Populations in Spatial Equilibrium*

Matthew Easton
Cornerstone Research

Patrick W. Farrell
Columbia University

This Version: June 2025

Click here for latest version

Abstract

Power law-like distributions for city populations are a distinctive, recurring feature of human settlement patterns. We propose a novel explanation for this phenomenon that reflects the qualities of a place (fundamentals) and its ability to benefit from trade based on its location (market access), two important forces that have not simultaneously been incorporated into an explanation of the city size distribution. Using random variation in geography to model these two terms within a quantitative spatial model results in lognormal population distributions which appear to follow a power law for the most populous locations (i.e., cities).

*Easton: matthew.d.easton@gmail.com. Farrell: pwf2108@columbia.edu. The authors thank Arslan Ali, Nadia Ali, Costas Arkolakis, Pierre-Philippe Combes, Donald R. Davis, Richard A. Davis, Jonathan Dingel, Thibault Fally, Matthieu Gomez, Brian Greaney, Madeline Hansen, Yannis Ioannides, Raphaël Lafrogne-Joussier, Jeffrey Lin, Kosha Modi, Serena Ng, Eshaan Patel, Stephen Redding, Frédéric Robert-Nicoud, Andrés Rodríguez-Clare, Esteban Rossi-Hansberg, Bernard Salanié, Stephan Thies, Lin Tian, Conor Walsh, David Weinstein, Natalie Yang, and seminar and conference participants at Columbia University, the 2024 European Meeting of the Urban Economics Association, and the 2024 North American Meeting of the Urban Economics Association 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 most remarkable empirical regularity in spatial economics is that the upper tail of the city size distribution appears to follow a power law. Why this regularity arises is a question that quantitative spatial models, which form the basis of modern spatial economics, have thus far failed to answer. While a key feature of these models is their ability to recover the unobserved productivities and amenities[1] of each location—the locational "fundamentals"—that rationalize the observed populations, they do not explain why the underlying locational fundamentals and market access should consistently generate the characteristic city size distribution. By contrast, existing explanations of the appearance of a power law are inconsistent with quantitative spatial models as these explanations either allow no role for a location's characteristics (Brakman et al. 1999; Hsu 2012; Rante et al. 2024), no role for a place's interactions with other locations (Lee and Li 2013; Behrens and Robert-Nicoud 2015), or no role for either factor to determine a city's population (Gabaix 1999b). That is, the literature that models the forces shaping cities cannot explain the regularity of the city size distribution, while the literature explaining city size distributions ignores the forces shaping cities.

Our paper shows that random variation in geography and trade costs can generate the characteristic city size distribution within quantitative spatial models, linking the canonical models of the spatial economy to the literature on the population distribution. We model a location's fundamentals as resulting from random variation in its geographic "attributes," exogenous aspects of its geography which affect locational fundamentals. Similarly, we model a location's market access as resulting from random variation in trade costs between locations. Locational fundamentals and market access, we then show, will be lognormally distributed. Together, these will generate a lognormal population distribution within quantitative spatial models. As a lognormal distribution approximates a power law in the upper tail, our framework naturally generates the characteristic power law-like city size distribution. Our key technical advance is to demonstrate that random variation generates a lognormal distribution for market access. This advances prior work, in particular Lee and Li (2013), that has used many random "factors" to generate lognormally distributed fundamentals in a model without trade.

Within the Allen and Arkolakis (2014) quantitative spatial framework, we show that the population of each location can be expressed as the product of two terms. One term represents that location's market access and is written as a trade cost-weighted sum over

[1] Productivities are output shifters in production functions. Amenities are utility shifters that affect the utility one earns at a certain real wage.

all locations. The other term solely reflects that location's fundamentals, which are a function of its attributes. Our major contribution is to characterize the distributions of both the fundamentals and the trade costs influencing market access resulting from random geographic variation, and we accomplish this in two steps. First, we incorporate insights form the prior literature and model the locational fundamentals within a quantitative spatial model as multiplicative aggregates of randomly distributed locational attributes. As multiplicative processes are additive in logs, the locational fundamentals will also be lognormally distributed by a central limit theorem. Second, we provide a novel proof for the distribution of our market access term, applying a result from Marlow (1967) about central limit theorems for sums of positive random variables. We apply this result to the market access summation to demonstrate that the market access terms will also be lognormally distributed. The population distribution, a product of these two lognormal terms, will thus be lognormally distributed.

The separation of the explanation into two parts has attractive economic interpretations because there is substantial empirical evidence that both a place's characteristics and its location in space influence its population. Within quantitative spatial models, the former is reflected by a location's fundamentals and the latter by its market access. Our paper is the first to show how these forces *jointly* produce the distinctive population distribution. This result holds even allowing for other important economic forces captured by these models, such as agglomeration and congestion externalities, to influence the population.

We show that an inversion of the Allen and Arkolakis (2014) model using data for the United States results in lognormally distributed fundamentals and market access, as predicted by our theory. We then simulate a quantitative spatial model to test our framework's robustness and ability to match results in the theoretical and empirical literatures. We show that differences in local productivity spillovers, intra-city congestion externalities, and inter-city transportation costs can explain variation in the city size distribution observed in data. We find that as agglomeration benefits rise the size distribution becomes more unequal, consistent with observed changes to the U.S. city size distribution (Gabaix and Ioannides 2004). Increasing inter-city transportation costs also results in a more unequal city size distribution. This provides a potential explanation for the very large "primate" cities and unequal city size distributions of developing countries, where domestic transportation costs are often much higher than in developed countries (Teravaninthorn and Gaël 2009; Atkin and Donaldson 2015).

Our paper touches on several branches of the spatial economics literature. Many theoretical explanations of the city size distribution are based on the similarly striking empirical observation that city growth rates often appear unrelated to city population, such

as Gabaix (1999a), Gabaix (1999b), Blank and Solomon (2000), Eeckhout (2004), Rossi-Hansberg and Wright (2007), and Córdoba (2008).[2] However, the assumption of random growth is inconsistent with the empirical evidence on the distribution of cities in significant ways. The growth of cities does not appear random in many important cases, particularly following major shocks.[3] Random growth explanations also fail to capture the influence of the characteristics of a place on the attractiveness of producing or residing there, implying that the large populations of New York City, Tokyo, and London are random and unrelated to their advantageous geographies. These theories are also aspatial and do not allow for interactions between different locations to shape settlement patterns, failing to capture the contribution of trade to the scale of the aforementioned global cities.[4]

There are papers within the literature on city size distributions that have accounted for either the role of a place's characteristics or its location in determining its population, but our paper is the first to include these important forces simultaneously. Papers that do account for the importance of geography or fundamentals, such as Lee and Li (2013) and Behrens and Robert-Nicoud (2015), have lacked trade between locations. In contrast, those papers which account for the importance of trade and market access, such as Brakman et al. (1999), Hsu (2012), and Rante et al. (2024), are models that do not allow for variation in the characteristics of locations. By placing our explanation for the distribution's emergence within a quantitative spatial model, we can include roles for both local characteristics and trade between locations. Further, given the properties of quantitative spatial models, we also demonstrate that population distributions will exhibit random growth in equilibrium in response to increases in the aggregate population level. This provides consistency with the observation of random growth in many contexts that motivated prior models.

By integrating the literature on population distributions with that on quantitative spatial models, we show the conditions under which these models generate the characteristic city size distribution. Models of the spatial economy, beginning with Krugman (1991), Helpman (1998), and Fujita et al. (1999), highlight the roles of space, local spillovers, and

[2]Random growth is often referred to in the literature as Gibrat's law. The "law" is an application of the central limit theorem to the log of the product of independent shocks, originally formulated to describe the growth of firms (Gibrat 1931).

[3]Notable instances of recovery from shocks are documented in Davis and Weinstein (2002), Brakman et al. (2004) and Davis and Weinstein (2008) following bombings, and in Johnson et al. (2019) following pandemics. Desmet and Rappaport (2017) also document the absence of the random growth phenomenon for cities during the settlement of the American West.

[4]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.

the importance of trade between locations in determining population distributions. The spatial literature is now based on quantitative spatial models as developed in Allen and Arkolakis (2014), Redding (2016) and Redding and Rossi-Hansberg (2017). The literature on quantitative spatial models suffers from much the opposite problem than that afflicting the literature on the city size distribution. While quantitative spatial models incorporate many of the forces that influence the size of cities, they do not explain the regularity of the population distribution. These models can recover the unobserved qualities of each location, their "fundamentals," which rationalize observed populations. Yet this inversion is possible for any arbitrary vector of "populations," even those with little resemblance to real-world population distributions.[5] We demonstrate how both heterogeneity in the attributes of different places and trade across space will generate lognormal population distributions in equilibrium as a result of random variation in geography. Thus, we show how the models within this literature can be used to provide a deep explanation for why the population tends to be distributed in a particular way within countries. As our result relies solely on variation in geography and applications of a central limit theorem within these models, its simplicity and generality can explain the consistent observation of this phenomenon in many contexts.

Our framework also provides a strong theoretical link between observable characteristics of the world and market access, which a broad literature demonstrates are important for local populations, and the characteristic shape of the city size distribution. The largest cities around the world tend to be in locations that are good for production, offer quality-of-life benefits to residents, and offer the opportunity to trade with other locations. The literature has found a large role for "first-nature" geographic characteristics in explaining local populations, as in Rappaport and Sachs (2003), Nordhaus (2006), Nunn and Puga (2012), Henderson et al. (2018), Bosker and Buringh (2017), and Alix-Garcia and Sellars (2020). Literature supporting the importance of market access and trade for determining populations and incomes includes Redding and Venables (2004), Redding and Sturm (2008) and Head and Mayer (2011). Further, basing the origin of the population distribution on slow-changing geography can explain the persistence and resilience of the city size distribution to negative shocks.[6]

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 appear-

[5] Adão et al. (2023) make a similar point, noting that modern spatial models are saturated with free parameters such that they are always able to exactly match the underlying data.

