Populations in Spatial Equilibrium

Matthew Easton and Patrick W. Farrell*

This Version: February 2024

Abstract

The appearance of power law-like distributions for city populations is a distinctive, recurring feature of human geography. We propose an explanation for this phenomenon that reflects both variation in geography and trade between locations. Realistically modeling geography as the determinant of a location's exogenous productivity and amenity value results in lognormally distributed locational "fundamentals." Given these fundamentals, populations are lognormally distributed within a broad class of quantitative spatial models and appear to follow a power law for the most populous locations (i.e., cities). Simulations mirror several empirical results in the literature on city population distributions.

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

The distribution of city populations in most countries appears to follow a power law. This distinctive empirical regularity is remarkable given the substantially different contexts in which it has been observed, having been documented extensively across countries,[1], varying definitions of cities.[2], and different periods of human history spanning millennia.[3] That is, a power law-like distribution appears even when the meaning of "city," the local externalities shaping cities, the level of development and structure of the economy, and the integration of cities into national or global networks all vary greatly.

The prevailing explanation for this remarkable pattern is the concomitant observation that, for many cities, population growth appears orthogonal to the population level. A "random growth" process reflecting this phenomenon can generate power law-like city population distributions, but the assumption of random growth is inconsistent with the empirical evidence on the distribution of cities in significant ways. Cities tend to recover rapidly following major shocks,[4] and growth does not appear random during transitions to new spatial equilibria.[5] Further, random growth explanations fail to capture the influence of the observable characteristics of a place on the attractiveness of producing or residing there. Random growth models imply the large populations of New York City, Tokyo, and London are unrelated to their advantageous locations. Moreover, many existing theories of the city size distribution are aspatial and

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

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

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

[4]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.

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

do not allow for interactions between locations to shape settlement patterns, failing to capture the contribution of trade to the scale of the aforementioned global cities.

We provide an explanation for the appearance of power law-like city size distributions that allows both place and space to shape settlement patterns, bridging the literatures on population distributions, economic geography, and modern quantitative spatial models. Modern quantitative spatial equilibrium models capture many of the forces that shape human geography.[6] While these models can rationalize population distributions in terms of recoverable locational "fundamentals," they do not explain why the distributions of fundamentals or population take a particular form.[7] The economic geography literature emphasizes that geographic attributes influence settlement patterns,[8] and we place this focus on observable geography within a modern spatial model. Modelling fundamentals as resulting from variation in geographic attributes, we demonstrate that the resulting city size distribution will appear to follow a power law within a broad class of spatial models. The result is robust to changes in model parameters, allowing us to explain the persistent appearance of, and variation in, this distribution in different contexts. Given the persistence of geography, our framework can explain the recovery of cities from negative shocks. We also show that "random growth" is a characteristic of the equilibrium, but not the force creating the distribution as in prior theories.

Our approach begins with a focus on the characteristics of population distributions. First, we argue that population distributions appear to be lognor-

[6]This literature includes Allen and Arkolakis (2014), Redding (2022), and Redding and Rossi-Hansberg (2017).

[7]Given certain parameter values, these models can be inverted to recover fundamentals to rationalize arbitrary, unrealistic vectors of "populations."

[8]See Henderson et al. (2018) for a detailed investigation of the role of geography.

mal and show why the lognormal distribution is difficult to distinguish from a true power law in the tail. We demonstrate that, given sufficiently high variance of the log population, the often-identified "Zipf's" power law will appear for some truncation of the population distribution consisting of only highly populous locations (i.e., cities).[9] We discuss the frequent appearance of the lognormal distribution, which results from many multiplicative processes (a consequence of the central limit theorem applied in logs) as well as, surprisingly, certain additive ones (applying a lemma from Marlow (1967)). We draw on both of these properties when characterizing the population distribution.

We then investigate how heterogeneity in observable geographic attributes influences the suitability of locations for production and habitation. We argue that a location's fundamentals should be determined by its many geographic attributes. Some geographic attributes are uncorrelated within places (like rainfall and topography), and the correlation of geographic attributes between places declines with distance. Modelling fundamentals as multiplicative functions of attributes with these two key properties results in locational fundamentals which are lognormally distributed and spatial correlated, with the degree of correlation declining with distance.[10] Using a granular panel of attributes, we show that the two properties are clear features of the real world.

Embedding this exogenous geography within a discrete version of the spatial model in Allen and Arkolakis (2014), which nests many spatial models, we demonstrate our key result that the equilibrium population will follow a lognormal distribution. This result follows from each location's own lognormal

[9]Zipf's law is a specific power law frequently identified at the country level, characterized by a slope of -1 on a plot of log-population and log-population rank. An example for the U.S. can be see in Figure 1. We discuss this power law and the role of truncation in more detail in Section 1.

[10]Our approach follows that of Lee and Li (2013), but with an additional focus on the spatial distribution of the resulting fundamentals.

fundamentals as well as trade with other locations. Within spatial models, other locations enter additively into the equilibrium population condition. By applying the lemma of Marlow (1967) to this sum, we are able to characterize the population distribution within a broad class of spatial models. We demonstrate the success of the model at generating both lognormal full population distributions and power law-like city population distributions through simulation. We explore how changes to local productivity spillovers, intra-city congestion externalities, and inter-city transportation costs influence the observed city-size distribution.

This paper touches on several topics within the spatial economics literature. First, it relates to work characterizing the city size distribution. Some work has argued that the city size distribution follows a Pareto distribution, such as Gabaix (1999a,b), ignoring smaller settlements and interpreting the truncated city size distribution as reflecting a true power law. Other work, such as Eeckhout (2004), has argued that a lognormal distribution better describes the full population distribution, appears similar to a Pareto distribution for large observations like cities, and naturally captures deviations from a true power law in the data. We argue in favor of the lognormal interpretation, and show how lognormality emerges within a broad class of spatial models.

Second, we relate to work on the origin of power law-like city size distributions. Many theoretical explanations of this phenomenon are based on the similarly striking empirical observation that city growth rates often appear unrelated to city population, referred to in the literature as Gibrat's law.[11] Much of the literature uses random growth as the basis for the appearance of

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

a power law-like distribution, as in Gabaix (1999a), Gabaix (1999b), Blank and Solomon (2000), Eeckhout (2004), Rossi-Hansberg and Wright (2007), and Córdoba (2008). However, when subject to large negative shocks such as war and disease, cities tend to recover rapidly to their prior position in the distribution (Davis and Weinstein, 2002; Brakman et al., 2004; Davis and Weinstein, 2008; Johnson et al., 2019). This "reversion" to the prior distribution is incompatible with random growth theories. Further, Desmet and Rappaport (2017) demonstrate that Gibrat's law did not hold throughout the historical settlement of the U.S. and only emerged after the end of westward expansion. These facts are consistent with locational fundamentals being a determinant of the observed population distribution in equilibrium, as argued in Davis and Weinstein (2002). We build on earlier aspatial models that have taken a fundamentals-based approach to explaining the population distribution (Lee and Li, 2013; Behrens and Robert-Nicoud, 2015; Desmet and Rappaport, 2017), while adding a role for space lacking in prior theoretical work by placing the exogenous geography within a spatial model.

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

graphic attributes.

A fourth related literature is that on spatial models, focused on the role of space, local spillovers, and the importance of trade and interactions between locations in determining population distributions (Fujita et al., 1999; Allen and Arkolakis, 2014). Indeed, the largest cities tend to be favorably located for trade with other locations.[12] Early spatial models such as Krugman (1991) primarily focused on the role of local population spillovers and trade, and assumed no differentiation in first-nature geography across locations.[13] Modern spatial equilibrium models Allen and Arkolakis (2014) also incorporate variation in exogenous fundamentals reflecting a differentiated geography. For certain parameter values, these models can be inverted to recover the fundamentals given any distribution of population. However, absent a theory for the distribution of fundamentals, this literature cannot explain why population tends to be distributed similarly in many different contexts. Our theory of geography-based fundamentals allows us to identify a mechanism for generating lognormal populations within these models.

At the intersection of these literatures, our work is the first to generate realistic city size distributions based on heterogeneous geography within a broad class of spatial models. The paper proceeds as follows. Section 1 argues that populations are best described by a lognormal distribution and establishes the link between this distribution and the appearance of a power law for cities. Section 2 investigates geographic attributes and their distribution

[12]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 almost surely 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.

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

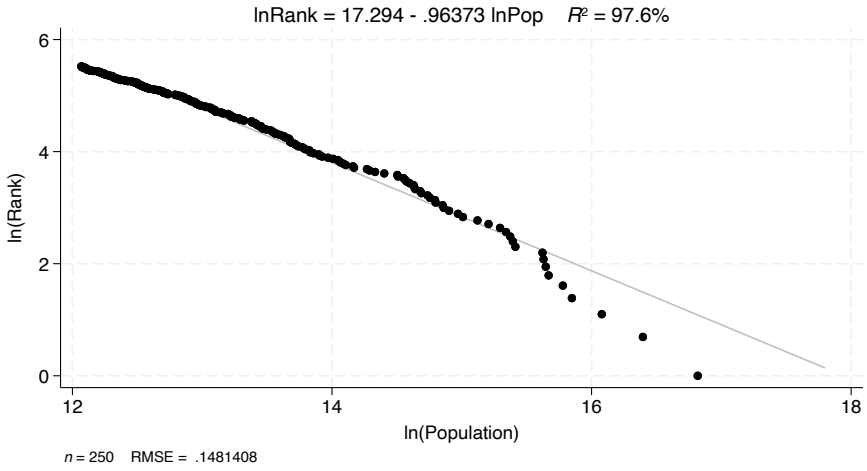


Figure 1: The appearance of a power law for top 250 U.S. metropolitan statistical areas (MSAs) in 2020. *Data Source: U.S. Census*

within places and across space, modelling locational fundamentals based on variation in geography. Section 3 demonstrates the lognormality of the population distribution within a broad class of quantitative spatial models when locational fundamentals are lognormally distributed. Section 4 uses numerical simulation of the model to demonstrate its ability to capture several results in the empirical literature on city size distributions. Section 5 concludes.

