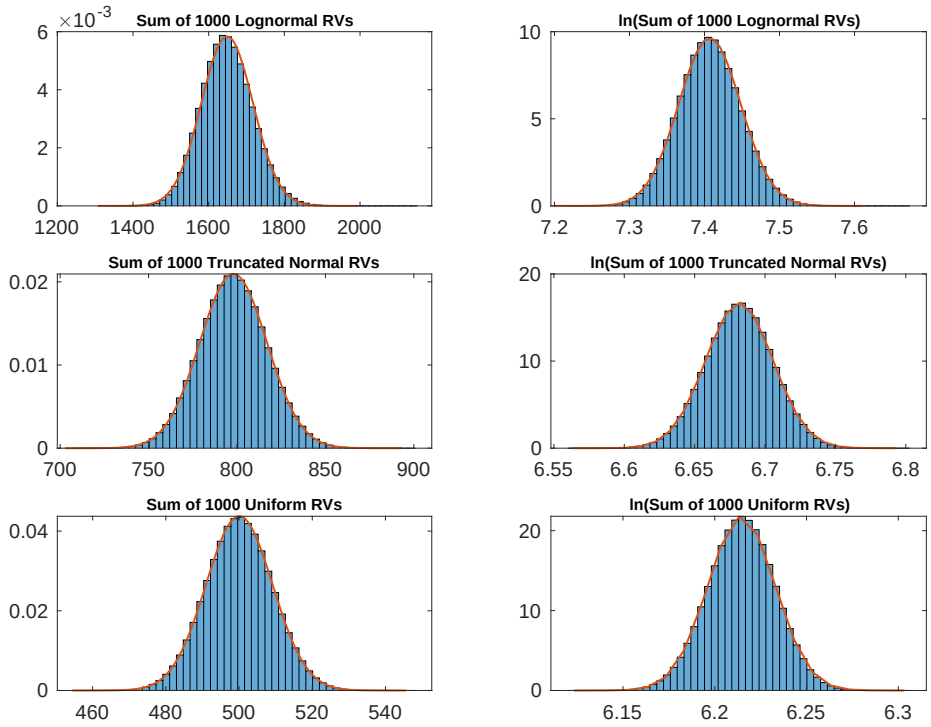
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) x^{-\left(\frac{\ln(x)-2\mu}{2\sigma^2}\right)-1}$$

Writing the constant term $\frac{1}{\sigma\sqrt{2\pi}}$ as Γ, the lognormal distribution can be written as:

$$f(x) = \Gamma x^{-\alpha(x)-1} \quad \text{, where } \alpha(x) = \frac{\ln(x)-2\mu}{2\sigma^2}$$

which is the same as Equation 3 in the main text. The representation of the lognormal PDF here appears in Malevergne, Pisarenko, and Sornette (2011), and is similar to that in Eeckhout (2009).

Figure A.I:
Sums of Positive Random Variables Drawn from Various Distributions



Note: Histograms show 1,000,000 replications. The random variables in the first row are drawn from a lognormal distribution with parameters $\mu_{LN} = 0, \sigma_{LN} = 1$, the middle row from a truncated normal distribution with parameters $\mu_{TN}, \sigma_{TN} = 0$ and minimum value $\alpha = 0.001$, and the bottom row from a uniform distribution on $(0, 1]$. The red overlaid line represents a normal distribution with the same mean and standard deviation as the underlying sums in each panel. The sums appear distributed normally in both levels (column 1) and in logs (column 2), as implied by Lemma 1.

A.2 Discussion of Lemma 2 (Marlow, 1967)

A demonstration of the apparent normality in both levels and logs of sum of positive random variables is included in the main text in Figure A.I, which shows the sums of lognormal ($\exp(N(0,1))$, truncated normal (a standard normal truncated at 0), and (0,1] uniform random variables. The sums of these random variables converge to normal distributions, while the log of the sum also appears to follow a normal distribution.

Examples of sums over several positive random variables (lognormal, truncated normal, and uniform) are presented in Figure A.I, exhibiting the appearance of normality in both levels and logs for these sums.

The Marlow result is interesting, as it appears to imply that a sum of positive random variables can be viewed as approaching a lognormal or normal distribution for large N. However, if we are considering both the normal and lognormal approximations for a sum of positive random variables we can demonstrate convergence of the lognormal approximation to the equivalent normal approximation—that is, the lognormal and normal approximation will be identical in the limit. The following discussion draws on Mazmanyan, Ohanyan, and Treitsch (2008).

For simplicity, consider an approximation of $i.i.d$ positive random variables with mean m and variance s^2. The normal approximation will have parameters $\mu_N = nm = M$ and $\sigma_N^2 = ns^2$. We will now define the parameters for the lognormal approximation of the sum.

First, define the coefficient of variation as

$$C_v = \frac{\sqrt{ns^2}}{nm} = \frac{\sqrt{n}s}{nm} \tag{24}$$

As n grows large, $C_v \to 0$.

The parameters μ_X and σ_X of the lognormal approximation can be found by

$$\sigma_X^2 = \ln(1 + C_v^2) \tag{25}$$

$$\mu_X = \ln(nm) - \frac{\sigma_X^2}{2} \tag{26}$$

As $C_v \to 0$ when $n \to \infty$, Equation (25) gives that as $n \to \infty$:

$$\sigma_X^2 \to C_v^2 \quad \text{, so} \quad \sigma_X \to C_v \quad \text{, and } \sigma_X \to 0 \tag{27}$$

Now we demonstrate that the lognormal approximation converges to the expected normal. As, for some m, $P(||x - M|| > \epsilon | n \geq m) = 0$, so $\frac{x}{M} \to_{a.s.} 1$. We can write $\frac{x}{M} \cdot \sigma_X \to \frac{\sqrt{n}s}{nm}$, and so $x\sigma_X \to \sqrt{n}s = \sigma_N$. This means $\frac{1}{x\sigma_X\sqrt{2\pi}} \to \frac{1}{\sigma_N\sqrt{2\pi}}$.

Similarly, $x = \frac{xM}{M}$, so $\ln(x) = \ln(M) + \ln(\frac{x}{M})$. As $\frac{x}{M} \to 1$, so $\ln(\frac{x}{M}) \to \frac{x-M}{M}$. As

$\mu_X = \ln(M) - \frac{\sigma_X^2}{2}$, and $\sigma_X \to C_v \to 0$, and $M = \mu_N$ then we have

$$\frac{\ln(x) - \mu_x}{\sigma_X} \to \frac{\ln(M) + (\frac{x-M}{M}) - \ln(M)}{\sigma_X} = \frac{x - M}{M\sigma_X} \to \frac{x - M}{M \cdot C_v} = \frac{x - \mu_X}{\sigma_N} \tag{28}$$

So we have shown, as $n \to \infty$,

$$f(x) = \frac{1}{x\sigma_X\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu_X}{\sigma_X}\right)^2} \to \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_N}{\sigma_N}\right)^2} \tag{29}$$

So as n increases, the lognormal approximation to the sum approaches the normal approximation.

A.3 Definitions of α-mixing, sequences and fields

Definition 1, α-mixing (sequences): *Suppose $X := (X_k, k \in \mathbb{Z})$ is a (not necessarily stationary) sequence of random variables. For $-\infty \leq J \leq L \leq \infty$, define the σ-field*

$$\mathcal{F}_J^L := \sigma\left(X_k, J \leq k \leq L(|k \in \mathbb{Z}\right).$$

The notation $\sigma(\ldots)$ means the σ-field $\subset \mathcal{F}$ generated by $(\ldots)$. Define:

$$\alpha\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right) = \sup_{A \in \mathcal{F}_{-\infty}^{J}, B \in \mathcal{F}_{j+n}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

For each $n \geq 1$, define:

$$\alpha(n) := \sup_{j \in \mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right),$$

The random sequence X is said to be "strongly mixing" (or " α-mixing") if $\alpha(n) \to 0$ as $n \to \infty$.

The definition above is drawn from Section 2.1 of Bradley (2005), which includes further discussion and the definitions of additional mixing conditions.[44] The ordering of the random variables X_k, in this definition, reflects the time-series nature of this formulation and reflects the "time" of the realization. We instead order the attribute type indices $g \in \mathcal{G}$ in terms of the similarity of the attributes, such that the difference in index reflects the "distance" between attributes $g, g' \in \mathcal{G}$.

The following definition of α-mixing for random fields is used in the more general version of our proofs for a two-dimensional geography in the Online Appendix, rather than the simpler "linear" geography presented in the main text.

Definition 2, α-mixing (fields): *Suppose $\{X_i\}, i \in \mathbb{Z}^2$ is a stationary random field where each X_i is drawn from a common distribution. For disjoint sets S, T, define the σ-fields:*

$$\mathcal{F}_S := \sigma\left(X_s, s \in S)\right), \mathcal{F}_T := \sigma\left(X_t, t \in T)\right).$$

[44] For additional discussion, see (Billingsley, 1995).

The notation $\sigma(\ldots)$ *means the* σ*-field* $\subset \mathcal{F}$ *generated by* $(\ldots)$*. Define:*

$$\alpha(S,T) = \sup_{A\in\mathcal{F}_S, B\in\mathcal{F}_T} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Define $dist(S,T) = \inf_{s\in S, t\in T} \|s-t\|$*, where* $\|\cdot\|$ *denotes the Euclidean norm. For each* $k \geq 1$ *and* $u, v \in \mathbb{R}_{++}$*, define:*

$$\alpha(k;u,v) := \sup_{S,T} \alpha(S,T),$$

where the supremum is taken over all disjoint subsets S, T *with* $|S| \leq u, |T| \leq v$ *such that* $dist(S,T) \geq k$*. The random field* $\{X_i\}$ *is said to be "strongly mixing" (or "*α*-mixing") if* $\alpha(k;\infty,\infty) \to 0$ *as* $k \to \infty$*.*

The definition above is drawn from Doukhan (1994) and Bradley (1993), both of which also include further discussion and additional mixing concepts for fields. The key distinction from the definition for sequences introduced earlier is the need to incorporate a concept of distance, which previously was summarized by the indices when considering α-mixing of a sequence.

A.4 Population as a random variable

The system of equations describing the population distribution, given for a particular location i in Equation 19 can be expressed in matrix form as:

$$\theta \mathbf{h} = \mathbf{J}[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}} \tag{30}$$

where $\theta = \bar{W}^{\sigma-1}$, each element of the vector $\mathbf{h}$ is given by $h_i = L_i^{\tilde{\sigma}\gamma_1}$, and $[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}}$ indicates raising each element of the vector $\mathbf{h}$ to the power $\frac{\gamma_2}{\gamma_1}$. The matrix $\mathbf{J}$, where elements $j_{i,n} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(\sigma-1)}$, is given by

$$J = \begin{bmatrix} j_{1,1} & j_{1,2} & \cdots & j_{1,N} \\ j_{2,1} & j_{2,2} & \cdots & j_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ j_{N,1} & j_{N,2} & \cdots & j_{N,N} \end{bmatrix}$$

where $\mathbf{J}$ is an $\bar{N} \times \bar{N}$ random matrix, with its elements consisting of the realizations for fundamentals and trade costs for all locations. The vector $\mathbf{h}$, which consists of transformations of the population vector, is the eigenvector corresponding to the leading eigenvalue of this random matrix as shown in the online appendix to Allen and Arkolakis (2014). The eigenvectors of random matrices are themselves random vectors, which motivates our treatment of each L_i as a random variable.[45]

Note that each sequence $s_{i,n}$ for each i can be written as the vector $\mathbf{s}_i$ resulting from the following matrix multiplication:

$$\mathbf{s}_i = \mathbf{K}[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}} \tag{31}$$

where the matrix $\mathbf{K}$ with elements $k_{i,n} = \tau_{i,n} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}1-\sigma}$ is given by:

[45] For a review of random matrices, see Anderson, Guionnet, and Zeitouni (2010) and for results on eigenvectors see Ben Arous and Guionnet (2010) and O'Rourke, Vu, and Wang (2016).