[6] Davis and Weinstein (2002) propose geography as a likely determinant of the population distribution, given the recovery of the cities to their prior place in the population distribution following devastating bombings.

ance 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 regression given by:

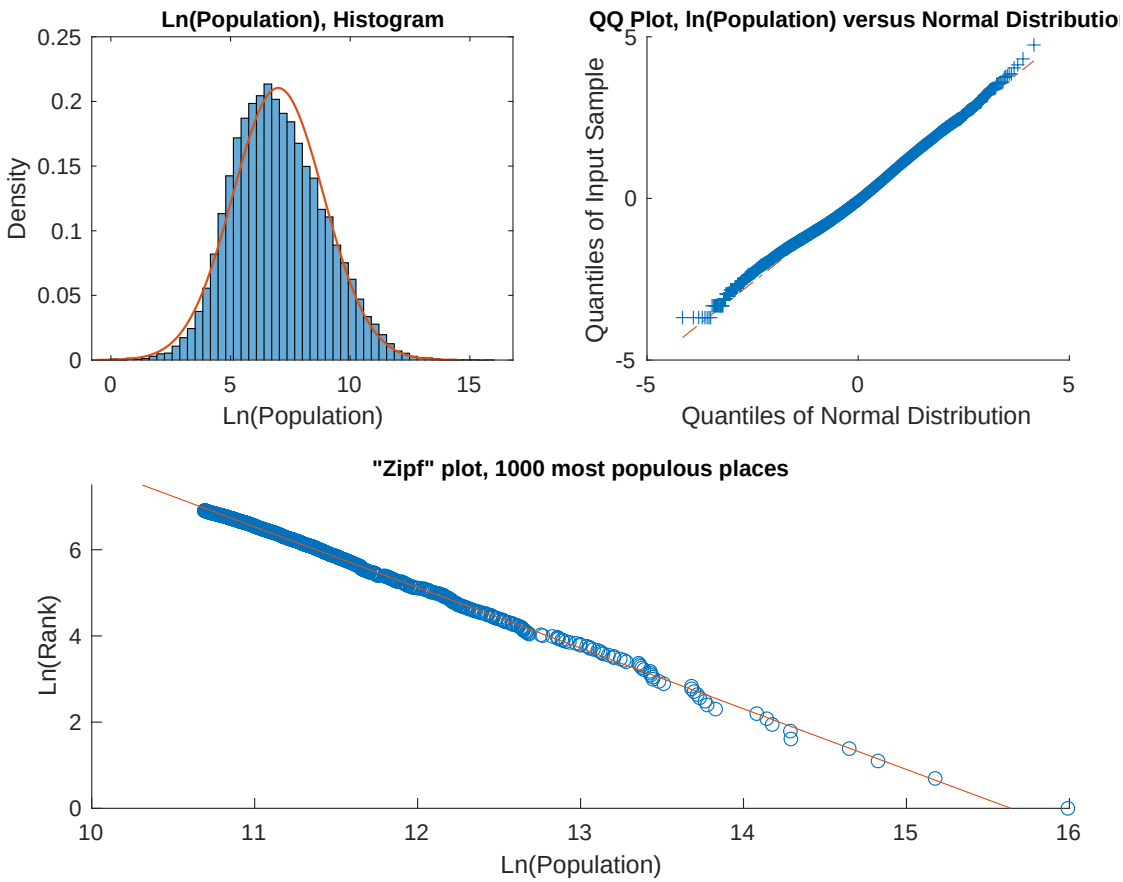
$$\ln(city\,rank_i) = \theta_0 + \theta_1 \ln(city\,pop_i) + \epsilon_i \tag{1}$$

for many countries delivers a high R^2 and frequently an estimate for θ_1 near -1. 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 nth-largest city, and the frequent appearance of this "law" motivated a substantial literature seeking to explain its origin. 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.[7]

Instead of being a true power law, the city size distribution may result from cities being a subset of a full population distribution which appears similar to a Pareto distribution for tail observations. Eeckhout (2004) demonstrates that the full population distribution for the U.S. appears lognormal. We construct an update to one of the key figures of Eeckhout (2004) in Figure I, 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.

[7]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).

Figure I:
Lognormal Populations Appear to Follow A Power Law in the Tail



Notes: The populations of U.S. Incorporated and Census Designated Places (N = 31,430) appears to follow a lognormal distribution. The right tail of this distribution (N=1000) appears to follow a power law. *Data Source:* U.S. Census

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 Supplemental 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 density function

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 density function in Equation 4, but with a scale-varying "shape parameter"-like term. As Malevergne et al. (2011) show, provided the σ parameter is not too small, 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.[8] With more complete 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 distribution, models generating a Pareto population distribution must rely on strong 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 deviations of the far right tail below the Pareto distribution 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 Supplemental 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 in expectation, roughly a quarter of the U.S. population.[9]

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

[8] This is true of early work, such as Auerbach (1913) and Zipf (1949). 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). 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).

[9] 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 Supplemental Appendix B.

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 fitted regression line.[10] Other features of real-world population distributions, such as the sensitivity of the estimated slope to the choice of truncation point, are consistent with the lognormal distribution as well.[11]

The appearance of a power law-like city population distribution is likely the result of focusing on the tail of a lognormal full population distribution. 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).

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 the canonical class of spatial models, and show that a realistic modeling of geography and trade will lead to lognormal population distributions in equilibrium.[12]

3.1 A Quantitative Spatial Equilibrium Model

The world consists of locations indexed by $i \in \mathcal{N}$, where $\mathcal{N} = \{i \mid i \in \mathbb{Z}^2, (-N, -N) \leq i \leq (N, N)\}$ where $N \in \mathbb{N}$ determines the extent of the plane. Trade between different locations is costly. We assume that locations have symmetric iceberg trade costs drawn from some distribution with finite support, such that $\tau_{i,i} = 1$, $\tau_{s,t} > 1$ for $s \neq t$, $\tau_{s,t} = \tau_{t,s}$,

[10] 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 Supplemental Appendix B.

[11] This property is discussed at length in Eeckhout (2004) and demonstrated in Supplemental 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.

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

and trade costs between any pair of locations $s, t \in \mathcal{N}$ are bounded above by a positive number.

The "distance" between locations is defined using the Euclidean norm such that $d_{i,j} = ||i - j||$. Distance is distinct from trade costs. Nearby location, in terms of their distance, will tend to be similar (in a sense to be made more precise in the following section) because of spatial correlation in geography, while trade costs can vary independently of distance to reflect varied geographies, trade routes linking certain locations (navigable rivers, railways, roads, etc.), or other idiosyncratic barriers to trade between 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:

$$\tilde{A}_i = A_i L_i^{\alpha} \tag{5}$$

$$\tilde{U}_i = U_i L_i^{\beta} \tag{6}$$

where the typical case will consist of $\alpha > 0$ and $\beta < 0$, reflecting positive productivity spillovers from agglomeration and the negative impact of overcrowding on amenities. 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 make the Armington assumption that each location produces a differentiated good. 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:

$$W_i = \left(\sum_{n \in \mathcal{N}} q_{n,i}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tilde{U}_i \tag{7}$$

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

Production is perfectly competitive.[13] A worker in location i can produce $\tilde{A}_i$ units of the

[13] As demonstrated in the appendix to Allen and Arkolakis (2014) the model nests cases of monopolistic

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 $w : N \to \mathbb{R}_{++}$.[14]

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

$$q_{i,n} = Q_n \left(\frac{p_{i,n}}{P_n} \right)^{-\sigma} \tag{8}$$

where Q_n is aggregate consumption in n, $Q_n = \frac{w_n L_n}{P_n}$, and P_n is the price index in location n, given by:

$$P_n = \left(\sum_{i \in \mathcal{N}} p_{i,n}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{9}$$

Given the assumption of perfect competition the price of the good produced in i consumed in n, can be expressed as:

$$p_{i,n} = \frac{\tau_{i,n} w_i}{\tilde{A}_i} \tag{10}$$

Combining the quantity (Equation 8) and price (Equation 10) expressions, we can write the value of the good produced in i consumed by n as:

$$X_{i,n} = \left(\frac{\tau_{i,n} w_i}{\tilde{A}_i P_n} \right)^{1-\sigma} w_n L_n \tag{11}$$

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

$$W_i = \frac{w_i}{P_i} \tilde{U}_i \tag{12}$$

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

$$w_i L_i = \sum_{n \in \mathcal{N}} X_{i,n} \tag{13}$$

The labor market clears:

$$\sum_{n \in \mathcal{N}} L_n = \bar{L} \tag{14}$$

We can then combine the welfare expression (Equation 12), value of consumption expres-

competition.

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

sion (Equation 11), and income expression (Equation 13) to get:

$$L_i w_i^{\sigma} = \sum_{n \in \mathcal{N}} W_n^{1-\sigma} \tau_{i,n}^{1-\sigma} \tilde{A}_i^{\sigma-1} \tilde{U}_n^{\sigma-1} L_n w_n^{\sigma} \tag{15}$$

The welfare expression (Equation 12) combined with the price index and and prices (Equations 9 and 10) yields:

$$w_i^{1-\sigma} = \sum_{n \in \mathcal{N}} W_i^{1-\sigma} \tau_{n,i}^{1-\sigma} \tilde{A}_n^{\sigma-1} \tilde{U}_i^{\sigma-1} w_n^{1-\sigma} \tag{16}$$

Given the form of the externalities in Equations 5 and 6, free movement between locations which ensures worker welfare is equal in all locations ($W_i = \bar{W}$ for all i), and symmetric trade costs we can combine Equations 15 and 16 into a single equation given by:

$$L_i^{\tilde{\sigma}\gamma_1} = \bar{W}^{1-\sigma} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in \mathcal{N}} \tau_{n,i}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \left(L_n^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \tag{17}$$

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 in our simulated results, which occurs when $\alpha + \beta \leq 0$ and is the empirically relevant case, but our result does not depend on this inequality. 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 17 in terms of the population in each location i in simplifying notation as:

$$L_i = F_i \times M_i^{\frac{1}{\tilde{\sigma}\gamma_1}} \tag{18}$$

where $F_i = \left(\bar{W}^{1-\sigma} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma}\right)^{\frac{1}{\tilde{\sigma}\gamma_1}}$, which we refer to as the "own-fundamental" term as it consists only of location i's fundamentals and the common positive constant $\bar{W}$, and $M_i = \sum_{n \in \mathcal{N}} m_{n,i}$ with $m_{n,i} = \tau_{n,i}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} L_n^{\tilde{\sigma}\gamma_2}$, which we refer to as the "market access" term of location i as it is a trade cost-weighted sum over all locations $n \in \mathcal{N}$. Each term $m_{n,i}$ of this summation thus represents the contribution of location n to the market access

of location i.