1 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. The regularity of its appearance across countries, the definition of a "city," and over time means it can reasonably be held as a minimum criterion for a model of the spatial economy[14] This distribution is typically illustrated with a simple plot and accompanying regression. For some truncation of the population distribution to include only

[14]As articulated in Gabaix (1999b).

the most populous locations ("cities"), the plot of the log population rank of a city and the log population of the city often appears strikingly linear, and a regression given by:

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

for many countries delivers a high R^2 (over 0.95) and frequently an estimate for θ_1 near -1, as in Figure 1 for U.S. cities. This slope is characteristic of a specific power law referred to as Zipf's law, which can be stated as the largest city in a given country being n times the size of the n^{th}-largest city. Interpreting this regression as describing the true city size distribution would mean that city populations follow a Pareto distribution with shape parameter $\alpha_P = 1$ and minimum city size x_m reflecting the choice of truncation point.[15]

Instead of being a true power law, the city population distribution may be better characterized by a different 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 2, which shows that this continues to hold for the U.S. in 2010. The tail of a lognormal distribution often appears similar to a Pareto distribution, which can be understood by considering the lognormal PDF:

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

[15]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).

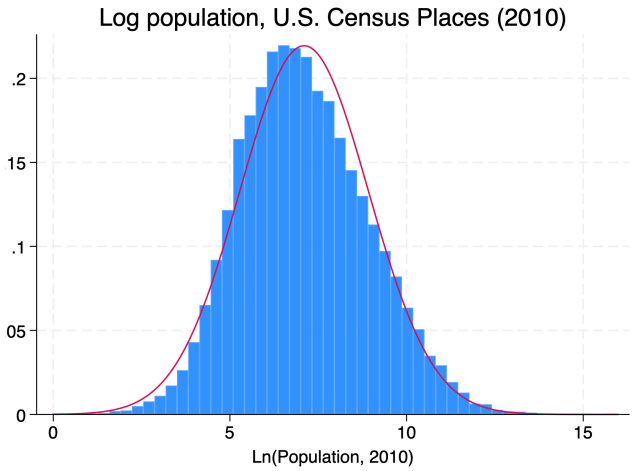


Figure 2: Histogram of log population for U.S. incorporated places and census designated places in 2010, with an overlaid normal distribution matching the moments of the empirical log distribution. The fit of the log population to the normal distribution distribution means the population distribution appears lognormal. This figure is an update to Figure 2 of Eeckhout (2004), which uses data from the 2000 Census. *Data Source: 2010 U.S. Census*

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

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

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

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

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

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.[16] With more complete data on population distributions the choice of a truncation point to support the Pareto interpretation becomes critical and there is no accepted method for determining such a cutoff. Many researchers rely on a visual test of the data to determine a cutoff (Gabaix, 2009). Second, beyond the need to truncate the data to fit a Pareto, models generating a Pareto population distribution must rely on unrealistic assumptions regarding city growth dynamics. 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.

The lognormal interpretation's attractive properties stand in direct contrast to the shortcomings of the Pareto. Regarding the Pareto distribution's inability to match the full population distribution, the lognormal distribution appears to fit both the body, obviating the need for arbitrary truncation, as well as the right tail. Further, the scale-varying "shape parameter"-like term of the lognormal (as shown in Equation 3) can explain commonly observed deviations in real-world city size distributions. The likelihood of very large cities is lower when the true distribution is lognormal than for a similar Pareto, because the scale-varying "shape parameter"-like term is increasing in x. This

[16]This is true of early work, such as Auerbach (1913) (while Auerbach had data on many small settlements, a table in his paper includes just the 94 largest; see the recent translation in Auerbach and Ciccone (2023)) and Zipf (1949). Even more recent investigation of Zipf's law in Krugman (1996), for instance, included just the top 135 cities as the *Statistical Abstract of the United States* included only those cities (Eeckhout, 2004).

appears to match the global city distribution (Rossi-Hansberg and Wright, 2007), as the largest cities in most countries tend to fall below the slope of the illustrative power law regression line (evident in Figure 1).[17] Other characteristics 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.[18]

Further, in contrast to the unrealistic conditions necessary to generate a Pareto distribution, the lognormal distribution can appear under very general conditions as the result of a central limit theorem. Many random variables resulting from a multiplicative process tend to lognormality as these processes are additive in logs (Roy, 1950).[19] This is the mechanism by which random many growth-based models generate a lognormal population distribution, such as that in Eeckhout (2004). We will use a multiplicative process of this type in Section 2 to characterize the distribution of locational "quality" based on random variation in geography.

Less commonly noted is that the lognormal distribution can also appear as a result of additive processes for certain sequences of positive random variables. A lemma from Marlow (1967), reproduced below, provides conditions under

[17]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 Appendix B.

[18]This property is discussed at length in Eeckhout (2004) and demonstrated in Appendix Figure A1 where we expand or reduce the number of cities relative to Figure 1. 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.

[19]For discussion of the many contexts in which a lognormal distribution appears, see Limpert et al. (2001).

which a lognormal distribution may appear given a summation of positive random variables:

***Lemma 1 (Marlow, 1967)**: Let $\{S_n\}$ 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, $\lim_{n\to\infty} P\left\{\frac{S_n - a_n}{b_n} \leq x\right\} = F(x)$*

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

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

Lemma 1 provides conditions for going from a central limit theorem in levels to one in logs. Condition (i) can reflect convergence under a central limit theorem, where $F(x)$ is the standard normal distribution and the sequences a_n and b_n are the mean and standard deviation of some S_n resulting from a sum of random variables. Condition (ii) then necessitates that the coefficient of variation of S_n is zero in the limit. Many sums of positive random variables fulfill this requirement. For a sum that satisfies the conditions for a central limit theorem and condition (ii), Lemma 1 states that the given normalization of the sum will converge in distribution to a lognormal. Examples of sums over several positive random variables (lognormal, truncated normal, and uniform) are presented in Figure 3, exhibiting the appearance of normality in both levels and logs for these sums.[20] This result is crucial for characterizing the population distribution within many quantitative spatial equilibrium models, which we return to in Section 3, as the contribution of trade (which is always positive) enters the equilibrium condition additively.

[20]We discuss Lemma 1 further in Appendix A.

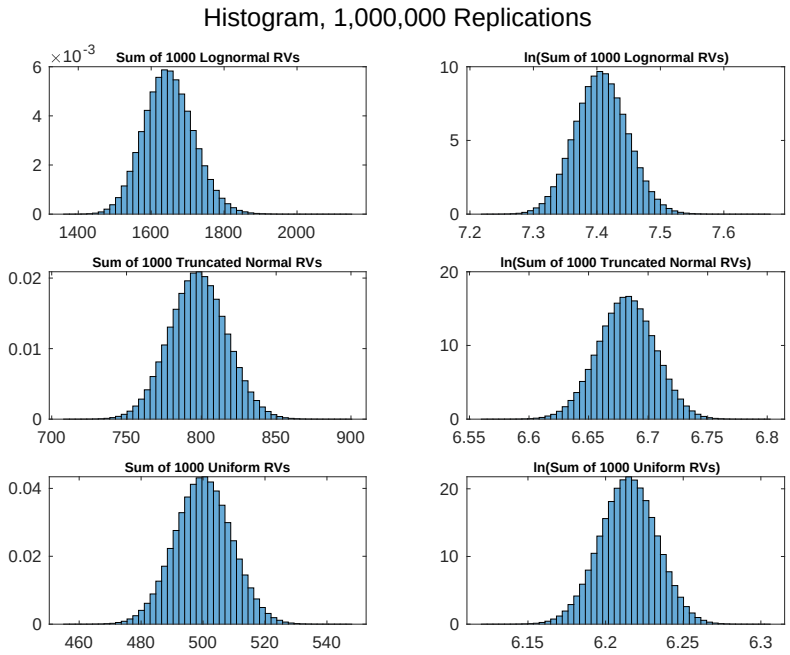


Figure 3: Sums of positive random variables drawn from various distributions. The random variables in the first row are draw from a lognormal distribution with parameters $\mu_{LN} = 0, \sigma_{LN} = 1$, the middle row from a truncated normal distribution with parameters $\mu_{TN}, \sigma_{TN} = 0$ and minimum value $\alpha = 0.001$, and the bottom row from a uniform distribution on $(0, 1]$. The sums appear distributed normally in both levels (column 1) and in logs (column 2), as implied by Lemma 1.

The appearance of a power law-like city population distribution is likely the result of a focusing on the tail of the true lognormal distribution of human populations. Such an interpretation requires fewer restrictive or arbitrary 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). The general conditions under which the lognormal distribution can appear due to random variation also make it a plausible candidate for the population distribution. We demonstrate that the equilibrium population distribution within many spatial models results is lognormal given a realistically modeled geography.

2 Geography, Attributes, and Fundamentals

An important starting point for understanding the distribution of population is noting that Earth's geography is highly varied. Climatic conditions, soil quality and type, and topography, among many other attributes of a place, vary greatly around the world and there is clear evidence for observable geographic attributes, alone and in combination, playing a role in shaping human settlement patterns. 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 the quality of a place that makes it good for human habitation. There are highly populous locations which have steep topographies (La Paz) and those with flat topographies (Houston), those which are on a coast (Rio de Janeiro) and those which are far from the ocean (Ulaanbataar), and those which have consistent rain (Dublin) and those which have little rain (Phoenix).