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,N} \\ k_{2,1} & k_{2,2} & \dots & k_{2,N} \\ \dots & \dots & \dots & \dots \\ k_{N,1} & k_{N,2} & \dots & k_{N,N} \end{bmatrix}$$

The random vector $\mathbf{h}$ is not the eigenvector of the random matrix $\mathbf{K}$. The vector $\mathbf{s}_i$ results from the multiplication of a random vector and a random matrix, motivating the treatment of elements $s_{i,n}$ as random variables.

A.5 Proofs for random fields

Several of our results generalize directly to geographies in two dimensions, as they necessitate an α-mixing sequence (as in the one-dimensional case) or otherwise do not depend on dimensions of the geography. This is true for Proposition 1 and Proposition 2 (which necessitate an α-mixing sequence) and Proposition 3 and Theorem 1 (which rely only on the properties of lognormal distributions).

The complication in considering geographies in two dimensions results when we must treat α-mixing random fields rather than α-mixing sequences. The additional complexity results from the need to be careful about the concept of "distance" within the sets and sequences when the index no longer neatly summarizes this information. The definition in Online Appendix A.3 for α-mixing fields, for instance, defines the partition of the field into subsets by the Euclidean distance between the nearest elements of each subset.

To extend the proofs of Proposition 4 and Theorem 2 we show that the key properties of α-mixing sequences from Bradley (2005) Theorem 5.2 (which provides the proof for the main text case of sequences) can be extended to the case of α-mixing fields.

Lemma A.1: *Suppose that for each* $n = 1, 2, 3, ..., X^{(n)} := \left(X_k^{(n)}, k \in \mathbb{Z}^2\right)$ *is a (not necessarily stationary) field of random variables. Suppose these fields* $X^{(n)}$, $n = 1, 2, 3, ...$, *are independent of each other. Suppose that for each* $k \in \mathbb{Z}^2, h_k : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \ldots \to \mathbb{R}$ *is a Borel function. Define the field* $X := (X_k, k \in \mathbb{Z}^2)$ *of random variables by* $X_k := h_k\left(X_k^{(1)}, X_k^{(2)}, X_k^{(3)}, \ldots\right), k \in \mathbb{Z}^2$. *Then for each* $m \geq 1$, $\alpha(X, m) \leq \sum_{n=1}^{\infty} \alpha\left(X^{(n)}, m\right)$

Proof: First, we want to show that α-mixing is preserved over measurable transformations of α-mixing random fields, and then we want to show that combinations of α-mixing random fields are also mixing. For brevity, we write $\alpha(k)$ in place of $\alpha(k; \infty, \infty)$ as appears in the definition of α-mixing fields in Online Appendix A.3.

1. Define the field $\{\bar{D}_i\}$ such that for all $i \in \mathbb{Z}^2$, $\bar{D}_i = j\left(D_i^{(1)}, D_i^{(2)}, D_i^{(3)}, ...\right)$, where $j(\cdot)$ is a measurable mapping that takes the field $\{D_i\}$, which is an n-tuple of random variables

for each $i \in \bar{Z}$, as input. The field $\{D_i\}$ is α-mixing with respect to distance k. We want to show that the field $\{\bar{D}_i\}$ will be mixing as well. First note that

$$\mathbb{P}\left(\bar{D}_{s\in S} \in A, \bar{D}_{t\in T} \in B\right) = \mathbb{P}\left((D^{(1)}_{s\in S}, D^{(2)}_{s\in S}, D^{(3)}_{s\in S}, ...) \in j^{-1}(A), (D^{(1)}_{t\in T}, D^{(2)}_{t\in T}, D^{(3)}_{t\in T}, ...) \in j^{-1}(B)\right)$$

And so,

$$\begin{aligned}
\alpha_{\bar{D}}(k) &= \sup_{A,B} \left|\mathbb{P}\left(\bar{D}_{s\in S*} \in A, \bar{D}_{t\in T*} \in B\right) - \mathbb{P}\left(\bar{D}_{s\in S*} \in A\right)\mathbb{P}\left(\bar{D}_{t\in T*} \in B\right)\right| \\
&= \sup_{A,B} |\mathbb{P}\left((D^{(1)}_{s\in S*}, ...) \in j^{-1}(A), (D^{(1)}_{t\in T*}, ...) \in j^{-1}(B)\right) \\
&\quad - \mathbb{P}\left((D^{(1)}_{s\in S*}, ...) \in j^{-1}(A)\right)\mathbb{P}\left((D^{(1)}_{t\in T*}, ...) \in j^{-1}(B)\right)| \\
&\leq \alpha_D(k),
\end{aligned}$$

as $\alpha_D(k)$ is defined as the supremum over sets S, T, not the fixed $S*, T*$ that correspond to $\{\bar{D}_i\}$. As $\alpha_D(k) \to 0$ as $k \to \infty$, then $\alpha_{\bar{D}}(k) \to 0$ as well and so $\{\bar{D}_i\}$ is α-mixing.

2. We now show that given two independent stationary random fields $\{Y_i\}, \{Z_i\}$ where $i \in \mathbb{Z}^2$ which are α-mixing with respect to distance k, the bivariate field $\{X_i\}$ where $X_i = (Y_i, Z_i)$ is also mixing.

   Define I:

$$\begin{aligned}
I &= |\mathbb{P}(X_{s\in S} \in A, X_{t\in T} \in B) - \mathbb{P}(X_{s\in S} \in A)\mathbb{P}(X_{t\in T} \in B)| \\
&= |\mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A, (Y_{t\in T}, Z_{t\in T}) \in B) \\
&\quad - \mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A)\mathbb{P}((Y_{t\in T}, Z_{t\in T}) \in B)|
\end{aligned}$$

   Define

$$\begin{aligned}
f(Z_{s\in S}, Z_{s\in T}) &= \mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A, (Y_{t\in T}, Z_{t\in T}) \in B) \\
g(Z_{S\in S}) &= \mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A) \\
h(Z_{t\in T}) &= \mathbb{P}((Y_{t\in T}, Z_{t\in T}) \in B)
\end{aligned}$$

Substituting in and taking expectations,

$$I = |\mathbb{E}[f(Z_{s\in S}, Z_{s\in T})] - \mathbb{E}[g(Z_{s\in S})h(Z_{t\in T})]|$$

Add and subtract $\mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})]$:

$$\begin{aligned} &= |\mathbb{E}[f(Z_{s\in S}, Z_{s\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})] \\ &\quad + \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})] - \mathbb{E}[g(Z_{s\in S})h(Z_{t\in T})]| \end{aligned}$$

Re-arrange and, using the $|\cdot|$,

$$\begin{aligned} &= |\mathbb{E}[f(Z_{s\in S}, Z_{s\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})] \\ &\quad - (\mathbb{E}[g(Z_{s\in S})h(Z_{t\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})])| \\ &\leq \underbrace{|\mathbb{E}[f(Z_{s\in S}, Z_{s\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})]|}_{II} \\ &\quad + \underbrace{|\mathbb{E}[g(Z_{s\in S})h(Z_{t\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})]|}_{III} \end{aligned}$$

Begin with II, where we use independence of Y, Z and mixing of Y to show

$$\begin{aligned} II &= |\mathbb{E}[f(Z_{s\in S}, Z_{s\in T})] - \mathbb{E}[g(Z_{s\in S})\mathbb{E}h(Z_{t\in T})]| \\ &= |\mathbb{E}[\mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A, (Y_{t\in T}, Z_{t\in T}) \in B)] \\ &\quad - \mathbb{E}[\mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A)]\mathbb{E}[\mathbb{P}((Y_{t\in T}, Z_{t\in T}) \in B)]| \\ &= |\mathbb{E}[\mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A, (Y_{t\in T}, Z_{t\in T}) \in B | Z_{s\in S} = z_{s\in S}, Z_{t\in T} = z_{t\in T})] \\ &\quad - \mathbb{E}[\mathbb{P}((Y_{s\in S}, Z_{s\in S}) \in A | Z_{s\in S} = z_{s\in S})]\mathbb{E}[\mathbb{P}((Y_{t\in T}, Z_{t\in T}) \in B | Z_{t\in T} = z_{t\in T})]| \\ &= |\mathbb{P}(Y_{s\in S} \in A, Y_{t\in T} \in B) - \mathbb{P}(Y_{s\in S} \in A)\mathbb{P}(Y_{t\in T} \in B)| \end{aligned}$$

Taking the supremum over A, B, we have

$$II \leq \alpha_Y(S, T)$$

Now consider III:

$$\begin{aligned} III &= |\mathbb{E}[g(Z_{s\in S})h(Z_{t\in T})] - \mathbb{E}[g(Z_{s\in S})]\mathbb{E}[h(Z_{t\in T})]| \\ &= |Cov(g(Z_{s\in S}), h(Z_{t\in T}))| \end{aligned}$$

Note that $\|g(Z_{s\in S})\|_\infty, \|h(Z_{t\in T})\|_\infty \leq 1$. By Lemma 3.1 of Doukhan (1994) and the α-mixing of $\{Z_t\}$ we have

$$III \leq 4\alpha_Z(S,T)$$

And so, putting together the above and, taking the supremum over A, B:

$$\alpha_X(S,T) \leq \alpha_Y(S,T) + 4\alpha_Z(S,T)$$

Taking the supremum again over all S, T such that $dist(S,T) \geq k$, we find

$$\alpha_X(k) \leq \alpha_Y(k) + 4\alpha_Z(k)$$

And, as we know $\alpha_Y(k), \alpha_Z(k) \to 0$ as $k \to \infty$, we have $\alpha_X(k) \to 0$ as $k \to \infty$ and so the field $\{X_i\}$ is mixing with respect to distance k.

Combining the results of Parts 1 and 2 establish the result. ■

With Lemma A.1, we can restate Proposition 4 and Theorem 2 for the case of random fields in two dimensions where $\{A_i\}$, $\{U_i\}$, Ω_i, and S_i are such that $i \in \mathbb{Z}^2$.

Proposition 4 (fields): *If $\{A_i\}$ and $\{U_i\}$, are independent α-mixing fields, then Ω_i is an α-mixing fields.*

Theorem 2 (fields): *If $\{\Omega_i\}\}$ and $\{S_i\}$ are independent α-mixing fields, then the population field $\{L_i\}$ is α-mixing.*

The proofs of both extensions follow directly from Lemma A.1.