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

3.2 Variation within Place and the Distribution of Fundamentals

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." We reframe their approach as modeling fundamentals as arising from geographic variation. Further, using random variation in geography to model locational fundamentals is the cross-sectional analog of random growth models based 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.

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.

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.[15] We also assume each individual attribute is drawn from a common distribution in all locations, while different attributes may differ in their respective distribution.

We focus our discussion on productivity fundamentals, as we define the amenity fundamentals in the same way. 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

[15] 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.

exhibit complementarities between each of the attributes. That is, 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 each g:[16]

$$A_i = \prod_{g \in \mathcal{G}} {a_{ig}}^{\xi_g} \tag{19}$$

Taking the natural log yields the following expression:

$$\ln(A_i) = \sum_{g \in \mathcal{G}} \xi_g \ln a_{ig} \tag{20}$$

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 Supplemental Appendix C.4 using data on geographic attributes from Henderson et al. (2018).

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 $i \in \mathcal{N}$. This concept, also referred to as "strong mixing" or "weak dependence," was introduced to the spatial literature in Lee and Li (2013). A formal definition of α-mixing for sequences is given below. 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

[16] 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.

between elements of the sequence.

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 $n \to \infty$.

We assume that attributes satisfy these properties and are α-mixing within locations. Lemma 1 is a central limit theorem for α-mixing sequences due to Herrndorf (1984). It allows us to demonstrate convergence of a sum over a sequence of non-i.i.d random variables to a normal distribution, provided 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).

We define the ordering of the sequence of attributes $\{a_{i1}, a_{i2}, ..., a_{iG}\}$ such that similar attributes are close together (July temperatures and growing days have indices 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 1: *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* $G \to \infty$*,* $\frac{1}{\sqrt{G}\sigma\left(\widehat{\ln A_i}\right)} \widehat{\ln A_i}$ *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:

$$\ln(U_i) = \sum_{g \in \mathcal{G}} \iota_g \ln a_{ig} \tag{21}$$

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 of a lognormal distribution for fundamentals in two ways. First, the fundamentals recovered by inverting the model in Allen and Arkolakis (2014), plotted in Supplemental 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.[17] The plots are reported in Supplemental Appendix C.6, and show that aggregating the eleven attributes results in a distribution of fundamentals that appears lognormal. 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.

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

[17] 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 Supplemental Appendix C.5. A similar exercise was earlier done by Behrens and Robert-Nicoud (2015) for U.S. MSAs.

distributed.

***Proposition 2**: If A_i and F_i have a bivariate normal distribution in logs, F_i will be lognormally distributed.*

Proposition 2 follows immediately from the properties of lognormals, as either raising a lognormal distribution to a power or multiplying by a positive constant begets another lognormally distributed random variable and multiplying two lognormal random variables which have a bivariate normal distribution in logs also results in another lognormally distributed random variable.

3.3 Variation over Space and the Distribution of Market Access

We now demonstrate 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 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 $M_i = \sum_{n \in \mathcal{N}} M_{n,i}$, consists of a summation over the random variables $m_{n,i}$, each a function of the random variables $\tau_{n,i}$, A_n, U_n, and L_n, for all $n \in \mathcal{N}$.[18] Many central limit theorems apply only to independent and identically distributed random variables, but the sequences given by $\{m_{n,i}\}$ are unlikely to consist of independent random variables. The fundamentals of a location are intended to reflect its geographic advantages, and geographic attributes (topography, rainfall, soil quality, and so on) will tend to be similar for nearby locations. This means it is possible for productivity and amenity fundamentals A_i and U_i as defined in the prior section to be similar for nearby locations and exhibit spatial correlation.

However, as the distance between locations increases, the similarity of attributes across locations falls. In Supplemental Appendix C.4 we provide evidence for this pattern of spatial correlation based on the geographic attributes data in Henderson et al. (2018). Nearby locations often have broadly similar geographic attributes, like the climatic 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.[19] As such, we assume that while fundamentals may be similar for nearby locations they may also differ substantially over greater distances.

[18] In Supplemental Appendix A.4, we motivate the treatment of $m_{n,i}$ 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.

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

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

An additional property of the summations M_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 M_i as lognormal. Each $m_{n,i}$ must be strictly positive, as A_n, U_n, $\tau_{n,i}$, and L_n are all strictly positive and so $m_{n,i} = \tau_{n,i}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} L_n^{\tilde{\sigma}\gamma_2} > 0$. This allows us to apply a useful lemma from 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 Supplemental Appendix A.2.

Lemma 2 allows us to apply a central limit theorem to the sum M_i for all $i \in \mathcal{N}$, provided the sequences $\{m^d_{n,i}\}$ fulfill the requirements of the lemma. 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.[20]

[20] We discuss Lemma 2 further in Supplemental 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.

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 3.

Proposition 3: *For each $i \in \mathcal{N}$, define the sequences $\{m^d_{1,i}, m^d_{2,i}, ...m^d_{N,i}\}$ and the demeaned sequences $\{\hat{m}^d_{n,i}\}$ such that $\mathbb{E}[\hat{m}^d_{n,i}] = 0$ for all $n \in \mathcal{N}$. Define $M_i^{(N)} = \sum_{n=1}^N m^d_{n,i}$, $\mu_i^{(N)} = \sum_{n=1}^N \mathbb{E}[m^d_{n,i}]$, and $\sigma\left(M_i^{(N)}\right)^2 = Var[M_i^{(N)}]$. If, for all i,*

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

ii. *The coefficients of variation associated with the sequences $\{M_i^{(1)}, M_i^{(2)}, ..., M_i^{(N)}\}$ fulfill condition (ii) of Lemma 2 as $N \to \infty$*

then the distribution of $\frac{\mu_i^{(N)}}{\sqrt{N}\sigma\left(M_i^{(N)}\right)} \ln\left(\frac{M_i^{(N)}}{\mu_i^{(N)}}\right)$ converges in distribution to a normal distribution 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 M_i, and Lemma 2 allows us to move to a central limit theorem in logs as all elements of this sum are positive. If the necessary assumptions on $m_{n,i}$ hold, the market access summation M_i converges in distribution to a lognormal as N grows large. As a lognormal raised to a power results in another lognormal distribution each $(M_i)^{\frac{1}{\tilde{\sigma}\gamma_1}}$ will converge in distribution to a lognormal distribution as N grows large.

It is important to note that this property of market access is not specific to the Allen and Arkolakis (2014) model. Any setting where market access is modeled as consisting of a summation over many locations should have this property, as market access will necessarily be made up of positive contributions from the incomes and populations of many locations.

3.4 The Distribution of the Population

We have now shown that F_i and $(M_i)^{\frac{1}{\tilde{\sigma}\gamma_1}}$ 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 $L_i = F_i M_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$. This result is given in Theorem 1.

Theorem 1: *If F_i and $M_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ 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 F_i and $M_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$, and the property that products of lognormal distributions which have a bivariate normal distribution in logs also follow a lognormal distribution. Additional technical details regarding the ergodicity of the population distribution are worked out in Appendix A.3.

4 Results

We now verify that the modeling assumptions are reasonable by recovering fundamentals by inverting the model on US data. The recovered fundamentals are lognormally distributed, supporting the modeling assumptions we made for these fundamentals and resulting in lognormal own-fundamental terms. The empirical market access terms we construct based on the theory and these recovered fundamentals are also lognormally distributed, as predicted by the theory.

We then test further features of the model in simulation. We show that for empirically relevant parameter values the particular population distribution we recover well-approximates Zipf's law. We provide simulated 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 note an additional theoretical result relating to the prior Zipf's law literature. Given the properties of the model, 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 Empirics

We now test the validity of our modeling assumptions and its predictions using real-world population data by inverting this model to recover fundamentals and construct market access.

We use the 2017 Commodity Flow Survey to estimate trade costs, and the 2020 Decennial Census for county-level populations and incomes. Our inversion differs from that in Allen and Arkolakis (2014) in that we do not take L_i to be the population density of

the counties, but rather the level of the population. This is more consistent with equation 13, which states that total income in a location must equal its production. We follow the geographic trade cost estimation procedure as in Allen and Arkolakis (2014), using the observed shares by various modes to estimate a discrete choice model of trade. We take travel maps from the Census TIGER/Line collection and the US Department of Transportation for various travel modes, and estimate the distances between CFS regions based on the county-county distances drawing on insights from Head and Mayer (2009). We then estimate a PPML regression to recover the parameters of this discrete choice problem to fix geographic trade costs and also estimate non-geographic trade costs. We recover the exogenous locational fundamentals by setting $\alpha = 0.03$, in line with the estimates in Combes et al. (2019), Rosenthal and Strange (2004), and Combes and Gobillon (2015), and fix $\beta = -\frac{1}{3}$ based on household spending on housing and an isomorphism in the model.[21] More details on our estimation procedure and our recovered parameter values are provided in the appendix.

The distribution of county populations is shown in the left-hand column of Figure II, with both the density and QQ plot exhibiting a close fit to a lognormal distribution. The recovered own-fundamental term F_i, plotted in the middle column, is also strikingly lognormal. The recovered empirical market access M_i is also lognormally distributed. These results provide empirical support for our decision to model fundamentals as lognormal, and verify that our theoretical result regarding the distribution of market access holds in the data.

4.2 Simulation of the Population Distribution

We next simulate the model to investigate the resulting population distributions and the robustness of our asymptotic theoretical results in a finite setting. We simulate the results within a two dimensional geography. Each location in the model is a place that can host a settlement.[22] Note that the definition of a "location" is not specified by the model, beyond being a place within which local spillovers occur.[23] 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

[21] The estimate is consitent with those in Combes et al. (2008) and Davis and Ortalo-Magné (2011).

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

[23] 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 and Lee (2010), Rozenfeld et al. (2011), and Mori et al. (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.