Despite such heterogeneity across individual attributes within any given location, those locations that are home to cities must offer some advantage to their inhabitants relative to other less populated locations. Even though it is located in a desert and has hot summer temperatures, Phoenix offers a favorable topography for a city, regular sunshine, and pleasant winter temperatures. Indeed, while Phoenix is often presented as an example of a fast-growing city in an unfavorable, unexpected, or even untenable location, the land where Phoenix exists now was once home to the Hohokam people who there developed one of the densest pre-Columbian settlements in North America (Doyel, 2001).[21] That is, even if some attribute of a place are not favorable, other

[21] The Arizona Museum of Natural History website offers a further discussion of the Hohokam people, who constructed complex irrigation canals to support an estimation population of as many as 100,000 people in the Phoenix basin 1000 years ago.

attributes may be—populous places have *something* going for them, but that *something* may be different across different locations. That there are many different combinations of attributes that can make a place habitable suggests that a wide variety of attributes contribute to the "quality" of a location.

In this section, we explore how to integrate a realistic geography into quantitative spatial models, noting both the geographic heterogeneity of populous places and two key aspects of the distribution of geographic attributes. First, while some attributes are correlated (such as high July temperatures and growing days) there are others that appear unrelated (such as topography and rainfall). Second, while geographic attributes tend to be similar for nearby places, over greater distances there are large differences in the geographic attributes of locations.[22] We explore the implications of variation in geographic attributes for determining the attractiveness of a location, and then empirically demonstrate that geographic attributes appear to satisfy key necessary assumptions to support our modelling decisions.

2.1 Linking Geography and Locational Fundamentals

We now formalize a means of incorporating a more realistic geography into spatial models, providing a microfoundation for the distribution of the locational fundamentals within these models based on random variation in geography. Our approach builds on that of Lee and Li (2013), who similarly model a location's quality as resulting from many random "factors," with an added focus on the spatial distribution of the resulting lognormal fundamentals we

[22]The geographic and climatic similarities of locations nearby in space reflects the "first law of geography," formulated by Tobler (1970) as "everything is related to everything else, but near things are more related than distant things."

integrate into a modern spatial model.[23]

Observable geographic *attributes* impact the exogenous productivity and amenity value of a location, which we refer to as *fundamentals*. Attributes a_{itk} are associated with a location i at time t with the type of attribute indexed by $k \in K$.[24] We assume all attributes are strictly positive in value—no place has less than zero access to water, or completely zero access—and for any k higher values of a_{ikt} reflect *better* draws.[25] 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 model the productivity and amenity fundamentals similarly, and for brevity focus our discussion on the productivity fundamental before returning to consider amenities. The locational productivity fundamental for a location i at time t, denoted A_{it}, should be a function of its many attributes a_{ikt}: $A_{it} = F(a_{i1t}, a_{i2t}, ..., a_{iKt})$. The productivity fundamental for location i should be increasing in each a_{ikt}, to reflect that better attribute draws increase productivity: $\frac{\partial F}{\partial a_{ikt}} > 0$ for all $k \in K$. Further, the aggregating function should exhibit complementarities between each of the attributes—the benefit of having reliable rainfall for production is increased when there is better arable land in a location, for instance. This means the aggregating function also needs a positive cross-partial for all arbitrary combinations of attributes: $\frac{\partial^2 F(\cdot)}{\partial a_{ijt} \partial a_{igt}} > 0$, for $j, g \in K, j \neq g$.

[23]Other papers that have adopted the Lee and Li (2013) approach to modeling fundamentals include Behrens and Robert-Nicoud (2015) and Desmet and Rappaport (2017), but these models are also aspatial and do not consider the distribution of the resulting fundamentals across space.

[24]We will assume that K is large, as many attributes impact productivity and amenity values. In applying a central limit theorem later, we will assume $K \to \infty$.

[25]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, where we instead want to consider a measure reflecting how positive the "shock" from a given attribute is.

Consistent with these assumptions, we can view the contribution of attributes to the fundamental as representing multiplicative "shocks." The varying importance of different attribute can be reflected by adopting a Cobb-Douglas form for the aggregating function, where the powers reflect the weight placed on the attributes. These weights may change over time to capture structural transformation or changing production technologies. Allowing the productivity-relevant weight at time t for attribute k to be denoted $\xi_{kt} > 0$:

$$A_{it} = \prod_{k \in K} {a_{ikt}}^{\xi_{kt}}$$

For simplicity, we suppress the t subscript as we are not considering change over time. Taking the natural log yields the following expression:

$$\ln(A_i) = \sum_{k \in K} \xi_k \ln a_{ik} \tag{5}$$

Before aggregating varied attributes to characterize the distribution of A_i, we impose some distributional assumptions on the attributes so we can apply a version of the central limit theorem that allows for some correlation among the attributes a_{ik} within each location i. While we can allow for some pairs of attributes to be correlated with each other within a location (e.g., July temperature and growing days), we require that, over the very large number of attributes of a place, there exist pairs of attributes which are nearly independent (e.g., topography and rainfall). This latter requirement is formalized by the concept of weak dependence or α-*mixing*, as defined below.

Definition,* α*-mixing.: *For a sequence $x_1, x_2, ...$ of random variables, let α_n be a number such that:*

$$|P(A \cap B) - P(A)P(B)| \leq \alpha_n$$

for $A \in \sigma(x_1, ..., x_k)$, $B \in \sigma(x_{k+n}, x_{k+n+1}, ...)$ *and* $k \geq 1, n \geq 1$. *Suppose* $\alpha_n \to 0$, *the idea being that* x_k *and* x_{k+n} *are then approximately independent for large* n. *In this case the sequence* $\{x_n\}$ *is said to be* α*-mixing (Billingsley, 1995).*

Together with further restrictions on the moments of the attributes a_{ik} and the rate of α-mixing, we can apply the central limit theorem as in Corollary 1 of Herrndorf (1984), introduced to the spatial literature in Lee and Li (2013), to characterize the distribution of A_i.

Lemma 2 (Herrndorf (1984))*: Let* $\{s_i\}$ *be an* α*-mixing sequence of random variables (denote the sequence* $\bar{\alpha}_i$*) satisfying the following conditions...*

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

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

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

iv. $\sum_{i=1}^{\infty} (\bar{\alpha}_i)^{1-\frac{2}{b}} < \infty$

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

Given these conditions, we can apply Lemma 2 (Corollary 1 of Herrndorf (1984), or Theorem 3 of Lee and Li (2013)), and 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 be lognormally distributed. Note that few assumptions are required on the precise distribution of the underlying geographic attributes. We can allow each attribute to differ in the distribution from which it is drawn (so long as the moment conditions above are fulfilled), and we can allow attributes to vary in their degree of correlation with other attributes within a location (so long as the α-mixing rate restriction holds).

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 $\iota_k > 0$, is:

$$\ln(U_i) = \sum_{k \in K} \iota_k \ln a_{ik} \tag{6}$$

and, given the same conditions as on the productivity fundamental, will also be lognormally distributed. We assume that the weights and attributes are such that $\ln(A_i)$ and $\ln(U_i)$ have a bivariate normal distribution.

Using random variation in geography to model locational fundamentals is the cross-sectional analog of random growth models based on Gibrat's law, like those of Eeckhout (2004) and Gabaix (1999b). Rather than random growth shocks, here locations receive shocks via random variation in geographic attributes. This approach has been used before in the urban economics literature but the resulting fundamentals have not previously been integrated into a spatial model.

To do so, we must accurately reflect the distribution of attributes across locations in modelling the fundamentals. We allow individual attributes to be *spatially correlated* across locations, such that nearby locations may have broadly similar attributes, but require that this spatial correlation declines with distance. The assumption of spatial correlation does not impact the application of the central limit theorem used to characterize the distribution of fundamentals, which applies only within a location, but has implications for the spatial distribution of the resulting locational fundamentals. As the unobservable fundamentals are a function of the observable features of place, we assume they will have a spatial correlation pattern similar to that of attributes and exhibit a spatial correlation that declines with distance—this assumption

help us to characterize the resulting population distribution.

2.2 Empirical Evidence

Next, 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 weak dependence between attributes within a place, used to apply the central limit theorem above to characterize the fundamentals, and the assumption of weak dependence of fundamentals 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.[26] The dataset is at the quarter-degree latitude and longitude cell level and we focus on the roughly 47,000 cells grid cells in the U.S., Mexico, and Canada, with the nearly 13,000 of those grid cells contained in the contiguous U.S. serving as our main sample.[27]

First, we explore the correlation between attributes within a given location and show that weak dependence of attributes is not an unreasonable 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

[26]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 which were categorical or discrete were excluded from our analysis. Refer to Appendix C for more information.