B Pareto vs. Lognormal Simulations

We provide additional evidence for the lognormality of the true population by focusing on the behavior of the distribution in the tail. As discussed in Eeckhout (2004), one characteristic of the lognormal as compared to the Pareto is the sensitivity of the estimated coefficient to the truncation point. This can be see in Figure A.II, where selecting alternative truncation points changes the estimated power law coefficient. The left column consists of three plots where we estimate the power law coefficient on three different subsets of the U.S. city distribution. Panel A presents the same plot as main text Figure I. When including more cities (Panel C) the coefficient rises, and when including few cities (Panel E) the coefficient falls. This is in line with the expected behavior of the scale-varying "shape parameter"-like term of the lognormal distribution in Equation 3. The R^2 of all of these regressions remains similarly high. The changes in the estimated coefficient are in line with expectations if the true underlying distribution were lognormal and demonstrate that, while the -1 exponent can be found for a particular truncation point (as in Figure I), it does not appear to be a meaningful feature of the distribution.

In the right column of Figure A.II, we consider deviations in the far tail by excluding the top quarter of cities within each subset of the city distribution. In all cases, nearly all top quarter cities fall below the trendline predicted based on the rest of the distribution.[46] The magnitude of the systematic divergence is very large, which is obscured on the log scale. As noted in the main text, the total deviation below the trendline in Panel A is 76 million people missing from the top 250 U.S. MSAs, roughly a quarter of the U.S. population. The deviations are larger when estimated on the subset of cities below the top quarter as in Panels B, D, and F. If the Pareto were the true distribution, panel B indicates a cumulative absence (in expectation) of 412 million people while panel D indicates a cumulative 494 million people missing from the sets of cities considered, both substantially more than the

[46] Of top-quarter MSAs, 62 of 62 MSAs in Panel B, 93 of 96 MSAs in panel D, and 27 of 27 MSAs in panel F are below the respective trendlines in Figure A.II

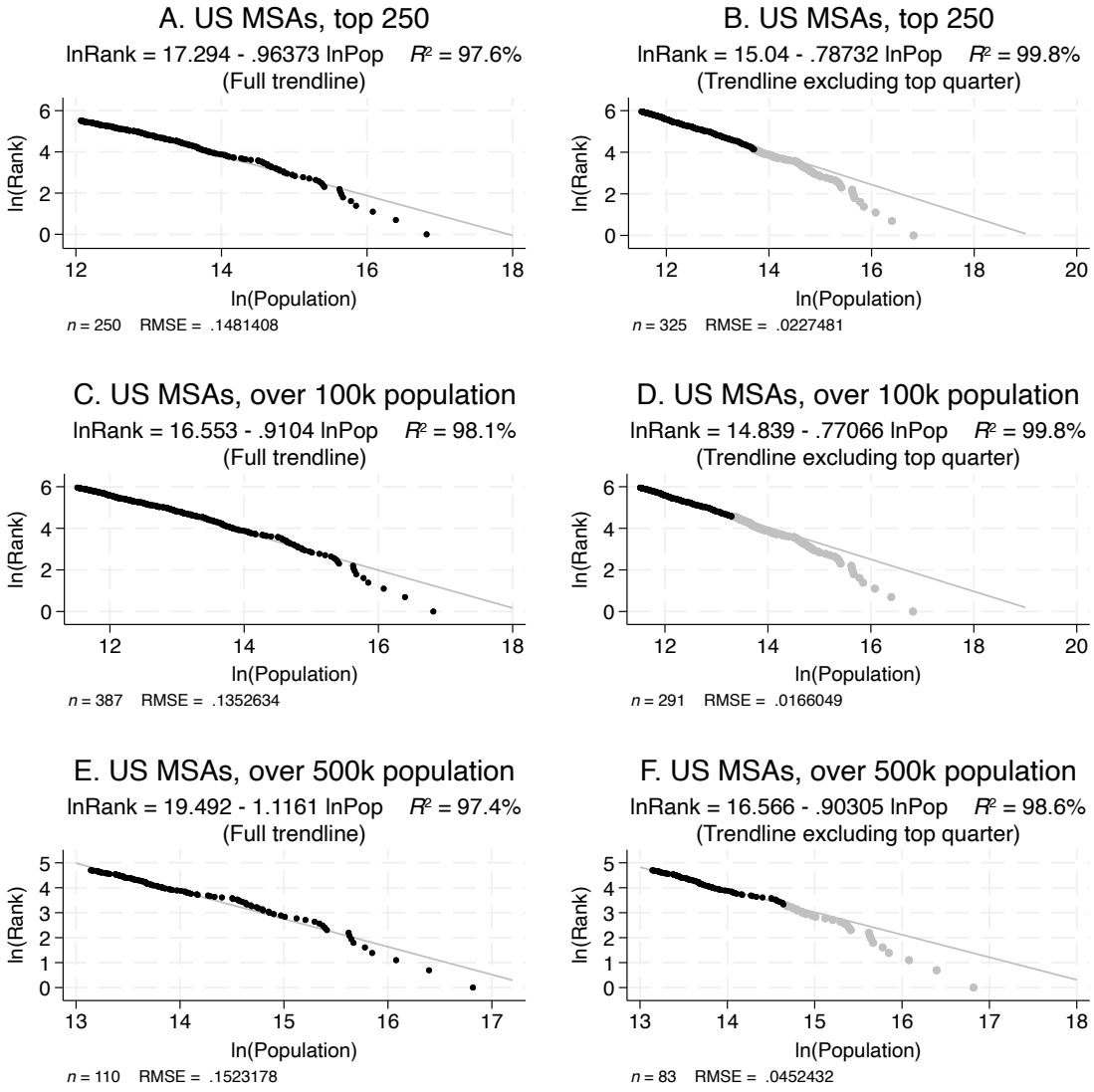
entire U.S. population, while panel F indicates an absence of 169 million people, roughly half the U.S. population.[47]

This scale variance offers evidence for the lognormal interpretation of the population distribution. When the true population is lognormal, large economies or regions (those containing many cities) should systematically contain smaller large cities than predicted by the estimated power law. We first demonstrate this property of the two distributions via simulation in Figure A.III. We calibrate a lognormal distribution to match a Pareto distribution with shape parameter $\alpha_P = 1$ in the tail.[48] The plots show the average value over 10000 draws from each distribution at each rank of the city size distribution, with the bands reflecting the 95 percent confidence interval. We calculate the slope at several scales excluding the top 25% of cities, to demonstrate the tail divergence of the lognormal resulting from its scale-variance, in contrast to the scale-invariant Pareto. When the tail is constructed to contain 100 cities, the difference between the two plots is minimal. However, when the tail is constructed to have 800 cities, cities in the far tail of the lognormal fall well below the estimated trendline.

[47]Repeating this exercise with other large countries (India, China, and Brazil) using standardized city definitions from Dingel, Miscio, and Davis (2021) indicates similarly large divergences in the tail, all in the expected direction (cumulative absences of 135 million, 53 million, and 8 million respectively).

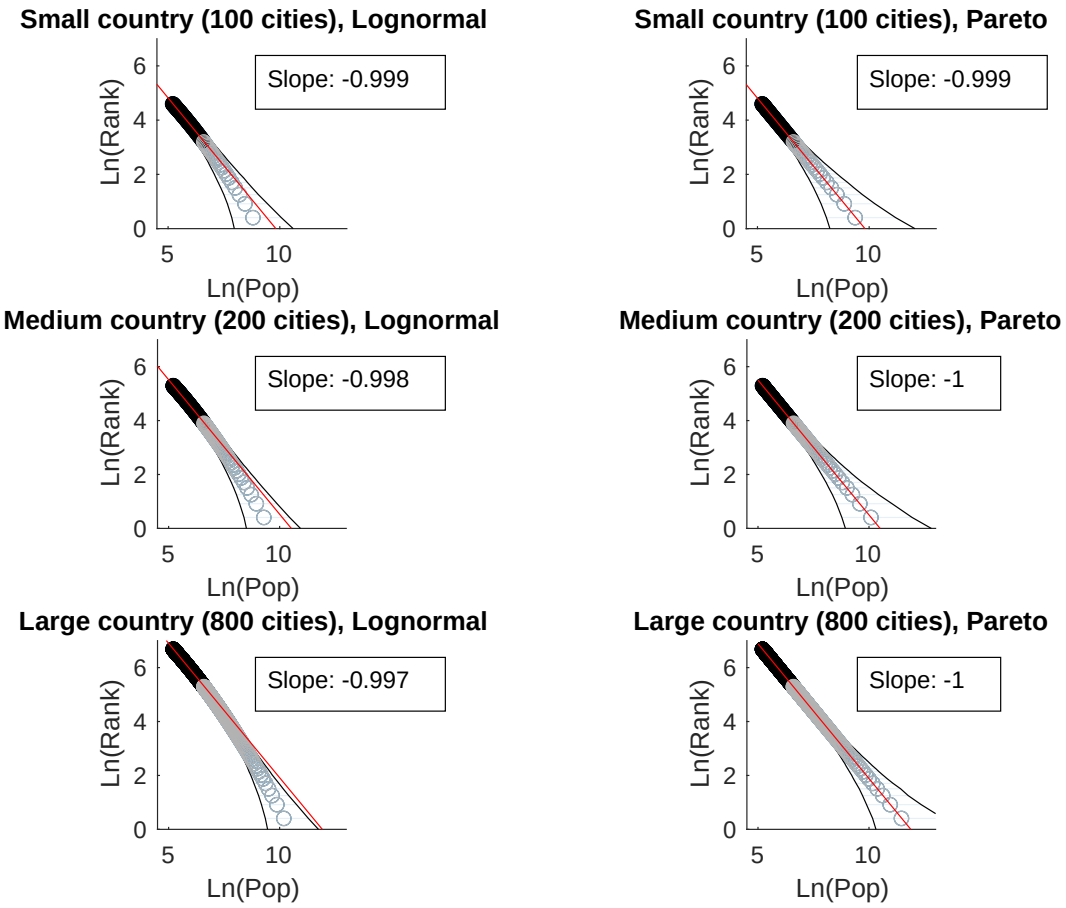
[48]The lognormal parameter $\sigma_{LN} = 2.6$ used for these simulations is similar to that resulting from simulation of the model (in Section 4) for standard parameter values in the literature. This value is larger than that identified by Eeckhout (2004) (who finds $\sigma_{LN} = 1.75$). The difference could partially be attributed to differing truncation points, along with the empirical difficulty of evaluating the population of small locations and the lower bound on real-world populations of 1.