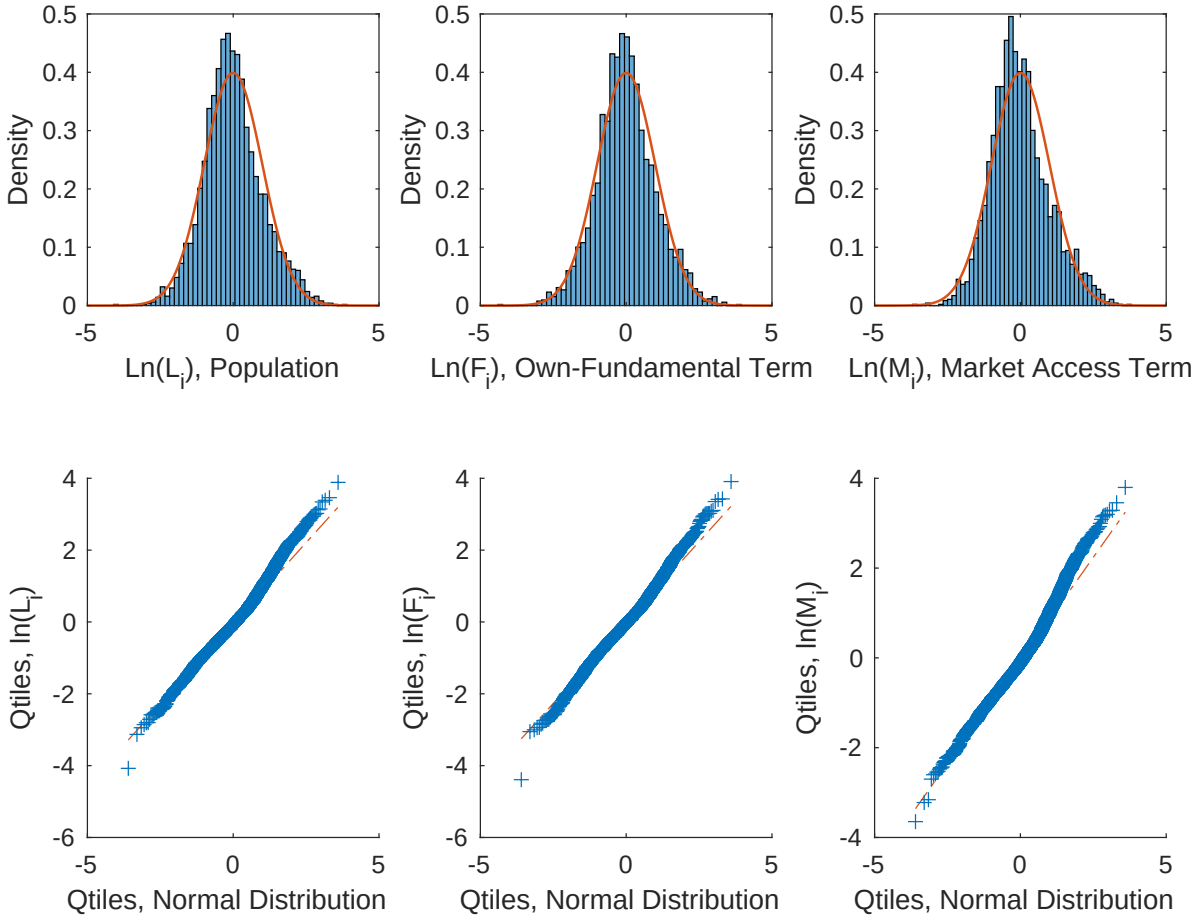


Figure II: The figure shows the distribution of population across US counties along with the distributions of the recovered own-fundamental terms F_i and the market access terms M_i. All of the distributions have been normalized for comparison against a standard normal distribution. The distribution of the recovered terms appear lognormal, as predicted by the theory.

distributions.

We take model parameters from the literature and from our empirical exercises where possible. The elasticity is set to $\sigma = 5$, and $\alpha = 0.03$ and $\beta = -\frac{1}{3}$. Our simulation is done on a dartboard geography. We draw exogenous fundamentals from lognormal distributions with $\mu_U = \mu_A = 0$ and $\sigma_A = \sigma_U = 1$. We allow these to have a correlation of 0.18 within locations, as found in our empirical inversion of US data, and we allow fundamentals to be spatial correlated and set the degree of spatial correlation at each distance $\rho^D = \exp(-\delta_{\rho_D} \cdot d_{ij})$. We set $\delta_{\rho_D} = 0.5$ reflect the decay of spatial correlation in attributes for US data, given our normalization of distance. We base trade costs on both a geographic component as well as bilateral and idiosyncratic shocks ϵ^B_{ij} and ϵ^I_i, setting $\tau^{sim}_{ij} = \exp(\delta^{TC} \cdot \epsilon^B_{ij} \cdot \epsilon^I_i \cdot d_{ij})$, where we draw the shocks from uniform distributions. We use the rate of increase of trade cost on roads from our regression on the US CFS data to parameterize the geographic portion of trade costs and set $\delta_{TC} = .001$. These parameters are listed in Table I.

Parameters		Fundamentals		Trade costs	
α	0.03	$\mu_U = \mu_A$	0	δ_{TC}	0.001
β	$-0.\bar{3}$	$\sigma_A = \sigma_U$	1	$\epsilon^B_{i,j}, \epsilon^I_i$	$\sim U(0.5, 1.5)$
σ	5	ρ_{AU}	0.12		
		δ_{ρ_D}	0.5		

Table I: Parameters used for simulation.

We demonstrate the robustness of the lognormal population distribution by performing 1000 simulations, each time drawing a new randomly generated distribution of fundamentals. Figure III displays smoothed results over 1000 simulations of the model.[24] 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, with an average coefficient across the 1000 simulations of -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 Supplemental Appendix B. Nonetheless, it is interesting to note that the estimated power law

[24]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.

Table II: Normality Tests

	Kolmogorov-Smirnov	Lilliefors	Jarque-Bera
Rejected at 1%	0.000	0.007	0.009
Rejected at 5%	0.001	0.047	0.041

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

Table III: Simulated changes in distribution

	$\frac{\partial \theta_1}{\partial \alpha}$	$\frac{\partial \theta_1}{\partial \beta}$	$\frac{\partial \theta_1}{\partial \delta_{TC}}$
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 Supplemental Appendix E.2)

exponent appears consistent with Zipf's Law for typical parameter values in the literature.

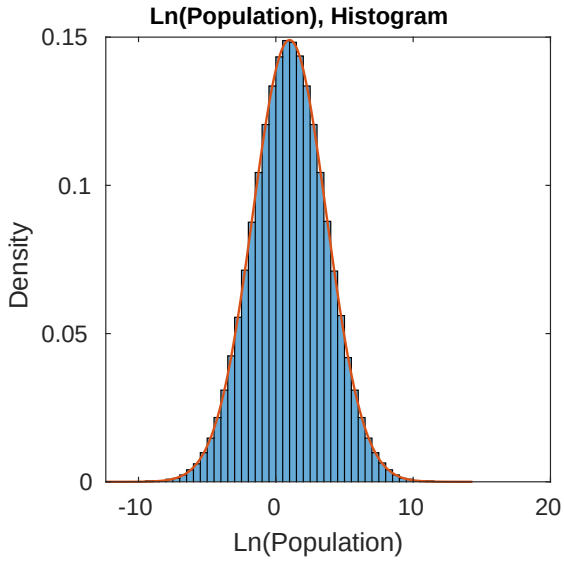
We test each of the 1000 simulated population distributions against the null hypothesis that the logged population distribution is normally distributed using the Kolmogorov-Smirnov, Lilliefors, and Jarque-Bera tests. The results of these tests are given in Table II. None of the tests reliably reject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

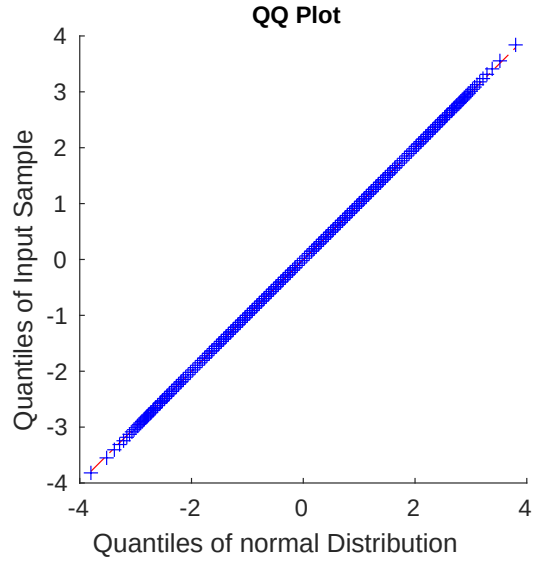
We next simulate the comparative statics of the estimated power law coefficient for the city size distribution, testing its sensitivity to changes in model parameters. Changing parameters alters the estimated coefficient of the log rank-size regression, which we denote θ_1 as in Equation 1. We perform 100 simulations for each of 150 combination of parameters.[25] A summary of the signs of changes (estimated by a regression given in Supplemental Appendix E.2) is provided in Table III.

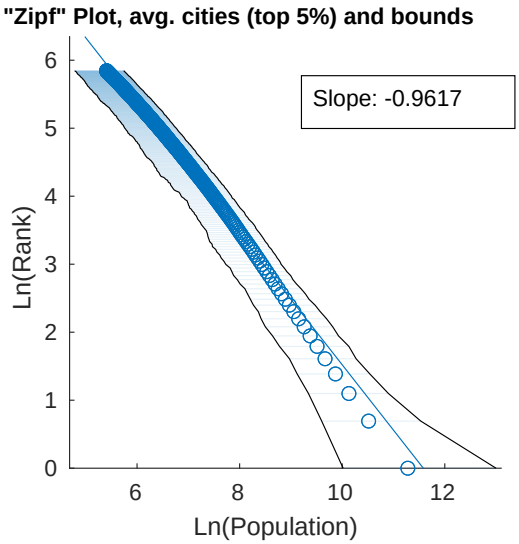
The comparative statics of our model demonstrate changes in the estimated power law coefficient in line with empirical evidence. Increasing the benefits of agglomeration by raising $\alpha > 0$ results in a more unequal city size distribution (greater dispersion, or a

[25]We simulate for 5 values of α, 6 values of β, and 5 values of δ_{TC} given in Supplemental Appendix E.2. 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.