[27]At the equator, a grid cell is ∼28-by-28 km; at 48 degrees latitude, ∼18-by-18 km. The reduction in the number of attributes and geographic scope does not drastically decrease the explanatory power of the attributes on economic activity relative to Henderson et al. (2018); see Appendix Table A1 for a regression showing that our eleven attributes explain 43% of the variance in economic activity in the contiguous U.S., in line with the 47% Henderson et al. (2018) found globally with their full set of attributes.

in the contiguous U.S., as shown in Figure 4a.[28] 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 4b.[29] 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.[30]

3 Lognormal Populations in Spatial Models

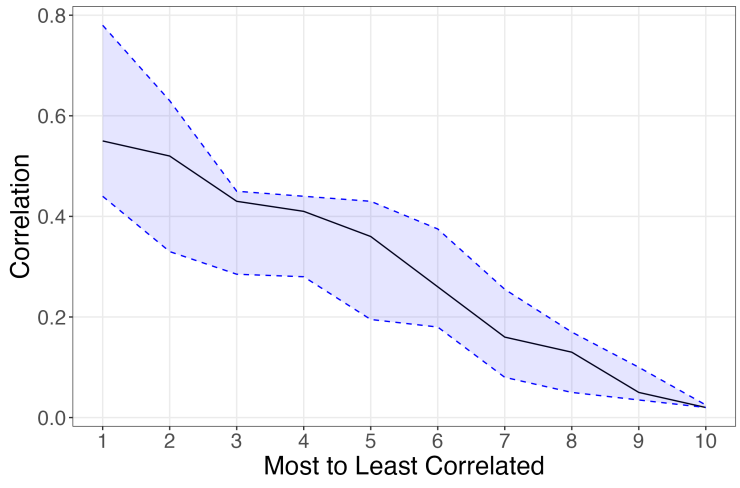
We now describe a quantitative spatial model with lognormally distributed productivity and amenity fundamentals, where the spatial correlation between these fundamentals declines with distance. It is a discretized version of the model in Allen and Arkolakis (2014), which nests a broad class of spatial models.[31] We demonstrate that the equilibrium population distribution within this model will be lognormally distributed. Populations do not just inherit the log-

[28] A description of how we calculated cross-correlation is provided in Appendix E

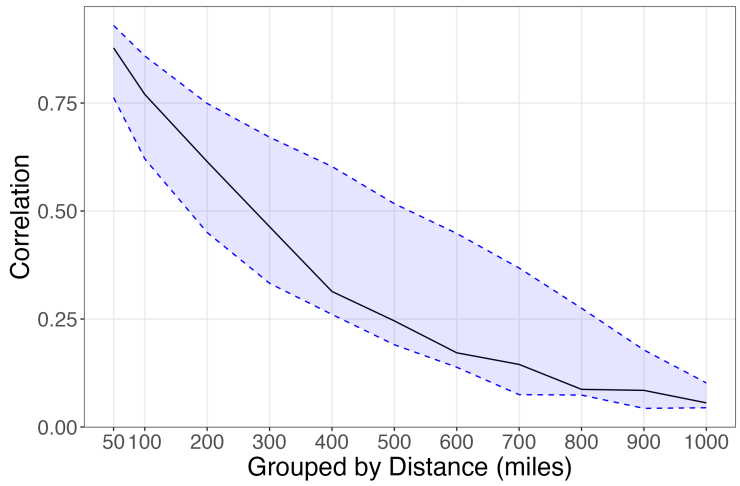
[29] A description of how we calculated spatial correlation is provided in Appendix E.

[30] In Appendix D, we also demonstrate that the eleven attributes in our dataset can be aggregated as described in this section to produce a lognormal "fundamental." Behrens and Robert-Nicoud (2015) showed that as few as six attributes can be aggregated to produce a fundamental that appears lognormal.

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



(a)



(b)

Figure 4: (a) Cross-correlation and (b) spatial correlation structure of U.S. geographic attributes. The solid black line represents the median correlation; the blue dashed lines represent the 25th and 75th percentile bands. *Data Source: Authors' calculations using data from Henderson et al. (2018)*

normal first-nature geography in each location, as in prior fundamentals-based work. Rather, the equilibrium condition in the model includes a weighted summation over all locations, and so trade between locations plays a key role. We use the lognormal central limit theorem from Marlow (1967) in order to

characterize the distribution of this summation and the resulting equilibrium population distribution.

The world consists of discrete habitable locations indexed by $n \in N$ distributed over a space S, where S is a closed bounded set of a two-dimensional Euclidean space with the Euclidean norm as its metric.[32] Trade between locations is costly, with trade costs between locations i and n denoted $\tau_{i,n}$ taking a symmetric iceberg form such that $\tau_{i,i} = 1$, $\tau_{i,n} = \tau_{n,i}$, and $\tau_{i,n} > 1$ for $i \neq n$. Trade costs are a function of the instantaneous trade costs incurred along the least cost path between locations in N while traveling across the surface S as in Allen and Arkolakis (2014). The instantaneous trade costs of traversing points in $s \in S$ are themselves a function of random variation in geography.[33] As a result, a location i's trade costs with two other locations n and m will be similar when n and m are close to one another and potentially very different with n and m are distant from one another. As these costs result from random variation in geography and exhibit spatial correlation, $\tau_{i,n}$ is a spatially correlated random variable.

Each location has a productivity fundamental A_i and an amenity fundamental U_i, which are functions of the geographic attributes a_{ik} of each location. Attributes are at most weakly dependent within a location and their spatial correlation declines with distance, as argued in Section 2. Aggregating over the attributes as in Section 2 results in lognormally distributed exogenous productivity and amenity fundamentals which exhibit declining spatial correlation with distance. A place's effective productivity and amenity value may

[32]The choice of a two-dimensional geography is a simplifying one to reflect a realistic geography, and as in Allen and Arkolakis (2014) the results hold in higher dimensional spaces. We do not take a stance on the distribution of habitable locations, but require that no two locations $i, n \in N$ exist at the same point in $s \in S$.

[33]For instance, travel across mountainous regions will incur higher trade cost than travel across plains or along rivers.

also be affected by negative or positive population externalities. We define the "composite fundamentals" as:

$$\tilde{U}_i = U_i L_i^{\beta} \tag{7}$$

$$\tilde{A}_i = A_i L_i^{\alpha} \tag{8}$$

where the typical case will consist of $\beta < 0$ and $\alpha > 0$, reflecting the negative impact of overcrowding on amenities and positive productivity spillovers from agglomeration.

Geography within this model is represented by the set of functions defining the locational fundamentals, $\tilde{A}$ and $\tilde{U}$, along with the trade costs function τ defining the spatial relationship between locations in the model. Within this discrete version of Allen and Arkolakis (2014), we use the term "regular geography" to describe a geography in which all locations have strictly positive, finite values for $\tilde{A}$ and $\tilde{U}$, and trade costs $\tau_{i,n}$ are similarly bounded above and below by strictly positive numbers between all locations.

We follow Allen and Arkolakis (2014) and assume Armington-style production of a differentiated good in each location. There is a population of homogeneous workers $\bar{L} \in \mathbb{R}_{++}$ who can freely move to any location. Workers have common constant elasticity of substitution preferences over goods in their welfare function given by:

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

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

Production is perfectly competitive.[34] A worker in location i can produce $\tilde{A}_i$ units of the local differentiated good, where $\tilde{A}_i$ is the composite productivity fundamental of location i. The number of workers and wages in a location are given by the functions $L : N \to \mathbb{R}_{++}$ and $w : N \to \mathbb{R}_{++}$.[35]

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

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

where P_n is the price index in location n, given by:

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

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

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

Combining the quantity (Equation 11) and price (Equation 12) expressions, we can write the value of the good produced in i consumed by n as:

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

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

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

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

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

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

[35] No location will be unpopulated in equilibrium given the range of parameters we consider.

The labor market clears:

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

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

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

The welfare expression combined with the price index yields:

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

We now focus on the case of the model with spillovers and externalities, and demonstrate that the resulting population distribution is lognormal.[36] Given the form of the externalities in Equations 7 and 8, along with free movement between locations that ensures welfare is equal in all locations ($W_i = \bar{W}$ for all i), we can rewrite Equations 17 and 18 as:

$$L_i^{1-\alpha(\sigma-1)} w_i^{\sigma} = \bar{W}^{1-\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} A_i^{\sigma-1} U_n^{\sigma-1} L_n^{1+\beta(\sigma-1)} w_n^{\sigma} \tag{19}$$

$$w_i^{1-\sigma} L_i^{\beta(1-\sigma)} = \bar{W}^{1-\sigma} \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} L_n^{\alpha(\sigma-1)} U_i^{\sigma-1} w_n^{1-\sigma} \tag{20}$$

With symmetric bilateral trade costs the system can be re-written such that the equilibrium is characterized by a single equation, as shown in Appendix A, given by:

$$\bar{W}^{\sigma-1} L_i^{\tilde{\sigma}\gamma_1} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \left(L_n^{\tilde{\sigma}\gamma_1} \right)^{\frac{\gamma_2}{\gamma_1}} \tag{21}$$

where:

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

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

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

To simplify notation, denote $\left(\bar{W}^{1-\sigma} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma}\right)^{\frac{1}{\tilde{\sigma}\gamma_1}} = \Omega_i$ and each element of the summation term $\tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} L_n^{\tilde{\sigma}\gamma_2} = s_{i,n}$, where $S_i = \sum_{n \in N} s_{i,n}$. The equilibrium condition for L_i is then:

$$L_i = \Omega_i \left(S_i\right)^{\frac{1}{\tilde{\sigma}\gamma_1}} \tag{22}$$

where Ω_i is lognormally distributed, as the lognormal distribution is maintained over multiplication by positive constants, exponentiation, and multiplication by other lognormal distributions. This result is provided by bivariate normality of the logged distributions, which we earlier assumed holds for all A_i and U_i.

The population distribution is invariant to the total population $\bar{L}$, so we set $\bar{L} = N$ so that the average population is equal to 1.[37] We make this normalization such that the population in any particular location does not fall to 0 as N grows large. We can then characterize the resulting population distribution by the following theorem.