Figure A.II:
Truncation Points and Power Law Coefficients



Notes: The top panels (A and B) show the 388 U.S. MSAs with a population over 100k in 2020. The bottom panels (C and D) show the 110 MSAs with a population over 500k in 2020. Panels A and C display the trendline for the full distribution and Panels B and D display the trendline excluding the top 25% of MSAs (in orange) in each panel. Altering the truncation point substantially influences the estimated coefficient, as can be seen by contrasting Panels A and C with Figure I Further, nearly all top quartile MSAs falling below the trendline (94 of 97 MSAs in panel B and 27 of 27 MSAs in panel D are below the respective trendlines). Both are consistent with the U.S. population distribution being lognormal.
Data Source: U.S. Census

Figure A.III:
Comparison of Lognormal and Pareto Distributions of Cities



Notes: Comparison of Lognormal (left) and Pareto (right) for simulated small, medium, and large "countries." The LN is truncated for cities 2 standard deviations above μ_{LN}, and the Pareto has a minimum value equivalent to this truncation point with shape parameter $\alpha_P = 1$. The slope in each plot is calculated excluding the top 25% of cities in each country, and the bands contain 95% of the observed values at each rank over 1000 simulations. At small scales, the lognormal distribution at Pareto distribution are largely indistinguishable. However, scale variance of the LN leads to substantial divergence in the tail. At larger scales (large countries with more large cities), if the distribution is draw from a LN distribution the large cities tend to fall below the trendline (with trend above the 95% band) while the the Pareto distribution does not diverge.

C Data and Empirical Results

C.1 Data

In this section, we list the variables we used in Section 3 for our correlation matrices and tables. The data come from the publicly-available data of Henderson et al. (2018).

1. *Ruggedness:* index measure of local variation in elevation. Originally computed by Nunn and Puga (2012) with corrections made in Henderson et al. (2018).
2. *Elevation:* above sea level, meters
3. *Temperature:* average from 1960-1990 of monthly temperatures, Celsius
4. *Precipitation:* average from 1960-1990 of monthly total precipitation, mm/month
5. *Land Suitability:* propensity of an area of land to be under cultivation based on separate measures of climate and soil quality
6. *Distance to Coast:* distance to the nearest coast, km
7. *Distance to Harbor:* distance to nearest natural harbor on the coast, km (great circle)
8. *Distance to River:* distance to nearest navigable river, km
9. *Malaria:* index of the stability of malaria transmission
10. *Land Area:* grid cell area covered by land, km^2
11. *Growing Days:* Length of agricultural growing period, days/year

C.2 Summary Statistics

We provide summary statistics of our attributes in Table A.I.

Table A.I: Summary Statistics for Attributes

Variable	N	Min	Max	Median	Q1	Q3	IQR	Mean	SD
Ruggedness	13,426	−2.65	1.79	0.06	−0.69	0.71	1.40	0.00	1.00
Elevation	13,426	−3.02	1.99	−0.06	−0.71	0.92	1.63	0.00	1.00
Land Suitability	13,426	−2.77	1.13	0.14	−0.55	0.91	1.46	0.00	1.00
Dist to River	13,426	−1.38	10.75	−0.25	−0.69	0.39	1.08	0.00	1.00
Dist to Coast	13,426	−0.90	4.98	−0.32	−0.60	0.23	0.84	0.00	1.00
Temperature	13,426	−3.15	2.29	−0.08	−0.76	0.83	1.60	0.00	1.00
Precipitation	13,426	−2.53	3.46	−0.12	−0.89	0.78	1.67	0.00	1.00
Dist to Harbor	13,426	−1.16	5.38	−0.32	−0.74	0.46	1.21	0.00	1.00
Growing Days	13,426	−3.12	1.54	−0.04	−0.84	0.88	1.72	0.00	1.00
Malaria Ecology	13,426	−5.60	0.56	0.51	−0.35	0.56	0.91	0.00	1.00
Land Area	13,426	−17.66	0.59	0.16	−0.03	0.32	0.34	0.00	1.00

Notes: Summary statistics for ordered geographic attributes for data points in the contiguous United States.
Data Source: Authors' calculations using data from Henderson et al. (2018)

C.3 Correlation Calculations

Cross-Correlations For every attribute type g, we calculate

$corr(\boldsymbol{a_i}, \boldsymbol{a_j}),\ \forall i, j \in \mathcal{G},\ j \neq g$, where $\boldsymbol{a_g}$ is a vector for each attribute type comprised of attribute values a_{ig} for every location i. This exercise tells us how correlated each attribute is with each other attribute within locations, giving an indication of how dependent realizations of geographic attributes may be on one another.[49]

Spatial Correlation To calculate spatial correlations within attributes, we construct rings at varying distances d (in miles) from every grid cell i in the contiguous U.S.; we refer to a location i around which rings are being drawn as a *centroid*. These rings define a collection of grid points in the U.S., Canada, and Mexico at a given buffered distance $(d-10, d+10)$ for each centroid.[50] We then select a random point, called $i^*(d, g)$, from within each ring of distance d from every centroid i, to construct our sets of points to calculate the correlations; we re-draw a random point for each attribute type g for every centroid. Mathematically, our calculation for the correlation within an attribute type g between our set of centroids and our set of points at distance d takes the form $corr(\boldsymbol{a_g}, \boldsymbol{a_{dg}}),\ \forall g \in \mathcal{G},\ \forall d$, where $\boldsymbol{a_g}$ is a vector of attribute values a_{ig} for attribute type g for all centroid locations i in our sample, and $\boldsymbol{a_{dg}}$ is a vector of all attribute values $a_{i^*(d,g)g}$, the randomly-selected points for each centroid i at distance d for each attribute type g.

[49] We do not know the full suite of attributes that characterize a location's productivity, and in our limited panel we have some attributes which are mechanically correlated within a location (such as average temperature and growing days).

[50] The spatial correlation in attributes between points at distance $d = 100$ miles should be interpreted as "the correlation between a point and a randomly-selected point 90–110 miles away". The buffer is to ensure there are eligible points at roughly the desired distance.

C.4 Empirics of Attribute Correlations

We empirically investigate the correlation of pairs of attributes within locations and the correlation of attributes across space. We provide support for our assumption of α-mixing for attributes within a place, used to apply the central limit theorem above to characterize the fundamentals, and the assumption of α-mixing of attributes across space, which will be used in Section 3 to characterize the population distribution.

We use gridded geographic data from Henderson et al. (2018), which includes a wide variety of first-nature geographic attributes of which we use the eleven continuous variables.[51] The dataset is at the quarter-degree latitude and longitude cell level.[52] We focus on the roughly 47,000 cells grid cells in the U.S., Mexico, and Canada, with the nearly 13,000 of those grid cells contained in the contiguous U.S. serving as our main sample.[53]

First, we calculate the correlation between attributes within a given location and show that weak dependence of attributes is a reasonable assumption, as there exist pairs of attributes which do not appear correlated within places. We calculate cross-correlations between our attributes for all grid cells in the contiguous U.S., as shown in Figure A.IVa.[54] Our results show that the assumption of weak dependence of attributes appears reasonable given the pattern of correlations of attributes within locations. While there appears to be some correlation between some pairs of attributes within locations, the median correlation among the least-correlated attribute pairs is very near 0.

Next, we demonstrate that while there is correlation within each attribute across space,

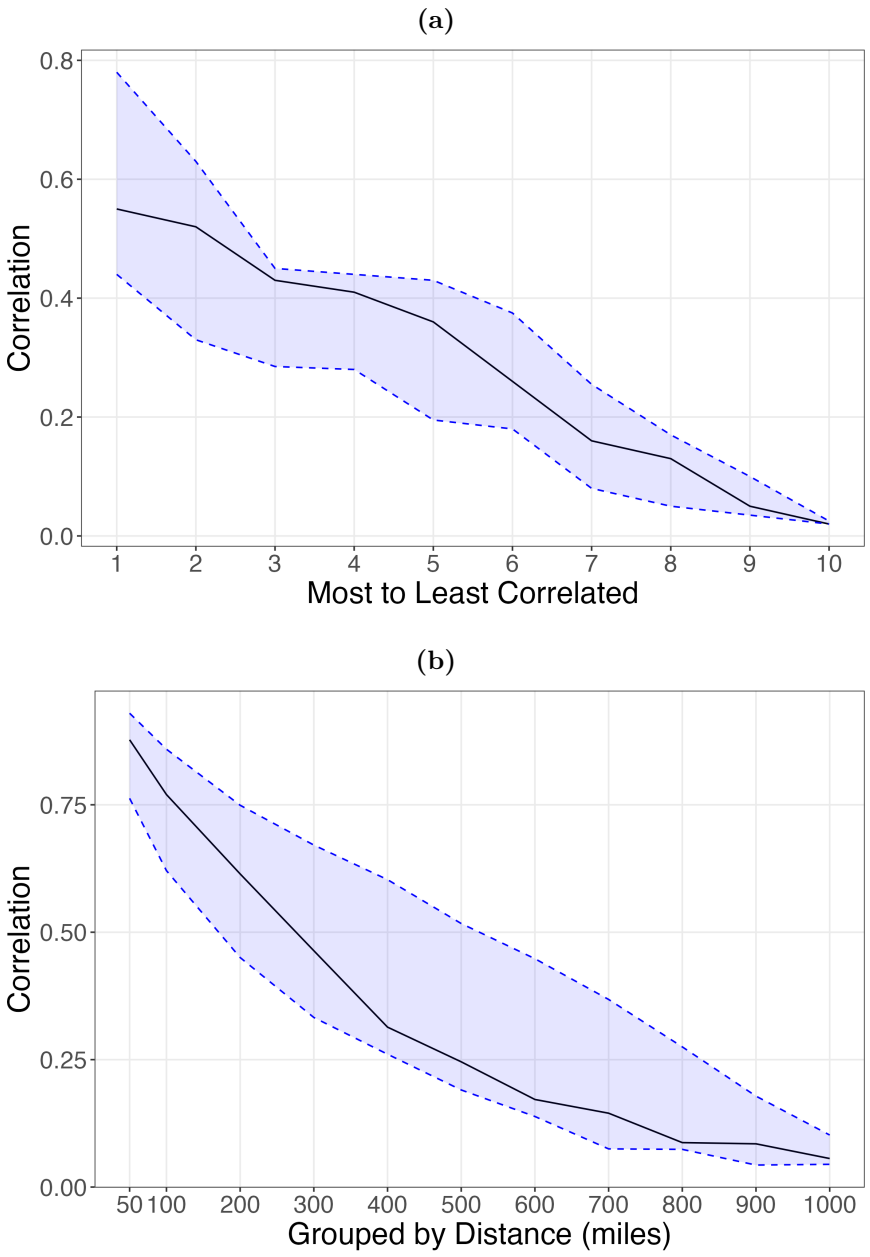
[51] The variables are ruggedness, elevation, land suitability for cultivation, distance to a river, distance to an ocean coast, average monthly temperature, average monthly precipitation, distance to a natural harbor, growing days per year, an index of malaria, and total land area of the grid cell. Variables in Henderson et al. (2018) which were either categorical or discrete transformations of the continuous data were excluded from our analysis.

[52] At the equator, a grid cell is ≈28-by-28 km; at 48 degrees latitude, ≈18-by-18 km.

[53] The reduction in the number of attributes and geographic scope does not drastically decrease the explanatory power of the attributes on economic activity relative to Henderson et al. (2018); see Online Appendix Table A.III for a regression showing that our eleven attributes explain 43% of the variance in economic activity in the contiguous U.S., in line with the 47% Henderson et al. (2018) found globally with their full set of attributes.