Figure III:
Smoothed Output Over 1000 MC Simulations







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 the 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 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.

flatter slope) and increasing local congestion costs by reducing $\beta < 0$ results in a more equal distribution (less dispersion, or a steeper slope). Increasing trade costs by increasing the rate at which these costs grow with distance, $\delta_{TC} > 0$, likewise results in a more unequal city size distribution. In many developing countries, the city size distributions are highly unequal, generating the flatter slopes in the log-log regression documented in Duben and Krause (2021). These unequal city size distributions 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 very large metropolises relative to secondary cities ("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 write Equation 29, the matrix form representation of Equation 17 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 Supplemental 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. As a result, 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 the underlying characteristics of place and trade. This view 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, 2008).

5 Conclusion

The power law-like distribution of city populations is a striking empirical regularity that holds across countries and over 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 the qualities of a location and its market access for its population 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

Adão, R., A. Costinot, and D. Donaldson (2023). Putting Quantitative Models to the Test: An Application to Trump's Trade War. Technical report, National Bureau of Economic Research.

Alix-Garcia, J. and E. A. Sellars (2020). Locational fundamentals, trade, and the changing urban landscape of mexico. *Journal of Urban Economics 116*, 103213.

Allen, T. and C. Arkolakis (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics 129*(3), 1085–1140.

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

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

Auerbach, F. and A. Ciccone (2023). The Law of Population Concentration. *Envmt. and Planning B: Urban Analytics and City Sc. 50*(2), 290–298.

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

Blank, A. and S. 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), 279–288.

Bosker, M. and E. Buringh (2017). City seeds: Geography and the origins of the european city system. *Journal of Urban Economics 98*, 139–157.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys 2*(N.A.).

Brakman, S., H. Garretsen, and M. Schramm (2004). The strategic bombing of German cities during World War II and its impact on city growth. *Journal of Economic Geography 4*(2), 201–218.

Brakman, S., H. Garretsen, C. Van Marrewijk, and M. Van Den Berg (1999). The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution. *Journal of Regional Science 39*(1), 183–213.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Combes, P.-P., G. Duranton, and L. Gobillon (2019). The Costs of Agglomeration: House and Land Prices in French Cities . *Review of Economic Studies*, 1556–1589.

Combes, P.-P. and L. Gobillon (2015). Chapter 5 - The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 247–348. Elsevier.

Cuberes, D., K. Desmet, and J. Rappaport (2021). Urban growth shadows. *Journal of Urban Economics 123*, 103334.

Córdoba, J.-C. (2008). On the distribution of city sizes. *Journal of Urban Economics 63*(1), 177–197.

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

Davis, D. R. and D. E. Weinstein (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science 48*(1), 29–65.

Davis, M. A. and F. Ortalo-Magné (2011). Household expenditures, wages, rents. *Review of Economic Dynamics 14*(2), 248–261.

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

Duben, C. and M. Krause (2021). Population, light, and the size distribution of cities. *Journal of Regional Science*, 189–211.

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

Eeckhout, J. (2009). Gibrat's Law for (All) Cities: Reply. *American Economic Review 99*(4), 1676–83.

Frigg, R., J. Berkovitz, and F. Kronz (2020). The Ergodic Hierarchy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University.

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

Gabaix, X. (1999a). Zipf's Law and the Growth of Cities. *The American Economic Review 89*(2), 129–132.

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

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics 1*(1), 255–294.

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

Gibrat, R. (1931). *Les inégalités économiques.* Paris: Librairie du Recueil Sirey.

Head, K. and T. Mayer (2009). Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade. In *The Gravity Model in International Trade: Advances and Applications*. Cambridge University Press.

Head, K. and T. Mayer (2011). Gravity, market potential and economic development. *Journal of Economic Geography 11*.

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

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The Global Distribution of Economic Activity: Nature, History, and the Role of Trade. *The Quarterly Journal of Economics 133*(1), 357–406.

Herrndorf, N. (1984). A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability 12*(1), 141 – 153.

Holmes, T. J. and S. Lee (2010). Cities as Six-by-six-mile Squares: Zipf's Law? In *Agglomeration Economics*. NBER.

Hsu, W.-T. (2012). Central place theory and city size distribution. *The Economic Journal 122*(563), 903–932.

Johnson, N., R. Jedwab, and M. Koyama (2019). Pandemics, Places, and Populations: Evidence from the Black Death. *Working Paper*.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–499.

Krugman, P. (1996). *The Self-Organizing Economy*. Blackwell.

Lee, S. and Q. Li (2013). Uneven landscapes and city size distributions. *Journal of Urban*

Economics 78, 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. (1967). A Normal Limit Theorem for Power Sums of Independent Random Variables. *Bell System Technical Journal 46*(9), 2081–2089.

Mori, T., T. E. Smith, and W.-T. Hsu (2020). Common power laws for cities and spatial fractal structures. *Proceedings of the National Academy of Sciences 117*(12), 6469–6475.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the Natl Academy of Sciences 103*(10), 3510–3517.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics 94*(1), 20–36.

Rante, R., F. Trionfetti, and P. Verma (2024). The Size Distribution of Cities: Evidence from the Lab. working paper or preprint.

Rappaport, J. and J. D. Sachs (2003). The United States as a Coastal Nation. *Journal of Economic Growth 8*(1), 5–46.

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

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

Redding, S. J. and E. Rossi-Hansberg (2017). Quantitative Spatial Economics. *Annual Review of Economics*.

Redding, S. J. and D. M. Sturm (2008, December). The costs of remoteness: Evidence from german division and reunification. *American Economic Review 98*(5), 1766–97.

Rosenthal, S. S. and W. C. Strange (2004). Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson and J.-F. Thisse (Eds.), *Cities and Geography*, Volume 4 of *Handbook of Regional and Urban Economics*, pp. 2119–2171. Elsevier.

Rossi-Hansberg, E. and M. Wright (2007). Urban Structure and Growth. *The Review of Economic Studies*, 597–624.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011). The Area and Population of Cities: New Insights from a Different Perspective on Cities. *American Economic Review 101*(5), 2205–25.

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

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort. *Addison-Wesley Press*.

SUPPLEMENTAL APPENDIX

for "Populations in Spatial Equilibrium"

Matthew Easton and Patrick W. Farrell

This appendix includes proofs and additional material.

1. Discussion of Main Text Results and Proofs.
2. Pareto vs. Lognormal Simulations
3. Attributes and Fundamentals
4. Model Inversion
5. Additional Details on Simulation

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 et al. (2011), and is similar to that in Eeckhout (2009).

The lognormal distribution, while fitting the body of the population distribution well, provides somewhat worse fit compared to similar distributions with more parameters, such as the double Pareto lognormal. We focus on describing the full population distribution as lognormal for relative simplicity of analysis. More robust discussions of goodness-of-fit tests and alternative statistical descriptions of population distributions can be found in Giesen et al. (2010) and Ioannides and Skouras (2013).

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 et al. (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

$$C_v = \frac{\sqrt{ns^2}}{nm} = \frac{\sqrt{n}s}{nm} \tag{22}$$

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

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

$$\sigma_X^2 = \ln(1 + C_v^2) \tag{23}$$

$$\mu_X = \ln(nm) - \frac{\sigma_X^2}{2} \tag{24}$$

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

$$\sigma_X^2 \to C_v^2 \quad \text{, so} \quad \sigma_X \to C_v \quad \text{, and } \sigma_X \to 0 \tag{25}$$

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 $\frac{x}{M} \rightarrow_{a.s.} 1$. We can write $\frac{x}{M} \cdot \sigma_X \rightarrow \frac{\sqrt{n}s}{nm}$, and so $x\sigma_X \rightarrow \sqrt{n}s = \sigma_N$. This means $\frac{1}{x\sigma_X\sqrt{2\pi}} \rightarrow \frac{1}{\sigma_N\sqrt{2\pi}}$.

Similarly, $x = \frac{xM}{M}$, so $\ln(x) = \ln(M) + \ln(\frac{x}{M})$. As $\frac{x}{M} \rightarrow 1$, so $\ln(\frac{x}{M}) \rightarrow \frac{x-M}{M}$. As $\mu_X = \ln(M) - \frac{\sigma_X^2}{2}$, and $\sigma_X \rightarrow C_v \rightarrow 0$, and $M = \mu_N$ then we have

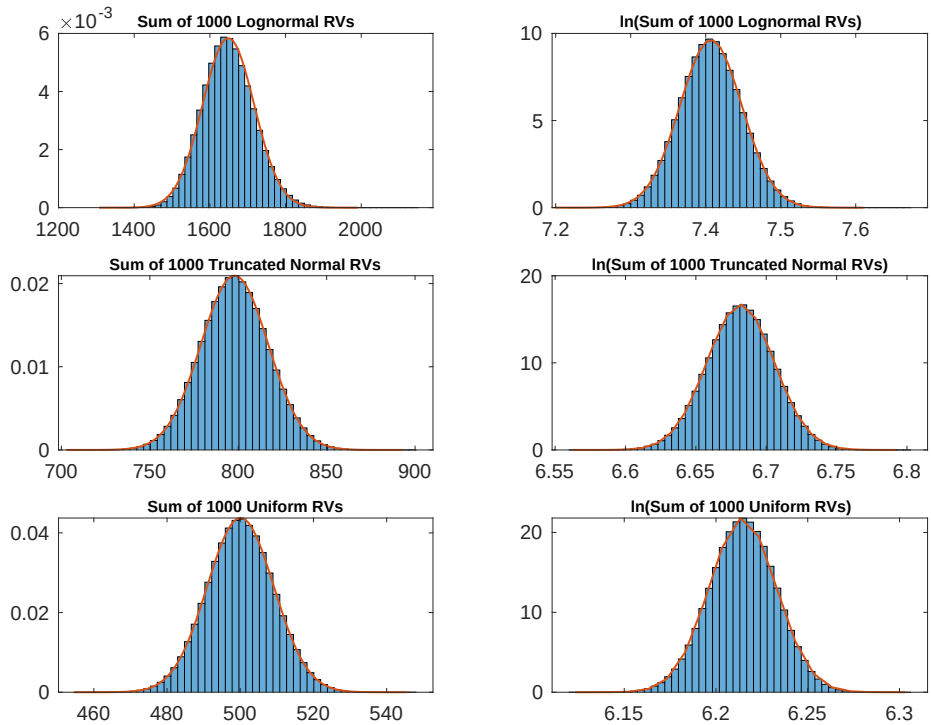
$$\frac{\ln(x) - \mu_x}{\sigma_X} \rightarrow \frac{\ln(M) + (\frac{x-M}{M}) - \ln(M)}{\sigma_X} = \frac{x-M}{M\sigma_X} \rightarrow \frac{x-M}{M \cdot C_v} = \frac{x-\mu_X}{\sigma_N} \tag{26}$$

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

$$f(x) = \frac{1}{x\sigma_X\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu_X}{\sigma_X}\right)^2} \rightarrow \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_N}{\sigma_N}\right)^2} \tag{27}$$

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

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



Notes: 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.3 Additional Proofs

A remaining point is the $\alpha-$mixing of the population distribution itself. We want to ensure that the distribution is such that the full distribution is realized— we further require that populations are not too correlated across locations. While we have proven that the distribution of the population in each i approaches a lognormal, we did not formally rule out the case where the population is perfectly correlated such that the PDF of the realized population distribution appears degenerate.