Theorem 1: *Consider a geography as described above with lognormally distributed fundamentals. For each $i \in N$, define the sequence $\{\omega_{i,0}, \omega_{i,1}, \omega_{i,2}, ...\omega_{i,N}\}$ such that for $n = 0$ we define $\omega_{i,0} = \Omega_i$ and $\omega_{i,n} = s_{i,n}$ for $n > 1$. We denote the demeaned sequence $\{\hat{\omega}_{i,n}\}$ such that $\mathbb{E}[\hat{\omega}_{i,n}] = 0$ for all i and n. If, for all i...*

[37] As proven in Allen and Arkolakis (2014), the population distribution is invariant to changes in scale, so this choice simply requires multiplying the population vector by a constant. We discuss this property, and its implications, in more detail in Section 4.

i. *The sequences* $\{\hat{\omega}_{i,n}\}$ *are* α*-mixing and fulfill the conditions in Lemma 2*

ii. *The coefficients of variation for sums over the sequences* $\{\omega_{i,n}\}$ *fulfill condition (ii) of Lemma 1*

...then, by an asymptotic argument, the population will approach a lognormal distribution.

Proof: The meaning of the necessary conditions on the demeaned $\{\hat{\omega}_{i,n}\}$ are, from Lemma 2, that (i) all elements are mean zero, (ii) the contribution of each element to the variance of the sum over the elements is vanishing (and the average contribution to the variance of each element is $\bar{\sigma}^2$ as $N \to \infty$), (iii) a bound on the higher-order moments, and (iv) a bound on the α-mixing rate. The condition on the original sequence $\{\omega_{i,n}\}$ is, from Lemma 1, that (ii) the coefficient of variation of the sums tend to zero.

The proof proceeds as follows. We first show that 1) S_i will approach a log-normal distribution, and then we show that 2) S_i will approach independence from Ω_i, such that the population will itself approach a lognormal distribution.

1. If the sequences $\{\hat{\omega}_{i,n}\}$ given by $\{\hat{\Omega}_i, \hat{s}_{i,1}, \hat{s}_{i,2}, ...\hat{s}_{i,n}\}$ are α-mixing and fulfill the conditions of Lemma 2 for all i, then the sub-sequences $\{\hat{s}_{i,1}, \hat{s}_{i,2}, ...\hat{s}_{i,n}\}$ must also be α-mixing and fulfill the conditions of Lemma 2.

   For each i, these sequences define $\hat{S}_i = \frac{1}{\sqrt{N}\bar{\sigma}} \sum_{n \in N} \hat{s}_{i,n}$, a demeaned and normalized S_i from the equilibrium condition in Equation 22. By the assumptions on $\{\hat{\omega}_{i,n}\}$, $\hat{S}_i \to \mathcal{N}(0,1)$ as $N \to \infty$. This means that, defining $\bar{M}_i = \mathbb{E}[S_i]$, we have $\frac{1}{\sqrt{N}\bar{\sigma}}(S_i - \bar{M}_i) \to \mathcal{N}(0,1)$, where S_i is positive by construction, satisfying condition (i) of Lemma 1. As condition (ii) holds such that $\lim_{N\to\infty} \frac{\sqrt{N}\bar{\sigma}}{\bar{M}_i} \to 0$, then $\frac{\bar{M}_i}{\sqrt{N}\bar{\sigma}} \ln\left(\frac{S_i}{\bar{M}_i}\right) \to \mathcal{N}(0,1)$ as $N \to \infty$ by Lemma 1.

By this asymptotic argument, for some large enough N, S_i will approach a lognormal distribution. As the lognormal distribution is maintained under exponentiation, $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ will approach a lognormal distribution for large N.

2. We proceed by noting that if the sequences $\{\hat{\omega}_{i,n}\}$ fulfill the restrictions of Lemma 2, this means that the contribution of any particular element to the sum given by $\hat{O}_i = \frac{1}{\sqrt{N}\bar{\sigma}} \sum_{n \in N} \hat{\omega}_{i,n}$ is negligible, so no one element dominates the distribution. We define a truncated sum $\hat{O}_i^m = \frac{1}{\sqrt{N}\bar{\sigma}} \sum_{(n>m) \in N} \hat{\omega}_{i,n}$, and note that $\hat{S}_i = \hat{O}_i^0$.

   By α-mixing, as $m \to \infty$, we have $\hat{\omega}_{i,1} \perp\!\!\!\perp \hat{\omega}_{i,m}$ (alternatively expressed, $\hat{\Omega}_i \perp\!\!\!\perp \hat{s}_{i,m}$). Define $m = \lfloor N^{\frac{1}{3}} \rfloor$, so as $N \to \infty$, $m \to \infty$ but at a slower rate such that...

$$\begin{aligned} \lim_{N \to \infty} \hat{O}_i - \hat{O}_i^m &= \frac{\hat{\omega}_{i,1} + ... + \hat{\omega}_{i,m}}{\sqrt{N}\bar{\sigma}} \\ &\leq \frac{|\hat{\omega}_{i,1}| + ... + |\hat{\omega}_{i,m}|}{\sqrt{N}\bar{\sigma}} \\ &\leq \frac{m \cdot \max_{1 \leq j \leq m}(\hat{\omega}_{i,j})}{\sqrt{N}\bar{\sigma}} = 0 \end{aligned}$$

   ...as $\lim_{N \to \infty} \frac{m}{\sqrt{N}} \to 0$. Thus, as $N \to \infty$, $\hat{O}_i = \hat{O}_i^m$. As the above also holds for any $m' \in \mathbb{Z}$, it is also true that $\hat{O}_i = \hat{S}_i$ as $N \to \infty$.

   As $N \to \infty$, all elements of $\hat{O}_i^m$ are independent of $\hat{\Omega}_i$, and so $\hat{\Omega}_i \perp\!\!\!\perp \hat{O}_i^m$, which by equivalence with S_i in the limit means that $\hat{\Omega}_i \perp\!\!\!\perp \hat{S}_i$.

   As independence is maintained under the continuous transformation of raising S_i to a positive power, this asymptotic argument implies that, as N grows large, Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ approach independence.

Given the asymptotic arguments above, $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ will approach a lognormal and Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ will approach independence for all i as N grows large. As such,

these two lognormal terms have a bivariate normal distribution in logs. The population L_i will then appoach a lognormal as N grows large, as it results from the product of two independent lognormal random variables, as shown in Equation 22. ■

The assumption of α-mixing requires that, as the distance between two locations i and m increases, the dependence between the own-lognormal term $\Omega_i = \left(\bar{W}^{1-\sigma} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma}\right)^{\frac{1}{\tilde{\sigma}\gamma_1}}$ and the trade-related contribution from location m given by $s_{i,m} = \tau_{i,m}^{1-\sigma} U_m^{\tilde{\sigma}(\sigma-1)} A_m^{\tilde{\sigma}\sigma} L_m^{\tilde{\sigma}\gamma_2}$ falls to zero. This assumption appears reasonable, as the fundamentals and populations in one location will be virtually unrelated to those in distant locations.

The proof is an application of Lemma 1 and Lemma 2, allowing us to characterize the distribution of the sum S_i in the equilibrium condition as lognormal in the limit, and the independence of the own-lognormal term and the trade contribution in the limit. The population is then given by the product of two independent lognormally distributed random variables, and will itself follow a lognormal distribution. The conditions necessary are limited. We show via simulation that, given standard parameter values, these conditions appear to hold, result in lognormal population distributions, and fulfill the predictions of the theorem— the distribution of S_i appears normal in both levels and logs, as predicted by Lemma 1, and Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ are uncorrelated as implied by independence.

4 Results

We now demonstrate that the model is successful at generating lognormal population distributions and power law-like city size distributions via numer-

ical simulation. We provide comparative statics based on varying parameter values across simulations to document how changes in congestion, spillovers, and trade costs influence the observed power law. We identify changes to the power law in the directions implied by the empirical literature. Finally, we show that Gibrat's law holds within the model when the aggregate population increases, showing that size-invariant growth is a feature of the equilibrium distribution.

4.1 Simulation of the Population Distribution

We first simulate the model to demonstrate that the resulting populations are indeed lognormally distributed and that the city size distribution appears to follow a power law. Each location in the model should be interpreted as a settlement, which can potentially be any size.[38] We define the most populous 5% of locations as "cities" within the model, to demonstrate that the tail behavior of the resulting population distribution mirrors the appearance of a power law in empirical city size distributions.

Dispersion in trade costs play a key role in ensuring that realizations of the trade contribution for locations i, j become nearly independent as the distance between these locations d_{ij} grows large, as required by the proof. To ensure dispersion in trade costs which is consistent with the triangle inequality, we model settlements as occurring randomly over a large surface and take the euclidean distance between all settlements. This modelling choice can be interpreted as either the "effective" distance, which reflects the difficulty of travelling across some parts of the geography, or variation in seed locations of settlements over a geography with identical trade costs at all points in

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

the surface.[39] We simulate a large geography and discard those settlements in border regions to limit the impact of border effects on the distribution. We are left with a central geography consisting of approximately 20,000 settlements.[40]

In simulating the model, we take model parameters from the literature and from Allen and Arkolakis (2014) where possible.[41] We first create randomly-generated draws of exogenous productivity and amenity fundamentals with declining spatial correlation. Fundamentals are drawn from lognormal distributions with parameters $\sigma_{LN} = 1$ and $\mu_{LN} = 0$, and we induce spatial correlation using a Choleski decomposition.[42] We allow the productivity and amenity fundamentals to be correlated, as while these are constructed using different weights as in Section 2 there could still be some degree of correlation between the two fundamentals.[43]

The magnitude of local productivity spillovers is given by $\alpha = 0.03$.[44] The model contains an isomorphism which we use to parameterize congestion costs. As discussed in Allen and Arkolakis (2014), the model is isomorphic to one with a fixed quantity of housing where spending on housing is δ and $\beta = -\frac{\delta}{1-\delta}$. Congestion costs are parameterized to match a level of spending

[39] We could alternatively model trade costs as having an idiosyncratic component to ensure dispersion. We choose the more restrictive setting without an idiosyncratic component to demonstrate that only a limited degree of dispersion is necessary.