[54] A description of how we calculated cross-correlation is provided in Online Appendix C.3

Figure A.IV:
Correlations Within and Across Locations



Note: (a) Cross-correlation and (b) spatial correlation structure of U.S. geographic attributes. The solid black line represents the median correlation; the blue dashed lines represent the 25th and 75th percentile bands.

Data Source: Authors' calculations using data from Henderson et al. (2018)

this correlation declines to zero as distance increases. We calculate spatial correlation at

various distances for grid points within the contiguous U.S., as seen in Figure A.IVb.[55] The spatial correlation of attributes is high over short distances but as distance increases spatial correlation falls to near zero. These results suggest spatial correlation of geographic attributes does decline with distance, supporting the assumption that the fundamentals will exhibit a similar pattern of declining correlation across space.

[55]A description of how we calculated spatial correlation is provided in Online Appendix C.3.

C.5 Calculating the Fundamental

For every attribute in our data set which has a minimum value less than or equal to 0, we re-define the attribute using an affine transformation to put the minimum ≈0.1. We then construct the "worst-to-best" ordering of our attribute values according to the sign on each attribute from a regression on attribute influence on economic activity as found in Henderson et al. (2018), Table 1. Attributes whose sign was positive we perform no additional transformations to. Attributes whose sign was negative we invert. We use the signs present in that table, as opposed to signs from a smaller regression of our subset of attributes on U.S. economic activity, because we believe the signs in their regression could plausibly be more robust world-wide.

After choosing our attribute value ordering, we then standardize the natural log of our attributes:

$$\frac{\ln(a_{ig}) - \text{mean}\left(\ln(a_g)\right)}{\text{sd}(\ln(a_g))}$$

where mean $(\ln(a_g))$ is the mean of that attribute across all locations and sd $\ln(a_g)$ is the standard deviation. This produces logged attributes which are mean 0 and standard deviation 1.

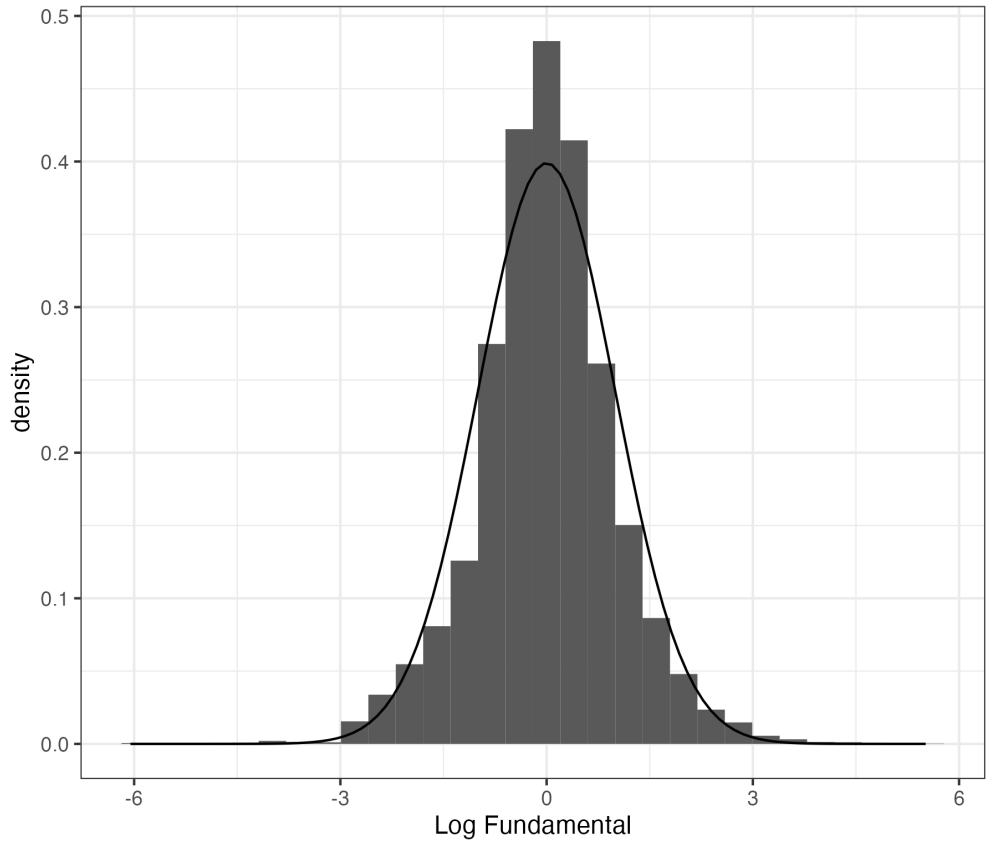
We then aggregate our attributes into a fundamental given by Equation 22:

$$\ln(A_i) = \sum_{g \in \mathcal{G}} \xi_g \ln(a_{ig})$$

setting $\xi_g = 1,\ \forall g$.

C.6 Results: Fundamental and Correlations

Figure A.V:
Lognormal Distribution of Locational Fundamentals for Contiguous United States



Notes: Lognormal distribution of locational fundamentals. All eleven attributes were ordered worst to best in terms of contribution to economic activity, logged, then standardized. The fundamental is calculated as the standardized sum of the standardized, ordered log attributes. The mean and variance are standardized to zero and one and a standard normal curve is overlaid.
Data Source: Authors' calculations using data from Henderson et al. (2018)

We plot in Figure A.V the empirical PDF of the resulting distribution of productivity fundamentals, calculated according to Equation 22 with $\xi_g = 1$, $\forall g$. The log of the empirical "fundamental" here is closely fit by a normal distribution, supporting the claim that aggregating weakly dependent and spatially correlated attributes can result in lognormal fundamentals, both in theory and in the data.

Additionally, we calculate the spatial correlation of the logged fundamental over distance

Table A.II: Spatial Correlation of Calculated Fundamental for the Contiguous United States

Distance (miles)	Correlation
50	0.66
100	0.50
200	0.30
300	0.17
400	0.05
500	0.02
600	-0.01
700	-0.06
800	-0.07
900	-0.04
1000	-0.03

Notes: The table indicates declining correlation towards 0 over distance of our logged fundamental, in line with theoretical predictions. Given randomness in the correlation calculation process, spurious and small deviations from 0 at large distances are possible.
Data Source: Authors' calculations based on data from Henderson et al. (2018).

for the contiguous U.S. The table of correlation values, provided in Table A.II, shows declining correlation over distance, consistent with our theoretical predictions and in line with the spatial correlation declines which appear for geographic attributes (provided in the main text).

C.7 Regression of Economic Activity on Attributes

We provide regression results in Table A.III for economic activity on our eleven attributes for a sample including all points in the contiguous United States, interpreting the R^2 as suggestive evidence of the reasonably large explanatory power first nature geographic attributes on economic activity.

Table A.III: Regression of Economic Activity on Attributes

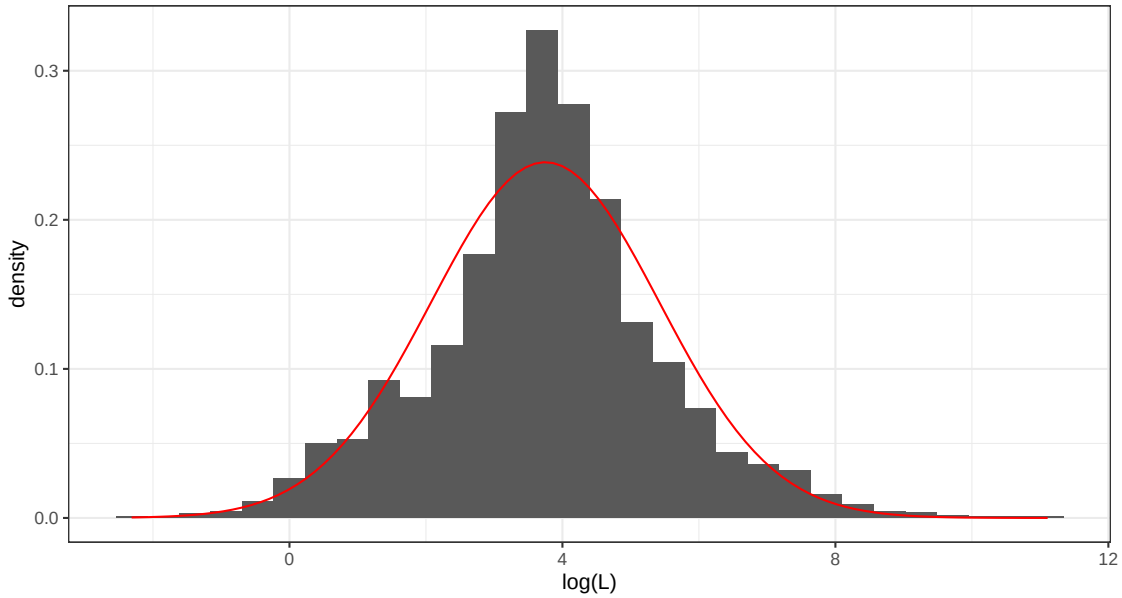
	Dependent variable:
	(Log of) Radiance
(Inv) Ruggedness	0.104
	(0.020)
Elevation	−1.279
	(0.057)
Land Suitability	0.298
	(0.039)
(Inv) Dist to River	−0.079
	(0.024)
(Inv) Dist to Coast	−0.135
	(0.034)
Temperature	−0.607
	(0.195)
(Inv) Precipitation	0.357
	(0.082)
(Inv) Dist to Harbor	0.006
	(0.049)
Growing Days	1.722
	(0.090)
(Inv) Malaria	−0.060
	(0.046)
Land Area	0.401
	(0.068)
Constant	−0.340
	(1.102)
Observations	13,426
R^2	0.431
Adjusted R^2	0.430
Residual Std. Error	2.294 (df = 13414)
F Statistic	922.395 (df = 11; 13414)

Notes: Grid-cell radiant lights on attributes, contiguous United States. (Inv) indicates the attribute data was inverted. The R^2 from this regression is comparable to the main regression from Henderson et al. (2018). *Data Source:* Authors' calculations based on data from Henderson et al. (2018)

C.8 Recovered fundamentals from Allen and Arkolakis (2014)