The mixing definition applied in the main text paplies to sequences. To consider the mixing of the population, we now require a shift from considering the mixing properties of points or summations defined over sequences for individual locations to considering the mixing properties of the random variables across the full plane. Population is defined on a plane and is a random field, a generalization of random sequences to a multidimensional space. A random field is a collection of X-valued random variables indexed by elements in a topological space T.

The following definition of α-mixing for random fields is used when considering α-mixing of the population across the two-dimensional geography.

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:*

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

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

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

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:

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

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.

Under certain conditions we can show that the population will be α-mixing, which will ensure that populations will be asymptotically independent across space. The restrictions necessary to establish α-mixing are strict, and α-mixing is not itself a necessary condition for the ergodicity of the population distribution as α-mixing is a stronger concept which implies ergodicity (Frigg et al. 2020).

To ensure this, we introduce a Lemma that generalizes α-mixing results for sequences.

Lemma 3: *Suppose that for each $n = 1, 2, 3, ..., X^{(n)} := \left(X_i^{(n)}, i \in \mathbb{Z}^2\right)$ 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 $i \in \mathbb{Z}^2, h_i : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \ldots \to \mathbb{R}$ is a Borel function. Define the field $X := (X_i, i \in \mathbb{Z}^2)$ of random variables by $X_i := h_i\left(X_i^{(1)}, X_i^{(2)}, X_i^{(3)}, \ldots\right), i \in \mathbb{Z}^2$. Then for each $k \geq 1$, $\alpha_X(k, \infty, \infty) \leq \sum_{n=1}^{\infty} \alpha_{X^{(n)}}(k, \infty, \infty)$, where α_Z indicates the mixing coefficient for field Z.*

The proof of Lemma 3 is stated below.

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 $\alpha(k; \infty, \infty)$.

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

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

And so,

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

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 establishes the result. ■

The proof extends the result for α-mixing sequences from Bradley (2005) Theorem 5.2 to α-mixing fields in two dimensions.

Lemma 3 states that measurable transformations and combinations of α-mixing fields are also α-mixing. If our geography was one-dimensional, such as a circle or line, the equivalent statement is given by Theorem 5.2 of Bradley (2005) which establishes that α-mixing is maintained over measurable transformations and over combinations of independent α-mixing random sequences. As our geography is defined in two-dimensions (to reflect a realistic geography), our Lemma 3 extends Theorem 5.2 of Bradley (2005) to two dimensions.

We assume that the sequences $\{A_i\}$ and $\{U_i\}$ are α-mixing with respect to distance to reflect the patterns of correlation in geographic attributes across space.[26] This assumption captures the potential similarity of nearby locations and how this similarity vanishes with increasing distance, and is consistent with the argument for α-mixing of the elements of the market access summation in the prior section.[27]

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

The proof follows directly from Lemma 3, as α-mixing is maintained over measurable transformations and over combinations of independent α-mixing random sequences. 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.[28]

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

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

The proof also follows directly from Lemma 3. The α-mixing of the population within the model is consistent with real-world population distributions, as large population cen-

[26] If geographic attributes were independent sequences the α-mixing of $\{A_i\}$ and $\{U_i\}$ would follow directly from Theorem 5.2 of Bradley (2005) and α-mixing of attributes with respect to distance. However, as many attributes are likely not independent, we must assume α-mixing of the fundamentals with respect to distance

[27] We find that the correlation of the constructed "naive" fundamentals across space declines with increasing distance, as shown in in Supplemental Appendix C.6, consistent with the assumption of α-mixing.

[28] This assumption is necessary for the application of Theorem 5.2 of Bradley (2005) in the proof of Proposition 4, 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 and our recovered fundamentals (discussed in the following section) have a similar correlation of 0.18.

ters tend to be surrounded by other populous areas rather than sparsely populated regions.[29] The α-mixing of the population also implies that the very largest cities will not cluster together despite the correlations in population, because the population of different locations will approach independence as distance between them increases.

[29]There is a literature on urban shadows which suggests that new cities tend not to form immediately next to existing cities, and which appears to be in tension with the correlation of the population distribution across space. One example of this literature is Bosker and Buringh (2017), which documents a "shadow" surrounding cities in Europe from 800-1800. This shadow is ascribed to forces beyond our model, such as the risk of armed conflict between cities, and results within Bosker and Buringh (2017) still demonstrate a high degree of spatial correlation in the population distribution (see, for example, their Figure 4). Related work on urban *growth* shadows, as in Cuberes et al. (2021) find that the growth of peripheral regions near large urban centers is influenced by the central city, with locations near urban centers tending to grow faster over the past century, which is also consistent with spatial correlation in populations.

A.4 Population as a random variable

The system of equations describing the population distribution, given for a particular location i in Equation 17 can be expressed in matrix form as:

$$\theta \mathbf{h} = \mathbf{J}[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}} \tag{28}$$

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.[30]

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

$$\mathbf{s}_i = \theta \mathbf{k}_i \mathbf{J}^{-1} \mathbf{h} \tag{29}$$

where $\mathbf{k}_i$ is a vector of elements $k_{n,i} = \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(1-\sigma)}$.

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 $s_{n,i}$ as random variables.

[30]For a review of random matrices, see Anderson et al. (2010) and for results on eigenvectors see Ben Arous and Guionnet (2010) and O'Rourke et al. (2016).

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.

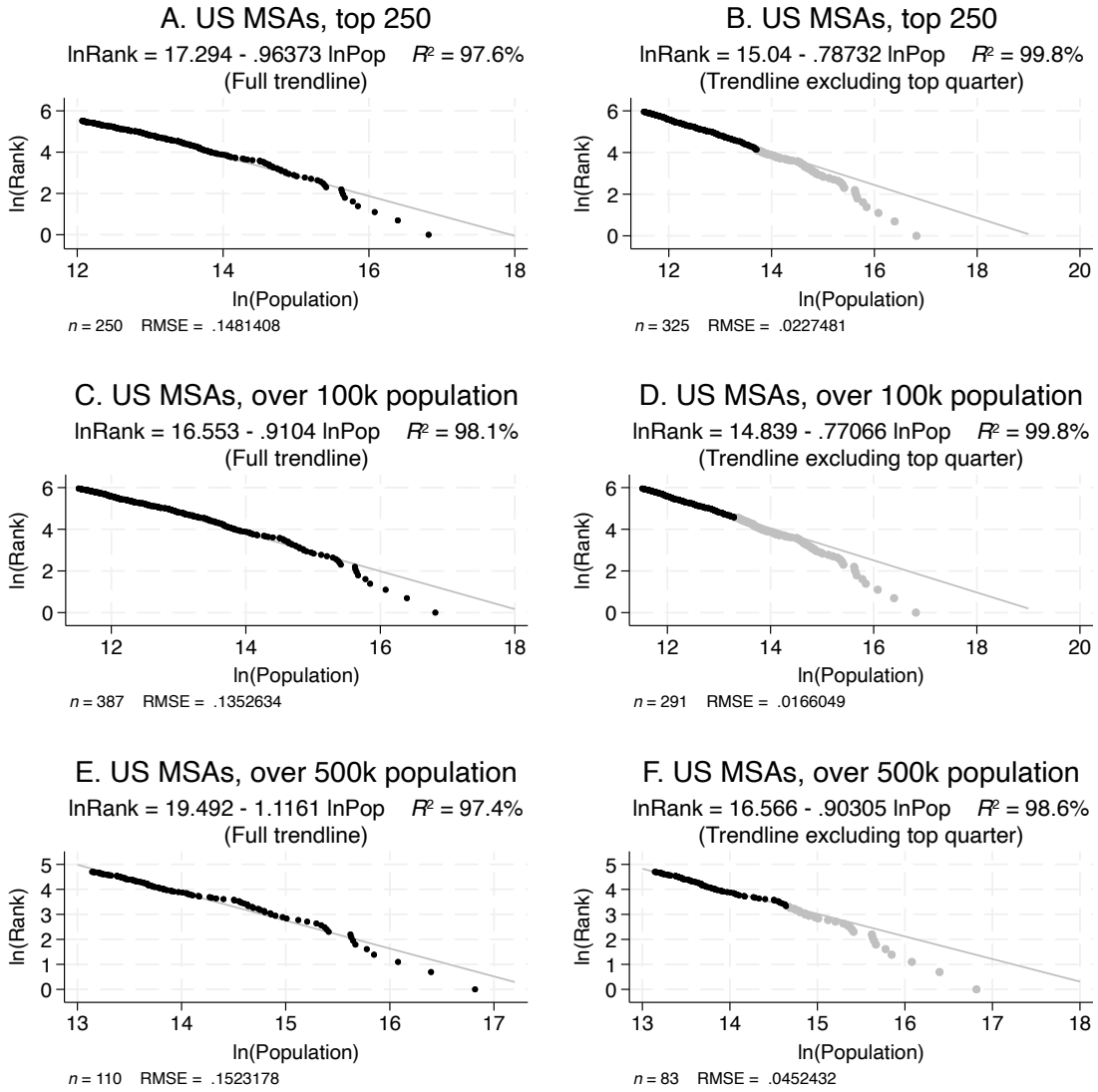
In the right column of Figure A.II, we consider deviations in the far tail by excluding the top quarter of cities within each subset of the city distribution. In all cases, nearly all top quarter cities fall below the trendline predicted based on the rest of the distribution.[31] The magnitude of the 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 entire U.S. population, while panel F indicates an absence of 169 million people, roughly half the U.S. population.[32]