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

[41] In Appendix D we additionally plot the recovered county-level exogenous fundamentals from Allen and Arkolakis (2014), which appear lognormal.

[42] We assume the degree of spatial correlation of the log-scale fundamental declines exponentially, so that $\rho_{ij} = e^{-\delta_\rho d_{ij}}$. For $j = i$, this gives $\rho_{ii} = 1$ as $d_{ii} = 0$. We set $\delta_\rho = 0.5$.

[43] We allow for correlation between productivity and amenity fundamentals of $\rho_{AU} = 0.12$ in the simulations here, in line with the correlation between inverted fundamentals from Allen and Arkolakis (2014). The result is not sensitive to the choice of correlation between productivity and amenities.

[44] This is in line with the estimates in Combes et al. (2008) and those surveyed in Rosenthal and Strange (2004) and Combes and Gobillon (2015).

on housing of 25% of income, which gives a congestion parameter of $\beta = -\frac{1}{3}$.[45] We model trade costs as an exponential function of distance, $\tau_{ij} = e^{\delta_{TC} d_{ij}}$ and set $\delta_{TC} = 0.001$.[46] We vary these parameters in Appendix F.

We take draws for productivity and amenity values, parameterized as above, and simulate population distributions within the model to demonstrate that the resulting population distribution appears lognormally distributed. First, we check that the summation term S_i behaves as predicted by Lemma 1 and the conditions in Theorem 1, which imply that this term should appear normal in both levels and logs and approach independence from the own-lognormal term for each location. We demonstrate this by plotting the histogram of values of S_i for a given geography in both logs and levels, along with the respective QQ plots comparing both to a normal distribution, in Figure 5. The distribution in Figure 5 displays the expected patterns, with a good fit to the normal distribution in both levels and logs as implied by Theorem 1.[47] The average correlation betweel Ω_i and $S_i^{\frac{1}{\tilde{\sigma}\gamma_1}}$ over the 100 replications is 0.018, indicating the two terms are virtually uncorrelated as implied by the proof.[48]

Figure 6 shows the equilibrium population distribution associated with a random draw of productivity and amenity fundamentals. The log of the population distribution very closely matches the overlaid normal distribution, demonstrating lognormality. The upper-right panel shows a quantile-quantile (QQ) plot of the log population and a normal distribution, demonstrating very close fit throughout the full distribution.

[45] This is consistent with the estimates in Combes et al. (2019) and Davis and Ortalo-Magné (2011).

[46] Allen and Arkolakis (2014) find that the cost of road travel is .56 when the width of the continental U.S. is normalized to 1. The geography we simulate has a width of 1000, which would imply a scaled parameter of 0.00056, near the value we choose.

[47] The smaller than expected right tail may be attributable to too little dispersion in trade costs or the size of the grid we simulate.

[48] While not a sufficient condition for independence, no correlation is necessary condition.

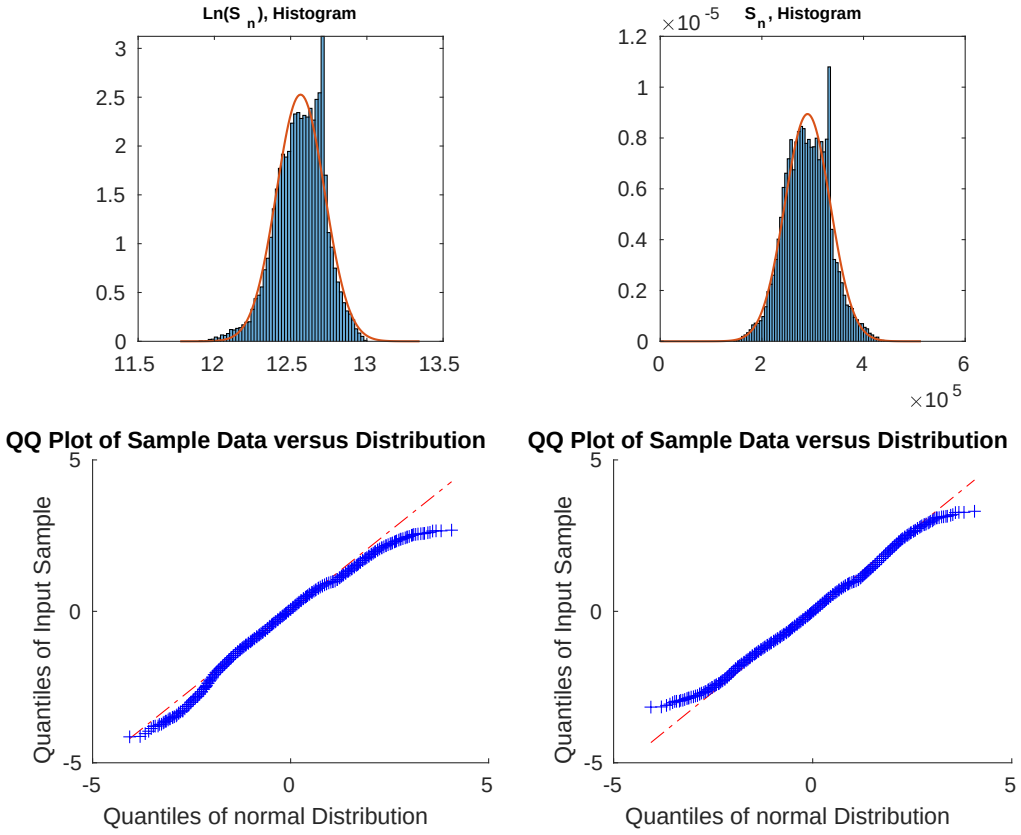


Figure 5: The figures above show a realization of the vector of S_i terms for some exogenous geography and associated population vector. The distribution of S_i appears normal in both levels and logs.

Concentrating only on the most populated 5% of locations, we find the model generates power law-like population distributions like those commonly identified in the data. Given the lognormality of the population distributions, the most populated locations in our model will appear to follow a power law distribution as demonstrated in Section 1. The log rank-size regression on this simulated data gives a slope of -0.977, close to the classic Zipf's law result of a slope of -1 for this particular random draw of fundamentals.

We next demonstrate the robustness of the approximate lognormal population distribution by performing 100 Monte Carlo simulations, each time

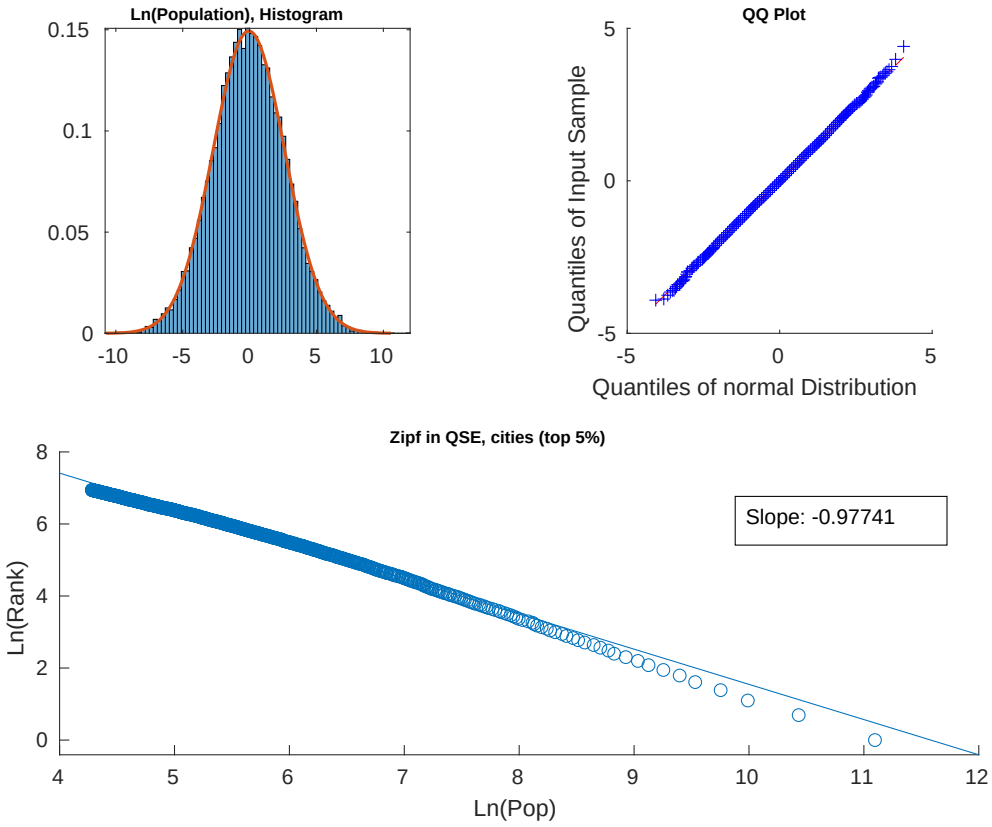


Figure 6: Example of the equilibrium population distribution. The top left panel shows the model's log population appears to follow a normal distribution. The top right panel contains a QQ plot of the model's log population distribution, indicating that it very closely matches a normal distribution. The power-law plot at the bottom shows a strong resemblance to the typical log rank-size plot along with the characteristic divergence of the largest locations below the trendline.