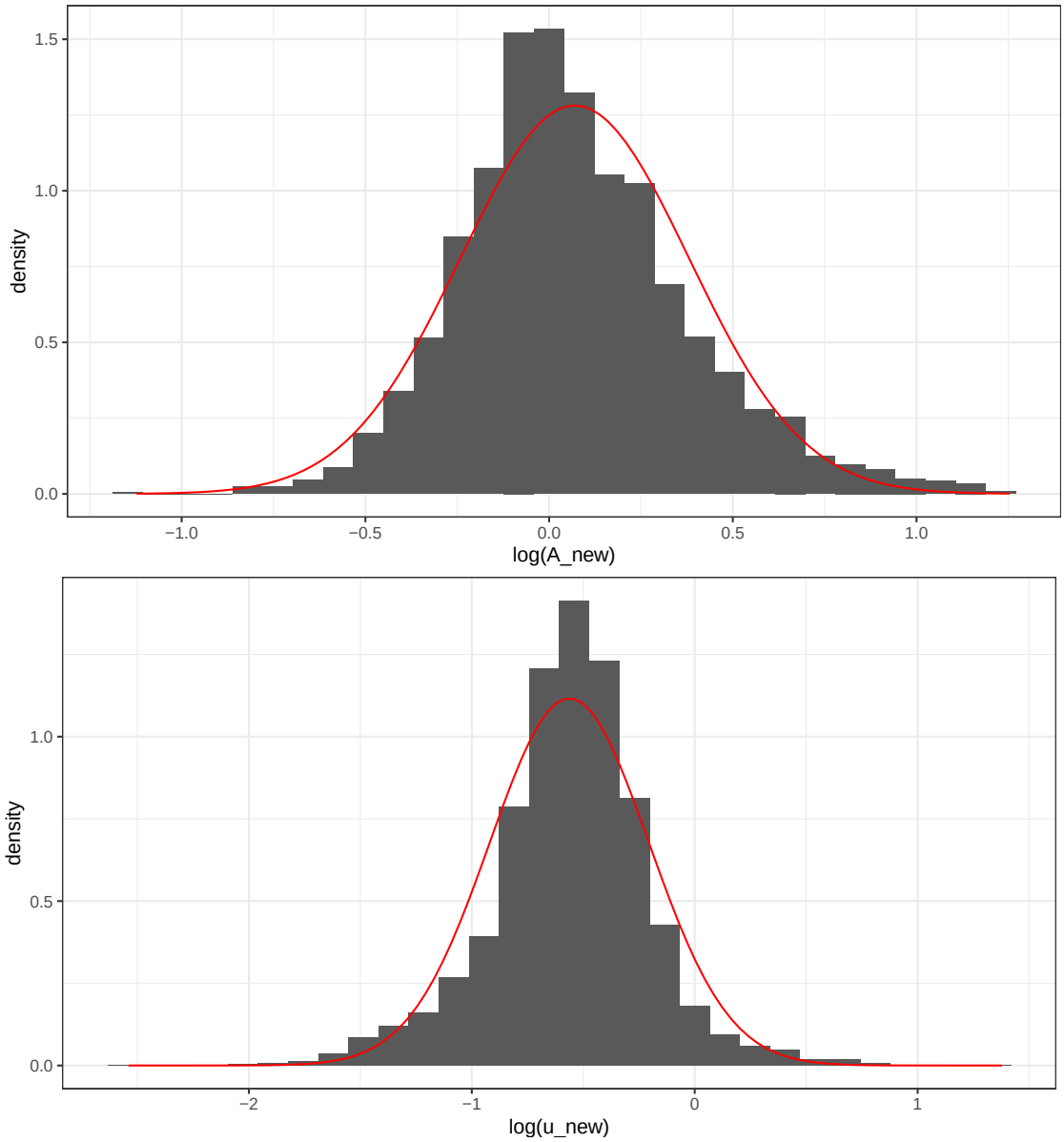
We plot the exogenous productivity and amenity values recovered in Allen and Arkolakis (2014). These are at the county-level. County populations are roughly lognormally distributed, as seen in Figure A.VI. The resulting fundamentals recovered by Allen and Arkolakis (2014), as seen in Figure A.VII, also appear roughly lognormally distributed. The correlation between A and u is $\approx$0.12, which one can use to parameterize our simulations.

Figure A.VI:
Distribution of Logged County Population in the United States



Notes: Distribution of logged county populations (L) in the United States, with super-imposed normal curve. Number of counties N=3,109.
Data Source: online replication package from Allen and Arkolakis (2014).

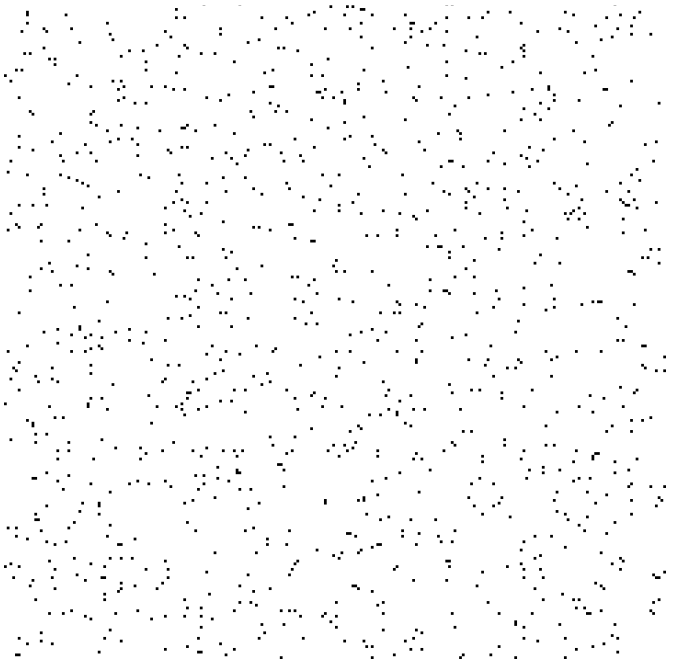
Figure A.VII:
Distribution of Fundamentals in United States



Notes: Top panel shows distribution of logged exogenous productivities (A_{new}) at the county-level in the United States, with super-imposed normal curve. Bottom panel shows distribution of logged exogenous amenities (u_{new}) at the county-level in the United States, with super-imposed normal curve. Number of counties N=3,109.

Data Source: online replication package from Allen and Arkolakis (2014).

Figure A.VIII:
Example of "dartboard" geography



Notes: The figure shows the middle settlements of our full geography to provide an example of how we induce variation in trade costs.

D Additional Simulated Results

D.1 Example of "dartboard" geography

In our simulated results we first randomly generate a dartboard geography to ensure dispersion in trade costs. We take draws from uniform distributions with parameters $[0, a]$ where a reflects the maximal horizontal and vertical dimension of our full square geography. We take draws until we have realized a set number of locations within the full geography. Figure A.VIII shows the locations in the center of our full geography to provide an example of the realized "dartboard." Using a dartboard geography is convenient because it allows us to ensure trade costs respect the triangle inequality, ensures random variation in trade costs, and obviates the need to calculate least-costs paths beyond taking the Euclidean distance between points.

D.2 The distribution of S_i

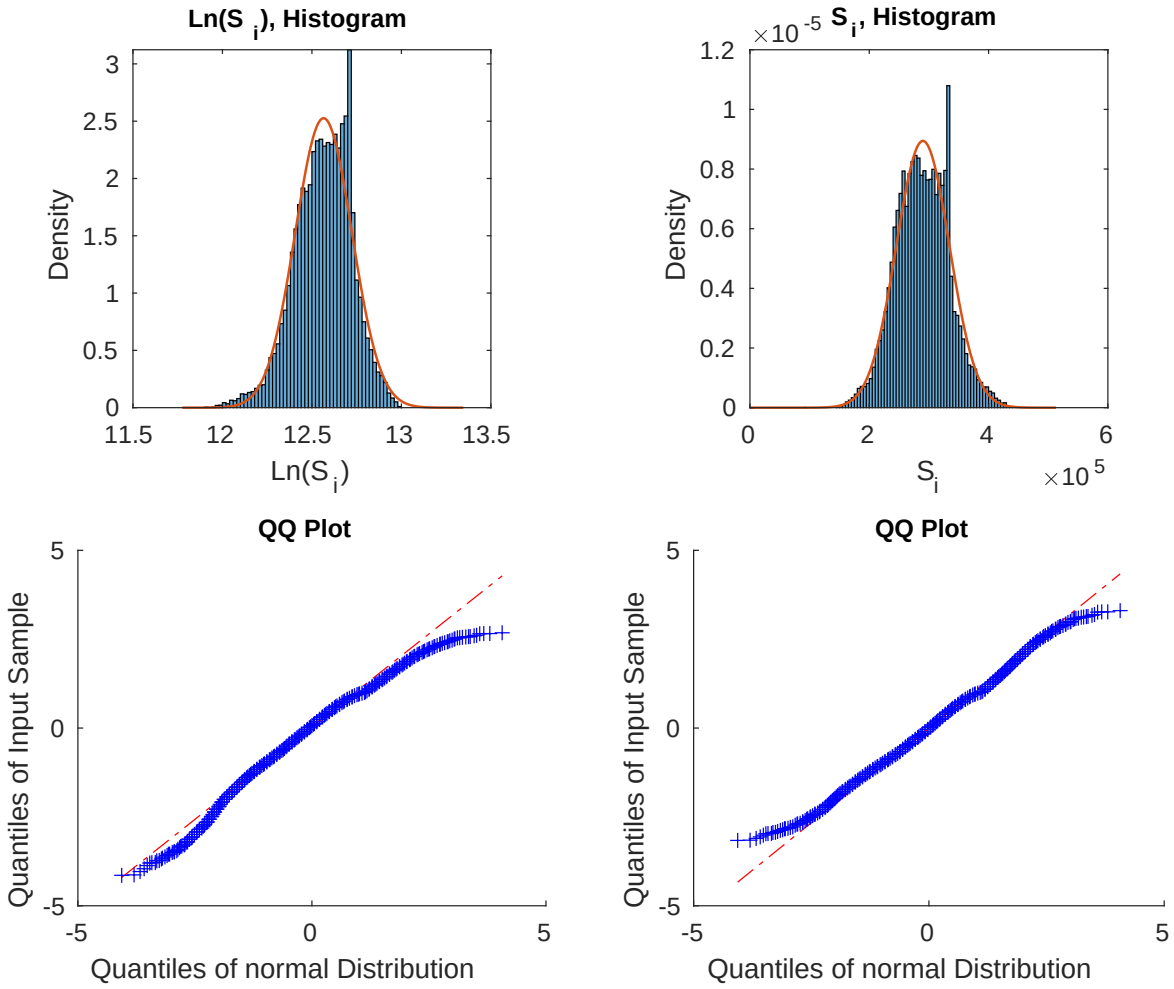
Our proof has two implications for the distribution of S_i, the "market access" term in the equilibrium population condition, in simulation that we verify here. First, by Lemma 1, Theorem 1 implies that S_i will appear distributed normally in both levels and logs. Second, Theorem 1 and Theorem 2 require that $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ is independent of Ω_i.

We demonstrate that the first property holds by plotting the histogram of values of S_i for a given geography in both logs and levels, along with the respective QQ plots comparing both to a normal distribution, in Figure A.IX. The distribution in Figure A.IX displays the expected patterns, with a good fit to the normal distribution in both levels and logs as implied by Theorem 1. The smaller than expected right tail may be attributable to too little dispersion in trade costs or the size of the grid we simulate.

To check independence, we verify that the average correlation between Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ over the 1000 replications is 0.017, indicating the two terms are virtually uncorrelated as implied by the proof. While not a sufficient condition for independence, no correlation is necessary condition.

We take these simulated results as evidence that the mechanism we emphasize in our proofs is the mechanism driving the result. Further, we note that the setting we used for our simulation has very little variation in trade costs compared to a more realistic geography. Greater variation should be expected to increase the fit of the simulated S_i to the expected distribution and reduce the correlation of Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$, as shown in Online Appendix D.4.