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

[31] 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

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

Figure A.II:
Truncation Points and Power Law Coefficients

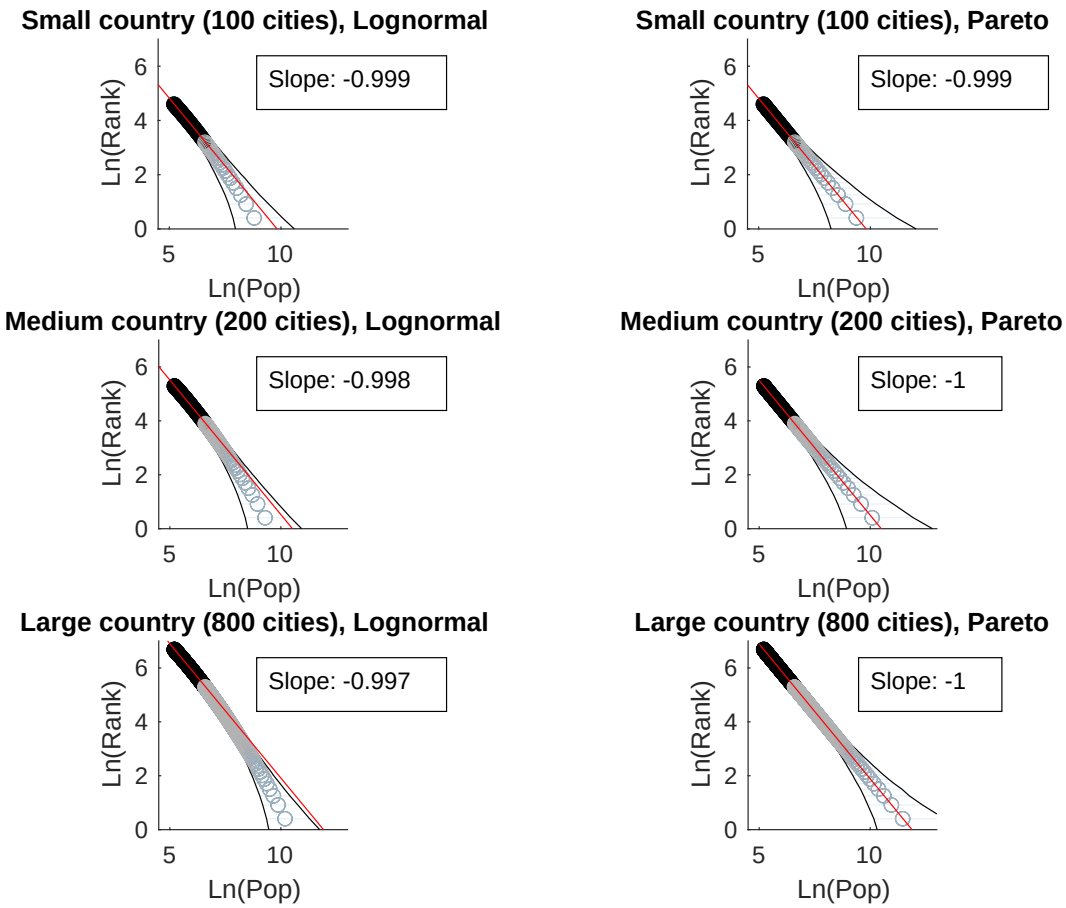


Notes: The top panels (A and B) show the 388 U.S. MSAs with a population over 100k in 2020. The bottom panels (C and D) show the 110 MSAs with a population over 500k in 2020. Panels A and C display the trendline for the full distribution and Panels B and D display the trendline excluding the top 25% 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

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.[33] 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.

[33]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.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 Attributes and Fundamentals

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 based on replication files of 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.[34]

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.[35] 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.

[34] 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).

[35] 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.[36] The dataset is at the quarter-degree latitude and longitude cell level.[37] We focus on the roughly 47,000 cells grid cells in the U.S., Mexico, and Canada, with the over 13,000 of those grid cells contained in the contiguous U.S. serving as our main sample.

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.[38] 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, 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.[39] 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.

[36] 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.

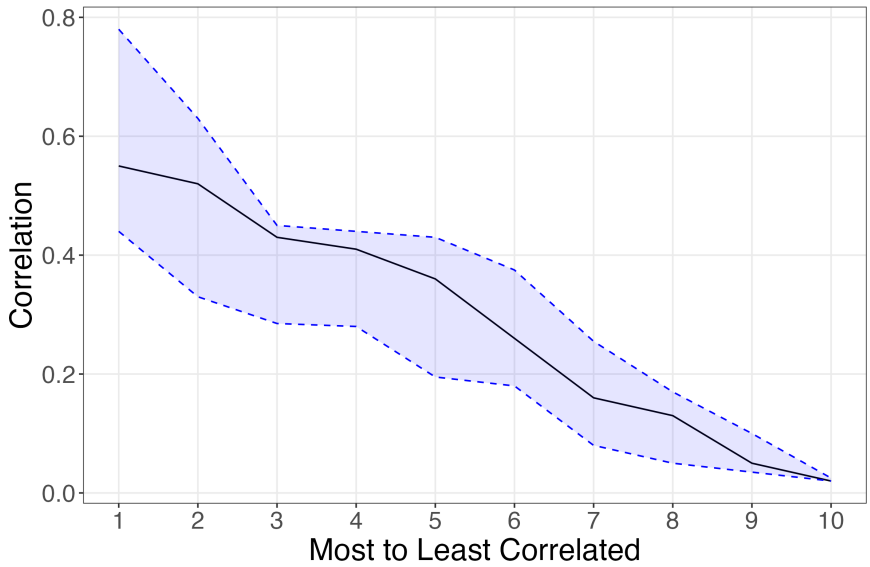
[37] At the equator, a grid cell is ≈28-by-28 km; at 48 degrees latitude, ≈18-by-18 km.

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

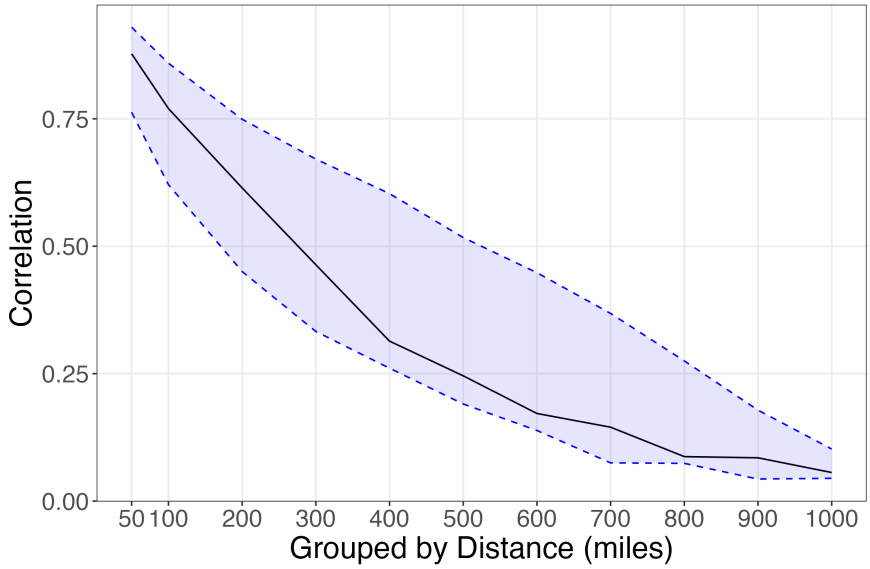
[39] A description of how we calculated spatial correlation is provided in Supplemental Appendix C.3.

Figure A.IV:
Correlations Within and Across Locations

(a)



(b)



Notes: (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 based on replication files of Henderson et al. (2018)

C.5 A "Naive" 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:

$$\hat{a}_{ig} = \frac{\ln(a_{ig}) - \text{mean}\,(\ln(a_g))}{\text{sd}(\ln(a_g))}$$

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.

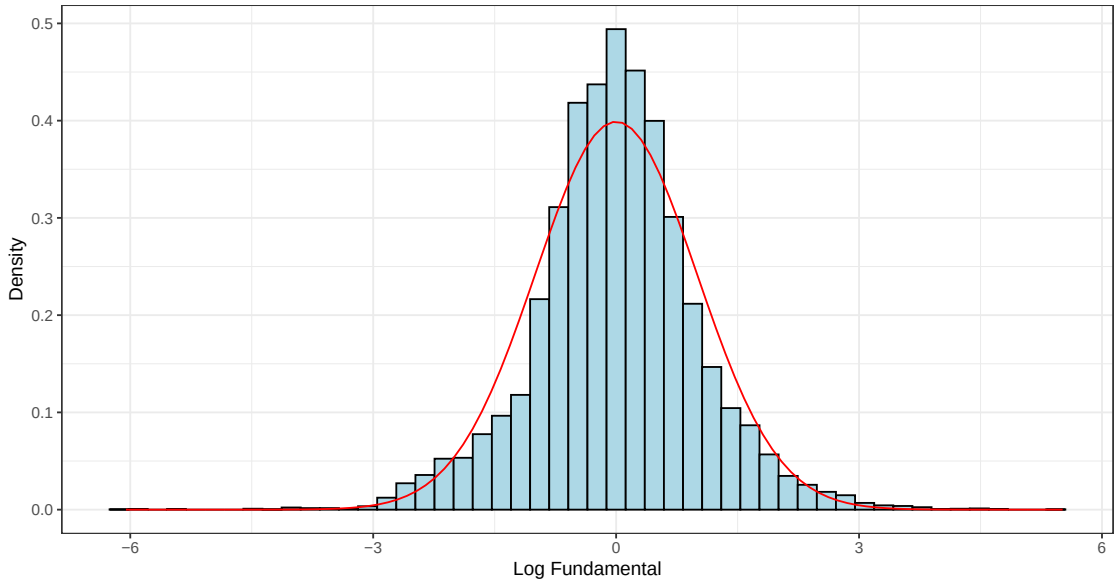
We then aggregate our attributes into a fundamental given by Equation 20:

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

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

C.6 Results: "Naive" Fundamental and Correlations

Figure A.V:
Lognormal Distribution of Locational Fundamentals for Contiguous U.S.



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 attributes are at the grid-cell level ($N = 13,426$). The mean and variance are standardized to zero and one and a standard normal curve is overlaid.
Data Source: Authors' calculations based on replication files of Henderson et al. (2018)

We plot in Figure A.V the empirical PDF of the resulting distribution of productivity fundamentals, calculated according to Equation 20 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 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).