	Kolmogorov–Smirnov	Lilliefors	Jarque–Bera
Rejected at 1%	0.00	0.02	0.22
Rejected at 5%	0.00	0.1	0.40

Table 1: Table shows the share of tests for a normal distribution rejected for the log equilibrium population of 100 Monte-Carlo simulations.

drawing a new randomly generated geography. Figure 7 displays smoothed results over 100 simulations of the model.[49] The QQ plot also demonstrates

[49]The log of population is averaged at each rank of the distribution over the 100 simulations. Results are similar when averaging the population and taking the log.

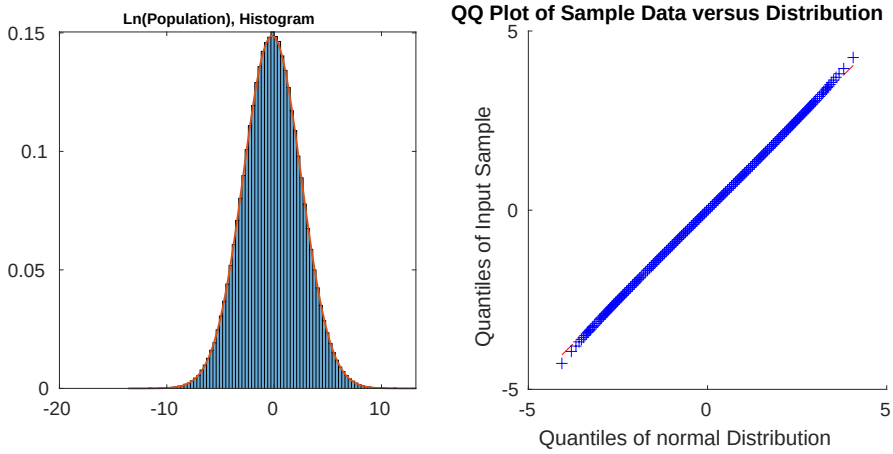


Figure 7: Results from smoothing the output over 100 Monte Carlo simulations of the model. The population distribution resulting from numerical simulation of model is in the left panel, and the resulting QQ plot is on the right. Both show that the equilibrium population distribution appears lognormal.

lognormality of the expected log population over these simulations. We test each of the 100 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 1. None of the tests reliably reject the normal distribution. Rejections of normality occur most often under the Jarque-Bera test, which tests for skewness and kurtosis. 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.[50]

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

	$\frac{\partial \theta_1}{\partial \alpha}$	$\frac{\partial \theta_1}{\partial \beta}$	$\frac{\partial \theta_1}{\partial \delta_{TC}}$
Sign of Change	+	−	+

Table 2: 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 has become flatter (larger largest cities), while "−" means the slope has become steeper (smaller largest cities).

4.2 Comparative Statics

We next perform comparative statics on the estimated power law coefficient for the city size distribution, testing the sensitivity to changes in model parameters. Changing these parameters alters the estimated coefficient of the log rank-size regression, which we denote θ_1 as in Equation 1. We perform 100 Monte Carlo simulations for each of 150 combination of parameters.[51] A summary of the signs of changes is provided in Table 2.

The comparative statics of our model demonstrate changes in the estimated power law coefficient in line with the empirical evidence. Increasing the benefits of agglomeration α results in a flatter slope (greater dispersion, or larger biggest cities) and increasing local congestion costs β results in a steeper slope (less dispersion, or smaller biggest cities). Increasing trade costs by increasing the rate at which these costs increase with distance, δ_{TC}, likewise results in a flatter slope. The flatter slopes for developing countries (documented in Duben and Krause (2021)) may be attributable to high domestic transportation costs, which are often substantially higher than in developed countries (Atkin and Donaldson, 2015). 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.

[51] We simulate for 5 values of α, 6 values of β, and 5 values of δ_{TC} given in Appendix F.

4.3 Gibrat's Law

Allen and Arkolakis (2014) demonstrate that the population vector is scaled by changes in the 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 $\bar{L}$.

The system of equations describing the population distribution, given for a particular location i in Equation 21 can be expressed in matrix form as:

$$\theta \mathbf{h} = \mathbf{J}[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}} \tag{23}$$

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 J, where $K_{i,n} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(\sigma-1)}$, is given by...

$$J = \begin{bmatrix} K_{1,1} & K_{1,2} & \dots & K_{1,N} \\ K_{2,1} & K_{2,2} & \dots & K_{2,N} \\ \dots & \dots & \dots & \dots \\ K_{N,1} & K_{N,2} & \dots & K_{N,N} \end{bmatrix}$$

We can write Equation 23 as

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

where $\tilde{h}_i = h_i \bar{W}^{\frac{\sigma-1}{1-\frac{\gamma_2}{\gamma_1}}}$. This must hold for any level of $\bar{L}$. As a result, changing $\bar{L}$ does not impact the resulting population distribution even as it impacts welfare ($\bar{W}$, which is the same across all locations). A percentage increase in overall population will result in each location experiencing population growth of the same percentage. That is, population growth rates are unrelated to initial population and Gibrat's law holds within the equilibrium of this model as the aggregate population increases.

This is a key difference between our explanation for observed population distributions based on locational fundamentals and trade and the prior literature on random growth models. Rather than being the force creating the equilibrium distribution, random growth is a feature of an equilibrium based on underlying geography. This view that is supported by the absence of Gibrat's law in systems that are in transition or have suffered disequilibrating 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 millennia. In this paper, we demonstrated that a broad class of economic geography models generate these characteristic population distributions when modeled with a realistic geography. We integrate insights from economic geography theory regarding the importance of both place and space into the extensive literature on power law-like population distributions and Zipf's law. Viewing population distributions as arising naturally in response to favorable geography and trade access provides a simple explanation for the emergence of the characteristic shape of the city size distribution. This explanation is consistent with the persistence of human settlements, the recovery of cities from disasters, and the random growth of cities in equilibrium.

References

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

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129(3):1085–1140.

Atkin, D. and Donaldson, D. (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 Ciccone, A. (2023). The law of population concentration. *Envmt. and Planning B: Urban Analytics and City Sc.*, 50(2):290–298.

Bai, J. and Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics*, 23(1):49–60.

Barjamovic, G., Chaney, T., Coşar, K., and Hortaçsu, A. (2019). Trade, merchants, and the lost cities of the bronze age. *The Quarterly Journal of Economics*, page 1455–1503.

Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogenous agents. *Handbook in Regional and Urban Economics*, Ch. 5.

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

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

Brakman, S., Garretsen, H., and Schramm, M. (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., Garretsen, H., Van Marrewijk, C., and Van Den Berg, M. (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., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P., Duranton, G., and Gobillon, L. (2019). The costs of agglomeration: House and land prices in french cities. *Review of Economic Studies*, page 1556–1589.

Combes, P.-P. and Gobillon, L. (2015). Chapter 5 - the empirics of agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 5, pages 247–348. Elsevier.

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

Davis, D. R. and Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, 92(5):1269–1289.

Davis, D. R. and Weinstein, D. E. (2008). A search for multiple equilibria in urban industrial structure. *Journal of Regional Science*, 48(1):29–65.

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

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

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

Doyel, D. (2001). *Late Hohokam*, pages 278–286. Springer US, Boston, MA.

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

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, pages 1429–1451.

Fujita, M., Krugman, P., and Venables, A. J. (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 Ioannides, Y. M. (2004). The evolution of city size distributions. *In: Handbook of Regional and Urban Economics*, pages 2341–2378.

Gibrat, R. (1931). Les inégalités économiques.

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

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.

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

Jiang, B., Yin, J., and Liu, Q. (2014). Zipf's law for all the natural cities in the world. *Working Paper*.

Johnson, N., Jedwab, R., and Koyama, M. (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 Li, Q. (2013). Uneven landscapes and city size distributions. *Journal of Urban Economics*, 78:19–29.

Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, 51(5):341–352.

Malevergne, Y., Pisarenko, V., and Sornette, D. (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.

Mazmanyan, L., Ohanyan, V., and Treitsch, D. (2008). The lognormal central limit theorem for positive random variables. *Working Paper*.

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 Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Rappaport, J. and Sachs, J. D. (2003). The united states as a coastal nation. *Journal of Economic Growth*, 8(1):5–46.

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

Redding, S. J. (2022). Chapter 3 - trade and geography. In *Handbook of International Economics: International Trade, Volume 5*, volume 5 of *Handbook of International Economics*, pages 147–217. Elsevier.

Rosenthal, S. S. and Strange, W. C. (2004). Chapter 49 - evidence on the nature and sources of agglomeration economies. In Henderson, J. V. and Thisse, J.-F., editors, *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2119–2171. Elsevier.

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

Roy, A. D. (1950). The distribution of earnings and of individual output. *The Economic Journal*, pages 489–505.

Soo, K. T. (2005). Zipf's law for cities: A cross-country investigation. *Regional Science and Urban Economics*, pages 239–263.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46.

Zipf, G. K. (1949). Human behavior and the principle of least effort. *Addison-Wesley Press*.

FOR ONLINE PUBLICATION

A Main Text 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).

A.2 Discussion of Lemma 1 (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 3, which shows the sums of lognormal ($\exp(N(0,1))$, truncated normal (a standard normal trunated 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.

There appears to be a contradiction in the Marlow result, as it implies a sum of positive random variables can be viewed as converging in distribution to either a lognormal or normal distribution.

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 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...

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

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

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

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

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

As $C_v \to 0$ when $n \to \infty$, eq. (31) gives that as $n \to \infty$...

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

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

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

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

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

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

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

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 A1, where selecting alternative truncation points changes the estimated power law coefficient. When including more cities (Panel A) the coefficient rises, and when including few cities 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.