Figure A.IX:
Distribution of Sums



Notes: The figures above show a realization of the vector of S_i terms for some exogenous geography and associated population vector. The distribution of S_i appears normal in both levels and logs.

D.3 Additional simulations: Varying Parameter Values

In the simulations for varying parameter values, we take all combinations of $\alpha = [0.02, 0.04, 0.06, 0.08, 0.1]$, $\beta = [-0.25, -0.30, -0.35, -0.40, -0.45, -0.50]$, and $\delta_{TC} = [0.0005, 0.001, 0.0015, 0.002, 0.0025]$. For each combination, we find the population distribution for 100 draws of the exogenous geography in a grid of the same dimensions as for our main results. We set the number of locations in the full geography to 10000.

We here report the full table of estimated parameter values for three values of δ_{TC}, at 0.0005, 0.0015, and 0.0025. When increasing α, the coefficient tends to increase (flatter slope). When increasing β, the coefficient tends to decrease (steeper slope). When increasing δ_{TC}, the coefficient tends to increase (flatter slope). As can be seen in Tables A.IV (showing average coefficients) and A.V (showing median coefficients) the change is nearly always in the anticipated direction.

To get the comparative statics reported in Table II, using the average estimated coefficient $\hat{\theta}_i$ from each of the 150 combinations of parameters we estimate the following regression:

$$\hat{\theta}_i = \psi_1 \alpha_i + \psi_2 \beta_i + \psi_3 \delta_{TC,i} + \epsilon_i \tag{32}$$

where the estimated coefficients $\hat{\psi}_1$, $\hat{\psi}_2$, and $\hat{\psi}_3$ are estimates of $\frac{\partial \theta_1}{\partial \alpha}$, $\frac{\partial \theta_1}{\partial \beta}$, and $\frac{\partial \theta_1}{\partial \theta}$, respectively. The signs of these coefficients are reflected in Table II, and the full regression output is included in Table A.VI.

Notably, the estimated coefficients are consistently in the neighborhood of -1 throughout the parameter space we simulate here. Given alternative truncations of the distribution (either expanding or reducing the number of locations included, as discussed in Gibrat (1931)) achieving a -1 slope is likely possible for most of these parameter combinations.

Table A.IV: Changes to Mean Power Law Coefficient with Respect to Model Parameters

Panel A: $\delta_{TC} = 0.0005$

	$\beta = 0.25$	$\beta = 0.30$	$\beta = 0.35$	$\beta = 0.40$	$\beta = 0.45$	$\beta = 0.50$
$\alpha = 0.02$	–0.750	–0.842	–0.928	–1.013	–1.109	–1.181
	(0.034)	(0.030)	(0.036)	(0.043)	(0.035)	(0.046)
$\alpha = 0.04$	–0.718	–0.813	–0.899	–0.980	–1.075	–1.152
	(0.027)	(0.032)	(0.037)	(0.036)	(0.044)	(0.040)
$\alpha = 0.06$	–0.696	–0.778	–0.875	–0.953	–1.050	–1.134
	(0.029)	(0.030)	(0.034)	(0.039)	(0.035)	(0.050)
$\alpha = 0.08$	–0.665	–0.757	–0.841	–0.928	–1.014	–1.101
	(0.026)	(0.028)	(0.035)	(0.042)	(0.038)	(0.042)
$\alpha = 0.10$	–0.640	–0.724	–0.819	–0.899	–0.990	–1.072
	(0.029)	(0.029)	(0.029)	(0.036)	(0.036)	(0.041)

Panel B: $\delta_{TC} = 0.0015$

	β = -0.25	β = -0.30	β = -0.35	β = -0.40	β = -0.45	β = -0.50
$\alpha = 0.02$	–0.736	–0.828	–0.917	–1.002	–1.091	–1.183
	(0.032)	(0.032)	(0.038)	(0.038)	(0.041)	(0.051)
$\alpha = 0.04$	–0.708	–0.794	–0.886	–0.969	–1.052	–1.142
	(0.032)	(0.032)	(0.037)	(0.042)	(0.042)	(0.046)
$\alpha = 0.06$	–0.683	–0.769	–0.852	–0.942	–1.025	–1.116
	(0.032)	(0.037)	(0.036)	(0.038)	(0.043)	(0.045)
$\alpha = 0.08$	–0.640	–0.738	–0.827	–0.918	–1.008	–1.094
	(0.030)	(0.031)	(0.032)	(0.041)	(0.043)	(0.043)
$\alpha = 0.10$	–0.614	–0.711	–0.799	–0.889	–0.977	–1.066
	(0.035)	(0.030)	(0.037)	(0.036)	(0.040)	(0.040)

Panel C: $\delta_{TC} = 0.0025$

	β = -0.25	β = -0.30	β = -0.35	β = -0.40	β = -0.45	β = -0.50
$\alpha = 0.02$	–0.714	–0.811	–0.901	–0.982	–1.080	–1.166
	(0.047)	(0.035)	(0.039)	(0.043)	(0.046)	(0.046)
$\alpha = 0.04$	–0.688	–0.772	–0.863	–0.959	–1.044	–1.127
	(0.033)	(0.035)	(0.038)	(0.038)	(0.045)	(0.049)
$\alpha = 0.06$	–0.657	–0.747	–0.830	–0.922	–1.021	–1.106
	(0.034)	(0.033)	(0.041)	(0.038)	(0.047)	(0.051)
$\alpha = 0.08$	–0.624	–0.717	–0.804	–0.886	–0.986	–1.075
	(0.032)	(0.035)	(0.035)	(0.034)	(0.041)	(0.050)
$\alpha = 0.10$	–0.593	–0.683	–0.776	–0.867	–0.954	–1.044
	(0.034)	(0.032)	(0.034)	(0.042)	(0.035)	(0.042)

Notes: Table containing the average coefficient over 100 simulations on the exogenous geography. The standard deviation of the estimated coefficient over the 100 simulations is in parenthesis.

Table A.V: Changes to Median Power Law Coefficient with Respect to Model Parameters

Panel A: $\delta_{TC} = 0.0005$

	β = -0.25	β = -0.30	β = -0.35	β = -0.40	β = -0.45	β = -0.50
$\alpha = 0.02$	–0.748	–0.850	–0.928	–1.023	–1.113	–1.191
$\alpha = 0.04$	–0.719	–0.813	–0.903	–0.986	–1.077	–1.156
$\alpha = 0.06$	–0.703	–0.789	–0.879	–0.955	–1.051	–1.139
$\alpha = 0.08$	–0.672	–0.763	–0.851	–0.930	–1.019	–1.109
$\alpha = 0.10$	–0.648	–0.725	–0.823	–0.904	–0.993	–1.078

Panel B: $\delta_{TC} = 0.0015$

	β = -0.25	β = -0.30	β = -0.35	β = -0.40	β = -0.45	β = -0.50
$\alpha = 0.02$	–0.742	–0.833	–0.918	–1.004	–1.093	–1.190
$\alpha = 0.04$	–0.714	–0.790	–0.898	–0.976	–1.057	–1.151
$\alpha = 0.06$	–0.689	–0.770	–0.859	–0.951	–1.027	–1.119
$\alpha = 0.08$	–0.658	–0.741	–0.829	–0.932	–1.015	–1.100
$\alpha = 0.10$	–0.618	–0.713	–0.801	–0.900	–0.987	–1.075

Panel C: $\delta_{TC} = 0.0025$

	β = -0.25	β = -0.30	β = -0.35	β = -0.40	β = -0.45	β = -0.50
$\alpha = 0.02$	–0.733	–0.817	–0.903	–0.992	–1.089	–1.170
$\alpha = 0.04$	–0.697	–0.765	–0.865	–0.965	–1.046	–1.141
$\alpha = 0.06$	–0.671	–0.756	–0.830	–0.925	–1.027	–1.117
$\alpha = 0.08$	–0.610	–0.728	–0.808	–0.895	–0.996	–1.089
$\alpha = 0.10$	–0.601	–0.692	–0.783	–0.881	–0.957	–1.061

Notes: Table containing the median coefficient over 100 simulations on the exogenous geography.

Table A.VI: Estimated coefficients from regression 32

	(1)
α	1.449***
	(0.0131)
β	1.769***
	(0.00433)
δ_{TC}	16.72***
	(0.523)
adj. R^2	0.999
N	150

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.VII: Normality Tests, "Idiosyncratic" Distance Simulations

	Kolmogorov-Smirnov	Lilliefors	Jarque-Bera
Rejected at 1%	0.000	0.010	0.013
Rejected at 5%	0.000	0.042	0.075

Notes: Table shows the share of tests for a normal distribution rejected for the log equilibrium population of 1000 Monte-Carlo simulations.

D.4 Simulation with idiosyncratic shocks to effective distance

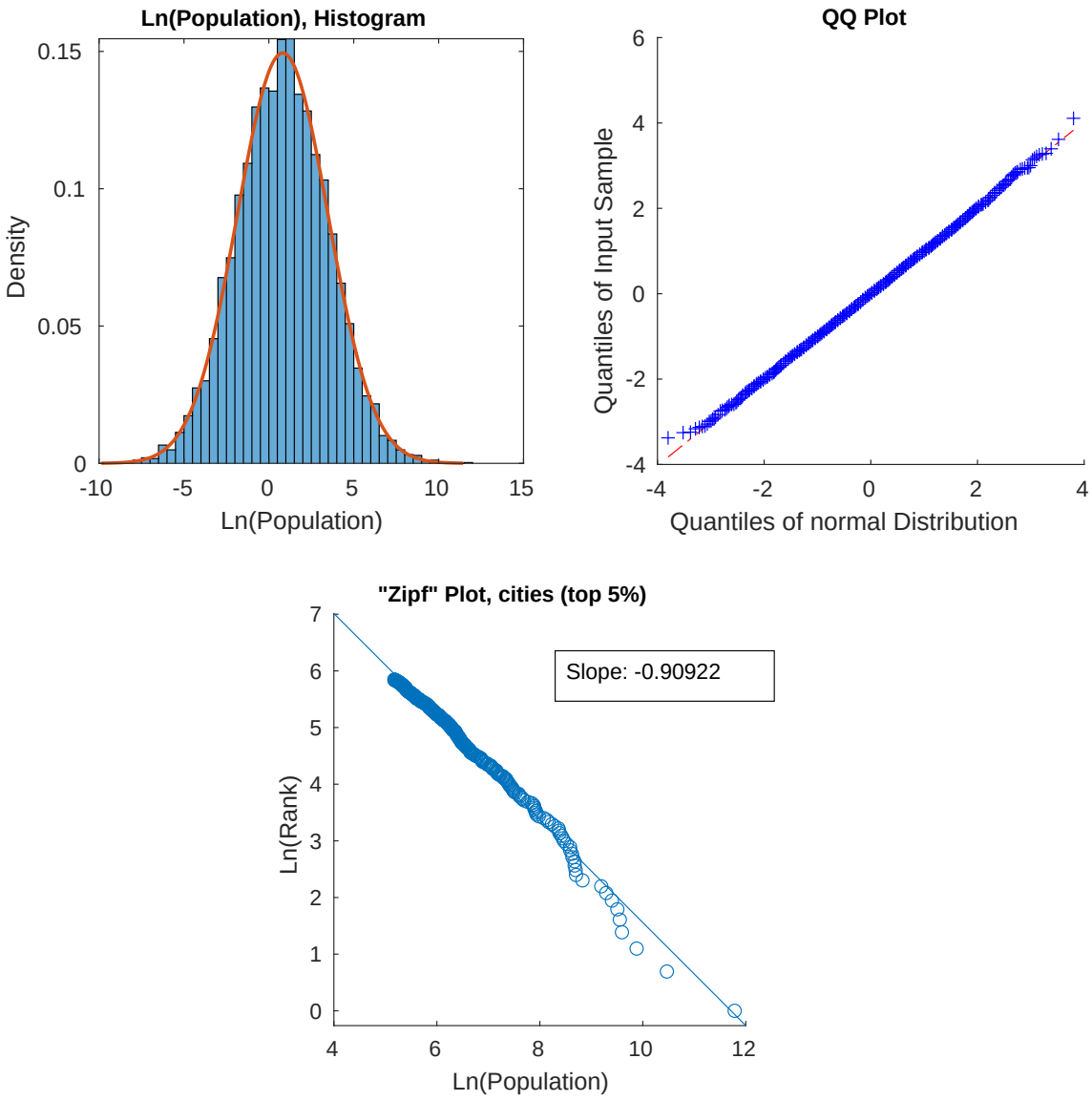
Variation in trade costs across locations are likely substantially greater than in the simple geography we simulate in our main specification. Consider the low trade costs for locations along natural bodies of water or along man-made transportation networks such as canals, train lines, or roadways. Trade costs can also vary for locations-specific idiosyncratic reasons such as policy (tariffs on broad categories of goods and other trade barriers) or bilateral factors that encourage or discourage trade (tariffs on particular countries or shared language).