Table A.II: Spatial Correlation of Calculated Fundamental for the Contiguous U.S.

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 replication files of Henderson et al. (2018).

D Model Inversion

This appendix discusses our empirical portion, which recovers fundamentals using data on wages and populations for US counties and estimated trade costs. Our model inversion largely follows Allen and Arkolakis (2014), with slight modifications to suit the data we use, our interpretation of the model, and our preferred parameter values.

We draw on three data sources for data on county-county. We use the decennial census for information on the total population in the county and on the demographic breakdown. We use the American Community Survey from 2020 for average county incomes. We use the Facebook Social Connectness Index for a measure of the degree of social connection between pairs of counties.

For estimating trade costs we draw on the 2017 Commodity Flow Survey. We follow Allen and Arkolakis (2014) in using a discrete choice estimation of the relative trade costs for various modes based on mode-specific shares. The three modes we consider are road, rail, and water. We exclude air, which was included in the original Allen and Arkolakis (2014) inversion, as very few CFS pairs record trade via this mode and the distances recorded in the 2017 CFS make it clear that most of this trade is not direct, such that it is difficult to calculate the distances on a travel map without knowledge of the hub-and-spoke system of air trade in the US.

We use three travel maps for our three modes. We draw on the TIGER/Line primary roads map for 2019, from which we extract the Interstate Highway System. We then complement this with the TIGER/Line primary-secondary road network for a complete road network. We use the TIGER/Line railway map for 2019 for rails. We use a map of water shipping routes from the US DOT for water routes. We convert each of these maps into an equal distance raster with cells 5km across, and assign each cell in the raster a travel speed. For roads, we assign the highest speed (70 mph) to interstate highways, a medium speed to other roads (35 mph), a low speed to travel off-road (10 mph), and a negligible speed to travel off land (0.1 mph). For trains we assign a high speed (50 mph) to travel on rail and the same land and off land speeds as for road travel. For water, we assign a high speed (20 mph) to travel on a designated water shipping route, a lower speed to travel on water outside a shipping route (10mph), and a low speed to travel on land (1 mph). We use the Matlab image processing tool to obtain the distance between all counties in the US by each of these three modes.

To construct the distance between CFS regions, we follow Head and Mayer (2009) and construct the CFS distances as a CES combination of the population-weighted distances between all of the constituent counties of each CFS region. This provides a better sense

of both the distances between two different CFS regions as well as an internal distance for each CFS region. Given our normalization of distances, we do not follow Head and Mayer (2009) in assign each county a distance to itself based on its land area but simply set this to the shortest such distance by each mode observed in the data.

Table A.III: PPML Regression Output, Trade Value

	(1)	(2)	(3)	(4)
Trade Costs	1.156***	0.585***	0.852***	0.484***
	(0.0478)	(0.0282)	(0.0286)	(0.0220)
Same state		2.157***		1.899***
		(0.0651)		(0.0592)
SCI connection likelihood			192.3***	114.9***
			(14.98)	(10.50)
Income difference			-2.543***	-2.267***
			(0.185)	(0.160)
Race difference			-2.562***	-1.097***
			(0.179)	(0.175)
Constant	9.850***	8.292***	10.27***	8.770***
	(0.0872)	(0.0665)	(0.0741)	(0.0738)
Origin-Destination FEs	Yes	Yes	Yes	Yes
Observations	16641	16641	16641	16641

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include origin and destination FEs.

θ	v_{rd}	v_{rl}	v_{wa}	f_{rl}	f_{wa}
8.2684	1.0148	0.0001	0.3338	0.9714	0.9675

Table A.IV: The table shows the estimated parameter θ from the discrete choice estimation as in Allen and Arkolakis (2014). The scaled trade costs recovered from the PPML regression given θ are displayed in columns 2-6.

We estimate the relative trave costs as in Allen and Arkolakis (2014) by finding the variable (for road, rail, and water) and fixed (for rail and water) costs that best rationalize the observed mode shares. We then use these costs to estimate a PPML regression to recover the parameter θ that allows us to calculate average geographic trade costs once we set the elasticity $\sigma = 5$. We also estimate four "non-geographic" trade costs using this regression. These are 1) the absolute value of the difference in log income between the CFS regions, 2) the difference in the demographic norm between the two counties, based on the six racial categories in the US census, 3) the Facebook Social Connectedness Index between the two counties, which reflects their degree of existing connection in social networks, and 4) an indicator for if the CFS regions are in the same state. The PPML regression coefficients are given in Table A.III.

We use the results in column (5) to recover the values θ and the fixed and variable trade costs, which gives us the parameter values in table XX. We use these trade costs and the distances between all counties to construct a matrix of county-county trade costs. The values generally half the size of those in the Allen and Arkolakis (2014) inversion, reflecting the different value of σ we adopt. The largest difference is for rail, which has a much smaller variable cost than estimated in the original Allen and Arkolakis (2014) inversion.

Given these trade costs, we invert the model and recover the exogenous A_i and U_i for each location, using $\alpha = 0.03$ and $\beta = -\frac{1}{3}$. The distributions of the resulting fundamentals, own-fundamentals F_i term, and market access M_i term are given in the main text.

E Additional Details on Simulation

This section gives more details on the simulation. We first discuss the main simulation and then discuss the version of the simulation where we vary parameter values.

E.1 Details on main simulation

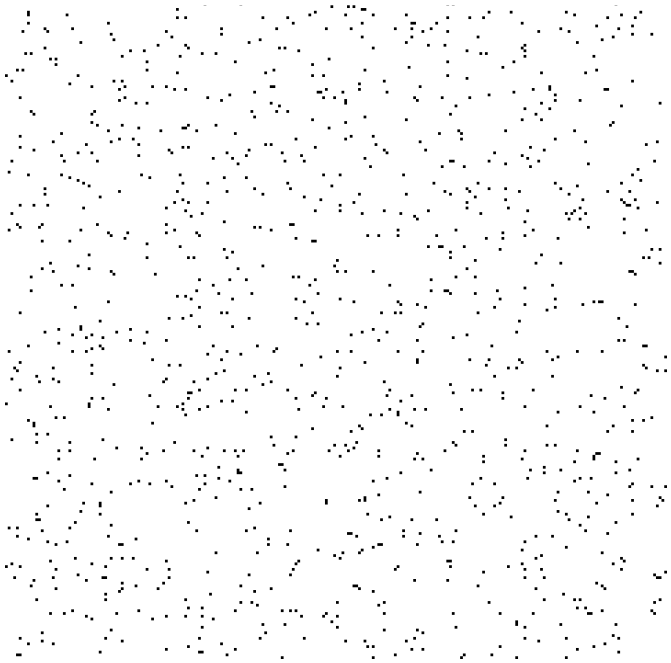
To ensure that geographic trade costs respect the triangle inequality, we model settlements as occurring randomly over a large surface and take the Euclidean distance between all settlements. We take draws from uniform distributions with parameters $[1, a]$ where a reflects the maximal horizontal and 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.VI 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.

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 uniformly distribute 10,000 settlements across a 1200-by-1200 grid and discard those within 100 cells of a border. This leaves an expected number of settlements of $\frac{100}{144} * 30000 = 7,944.\bar{3}$. We draw new randomly drawn fundamentals each simulation. We are left with a central geography consisting of approximately 7,000 settlements. We fix the distance across this central geography to equal 1.

Fundamentals are drawn from lognormal distributions with parameters as given in Table I. We induce spatial correlation in the fundamentals using a Choleski decomposition. We assume the degree of spatial correlation of the log-scale fundamental declines exponentially, consistent with the empirical attribute correlations we show in Supplemental 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 $\delta_\rho = 50$. 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 $\rho_{AU} = 0.18$ to match the correlation between the recovered productivity and amenity fundamentals in our empirical inversion.

The magnitude of local productivity spillovers is given by $\alpha = 0.03$, in line with the estimates in Combes et al. (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

Figure A.VI:
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.

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 et al. (2019) and Davis and Ortalo-Magné (2011).

E.2 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 do not include idiosyncratic shocks to trade costs in this section, and use only geographic trade costs such that $\tau_{ij} = \exp(\delta^{TC} \cdot d_{ij})$, so that an unusual realization of these costs does not influence the estimate. We set the number of locations in the full geography to 10,000.

We here report the full table of estimated parameter values for three values of δ_{TC}, at

0.5, 1.5, and 2.5. 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.V (showing average coefficients) and A.VI (showing median coefficients) the change is nearly always in the anticipated direction.

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

$$\hat{\theta}_i = \psi_1 \alpha_i + \psi_2 \beta_i + \psi_3 \delta_{TC,i} + \epsilon_i \tag{30}$$

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 III, and the full regression output is included in Table A.VII.

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.V: Changes to Mean Power Law Coefficient through 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.VI: Changes to Median Power Law Coefficient through 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.VII: Estimated Coefficients from Regression 30

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

Appendix References

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*, 129(3):1085–1140.

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

Ben Arous, G. and Guionnet, A. (2010). *Wigner matrices*, pages 433–451. Oxford University Press.

Bradley, R. C. (1993). Equivalent mixing conditions for random fields. *The Annals of Probability*, 21(4):1921–1926.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2(N.A.).

Dingel, J. I., Miscio, A., and Davis, D. R. (2021). Cities, lights, and skills in developing economies. *Journal of Urban Economics*, 125:103174.

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

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

Gibrat, R. (1931). *Les inégalités économiques.* Paris: Librairie du Recueil Sirey.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities–double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2018). The Global Distribution of Economic Activity: Nature, History, and the Role of Trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Ioannides, Y. and Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, 73(1):18–29.

Marlow, N. (1967). A Normal Limit Theorem for Power Sums of Independent Random Variables. *Bell System Technical Journal*, 46(9):2081–2089.

Mazmanyan, L., Ohanyan, V., and Treitsch, D. (2008). The Lognormal Central Limit Theorem for Positive Random Variables. *Working Paper*.

Nunn, N. and Puga, D. (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics*, 94(1):20–36.

O'Rourke, S., Vu, V., and Wang, K. (2016). Eigenvectors of random matrices: A survey. *Journal of Combinatorial Theory, Series A*, 144:361–442.