In this appendix section we show that inducing additional variation in realized trade costs improves the simulated results relative to the more restrictive baseline we presented in the main paper. We here consider the potential for trade costs to vary for idiosyncratic location-specific and bilateral reasons.

We here simulate with $\tau_{i,n} = \exp(\delta_{TC}\beta_i\beta_{i,n}d_{i,n})$, where $\beta_i \in (0,1)$ and $\beta_{i,n} \in (0,1)$. We draw all β_i and $\beta_{i,n} = \beta_{n,i}$ from a uniform (0,1) distribution. The parameter values are the same as in our baseline results and we simulate for 10,000 locations as in the Monte Carlo results.

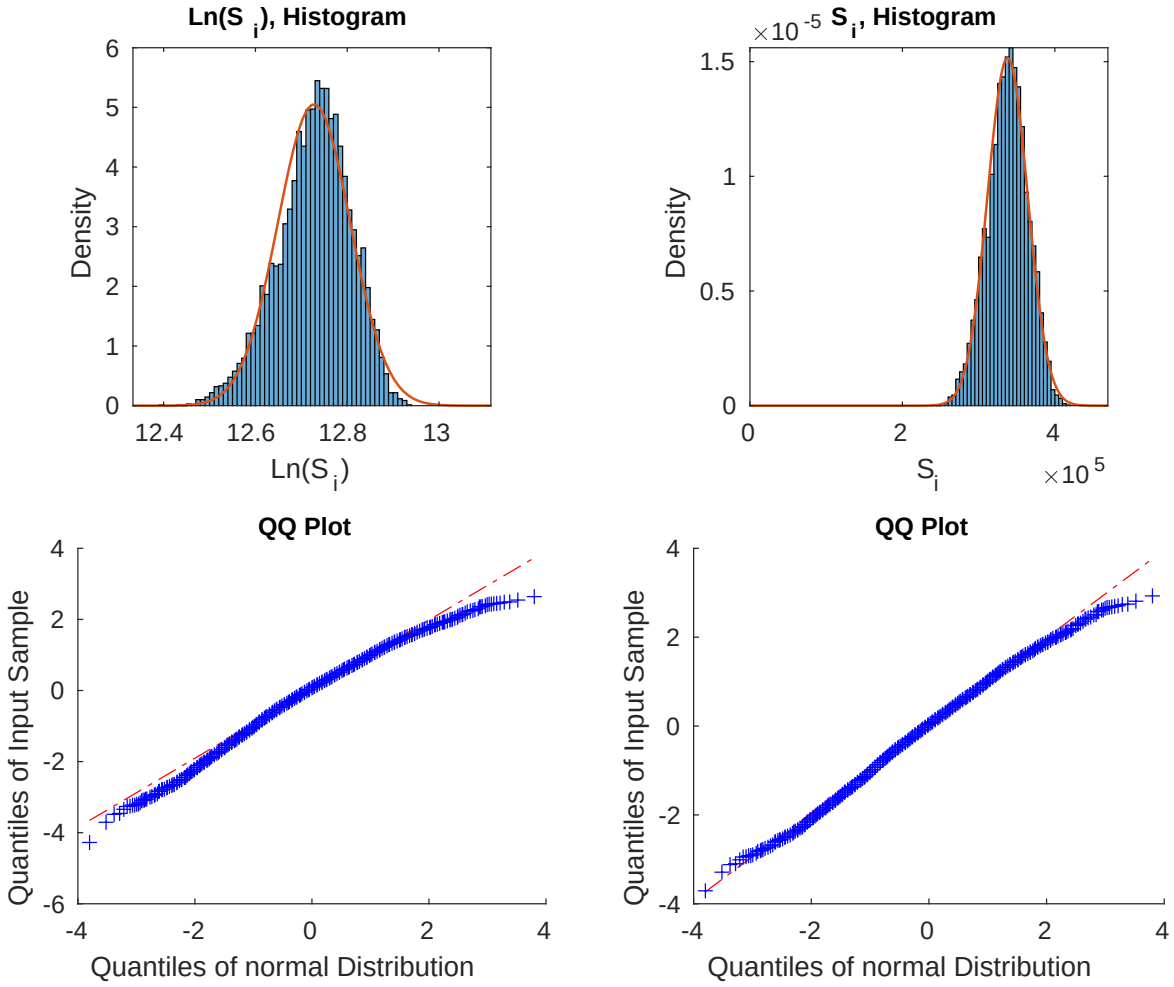
Notably, the lognormality of the resulting population distribution is rejected less frequently than in our baseline results. The distribution of S_i is also appreciably more lognormal when including additional variation in effective distance than in baseline.

Figure A.X:
Example of the Equilibrium Population Distribution, "Idiosyncratic" Distance



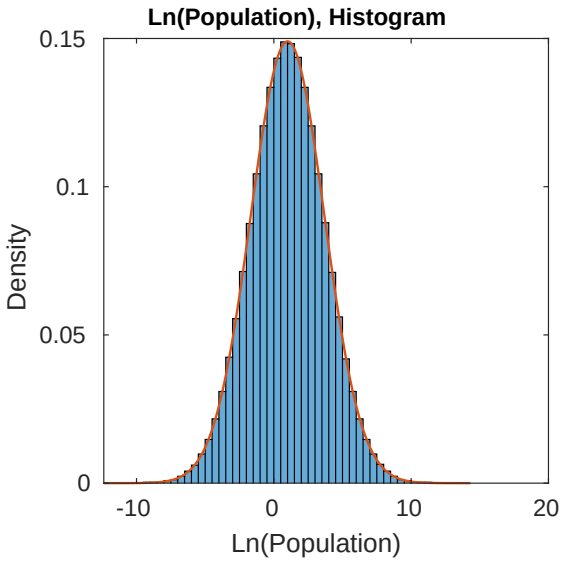
Notes: The top left panel shows the model's log population appears to follow a normal distribution. The top right panel contains a QQ plot of the model's log population distribution, indicating that it very closely matches a normal distribution. The power-law plot at the bottom shows a strong resemblance to the typical log rank-size plot along with the characteristic divergence of the largest locations below the trendline.

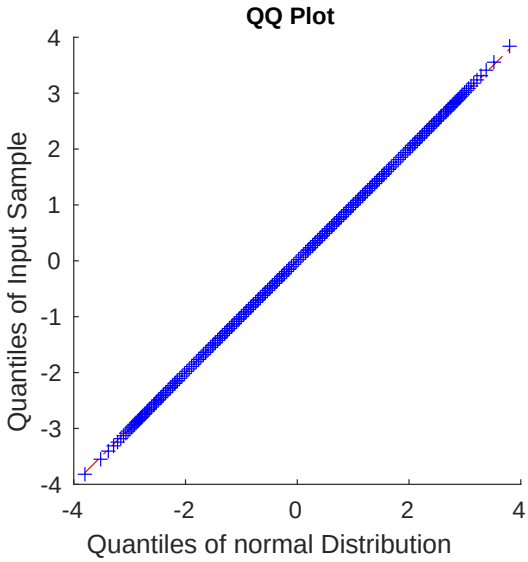
Figure A.XI:
Distribution of S_i, Market Access term, "Idiosyncratic" Distance Simulation

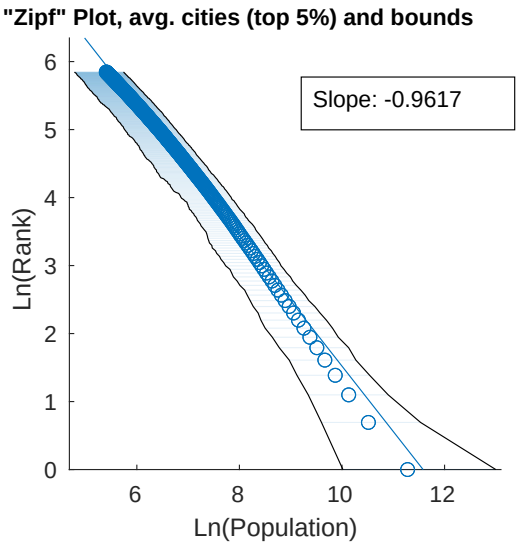


Notes: The figures above show a realization of the vector of S_i terms for some exogenous geography and associated population vector. The distribution of S_i appears normal in both levels and logs.

Figure A.XII:
Smoothed output over 1000 MC simulations, "Idiosyncratic" Dist.







1000-draw MC	Value
Coef. of Avg.	-.962
Avg. over Coefs	-.958
Std. Dev.	(.055)
Min	-1.138
5pct	-1.053
25pct	-.994
50pct	-.957
75pct	-.921
95pct	-.869
Max	-.797

Notes: The population distribution resulting from numerical simulation of model with "effective" distance is in the upper left panel, and the resulting QQ plot is on the upper right. Both show that the equilibrium population distribution appears lognormal. The city size distribution of the Monte Carlo output is in the lower left, and statistics over model simulations the lower right. The slope on the lower left left represents the slope taken over the average of log(pop) at each rank over 1000 simulations, and the bounds contain 95% of the log populations at each rank of the distribution. The table displays statistics over the 1000 estimated power law coefficients from the simulations.