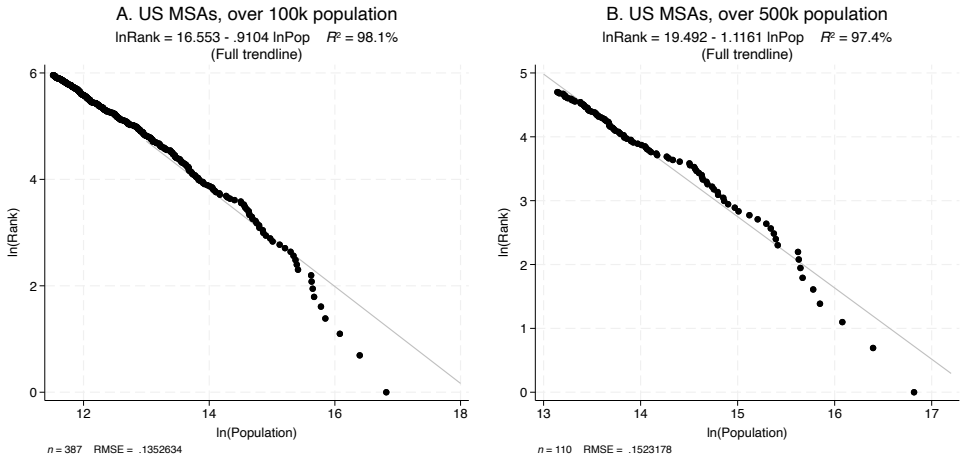


Figure A1: Alternative truncation points for the distribution of U.S. MSAs. Altering the truncation point substantially influences the estimated power law coefficient, as can be seen by contrasting with Figure 1. *Data Source: U.S. Census*

The scale variance of a lognormal distribution can offer 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 esti-

mated power law. We first demonstrate this property of the two distributions via Monte Carlo simulation in Figure A2. We calibrate a lognormal distribution to match a Pareto distribution with shape parameter $\alpha_P = 1$ in the tail.[52] 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.

We can repeat this exercise with data. In Figure A3, we again plot U.S. MSAs to illustrate the lognormality of the U.S. city distribution. We focus on two properties of the lognormal tail that contrast with the Pareto. The lognormal tail will be sensitive to the truncation point (as demonstrated by Eeckhout (2004)), and the largest cities will be smaller than predicted. Choosing two alternative truncation points from that in Figure 1, the estimated power law coefficient falls (when including more small cities) and increases (when including fewer small cities) while the R^2 remains similarly high. These deviations 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 1), it does not appear to be a meaningful feature of the distribution. Next, we consider deviations in the far tail by excluding the largest cities. In both cases, nearly all top quarter cities fall below the

[52]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.

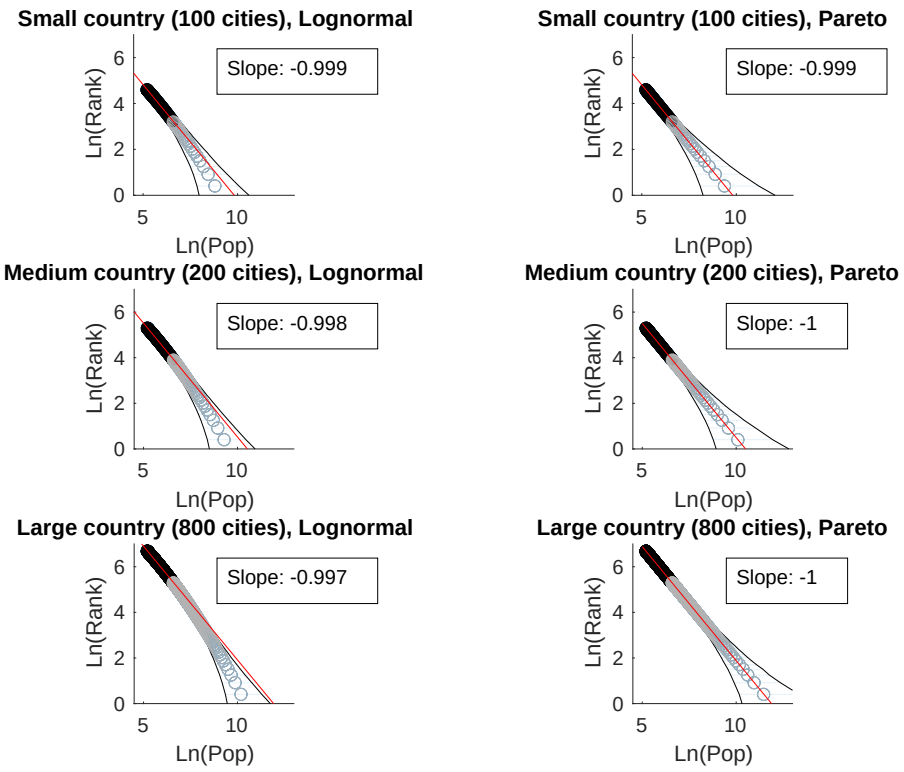


Figure A2: Comparison of Lognormal (left) and Pareto (right) for simulated small, medium, and large "countries." The LN is truncated for cities 2 standard deviations above μ_{LN}, and the Pareto has a minimum value equivalent to this truncation point with shape parameter $\alpha_P = 1$. The slope in 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.

trendline predicted based on the rest of the distribution.[53] The magnitude of the systematic divergence is very large, which is obscured on the log scale. If the Pareto were the true distribution, panel B indicates a cumulative 494 million people missing (in expectation) from the top 25% of U.S. cities, sub-

[53]Of top-quarter MSAs, 94 of 97 MSAs in panel B and 27 of 27 MSAs in panel D are below the respective trendlines in Figure A3

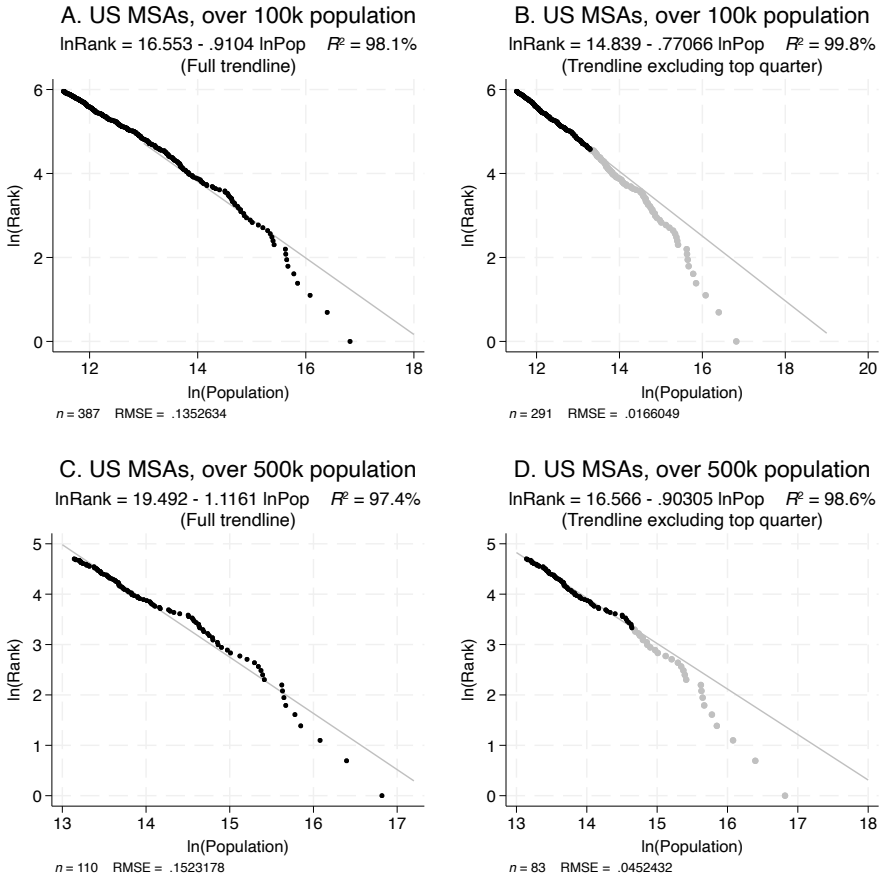


Figure A3: The top panels (A and B) show the 388 U.S. MSAs with a population over 100k in 2020. The bottom panels (C and D) show the 110 MSAs with a population over 500k in 2020. Panels A and C display the trendline for the full distribution and Panels B and D display the trendline excluding the top 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 1. 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*

stantially more than the entire U.S. population, while panel D indicates an absence of 169 million people, roughly half the U.S. population.[54]

[54]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

C Data

C.1 Data Description

In this section, we list the variables we used in Section 2 for our correlation matrices and tables. The data come from the publicly-available data associated with 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

8 million respectively).

D Aggregating Attributes to Fundamentals

D.1 Calculating the Fundamental

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

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

$$\frac{\ln(a_{ik}) - \text{mean}\left(\ln(a_k)\right)}{\text{sd}(\ln(a_k))}$$

where $\text{mean}\left(\ln(a_k)\right)$ is the mean of that attribute across all locations and $\text{sd}\ln(a_k)$ 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 5:

$$\ln(A_i) = \sum_{k \in K} \xi_k \ln(a_{ik})$$

setting $\xi_k = 1,\ \forall k$.

D.2 Summary Statistics

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

Figure A4: Summary statistics for ordered geographic attributes for data points in the contiguous United States. *Data Source: Authors' calculations using data from Henderson et al. (2018)*

D.3 Results

We plot in Figure A5 the empirical PDF of the resulting distribution of productivity fundamentals, calculated according to Equation 5 with $\xi_k = 1,\ \forall k$. 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.

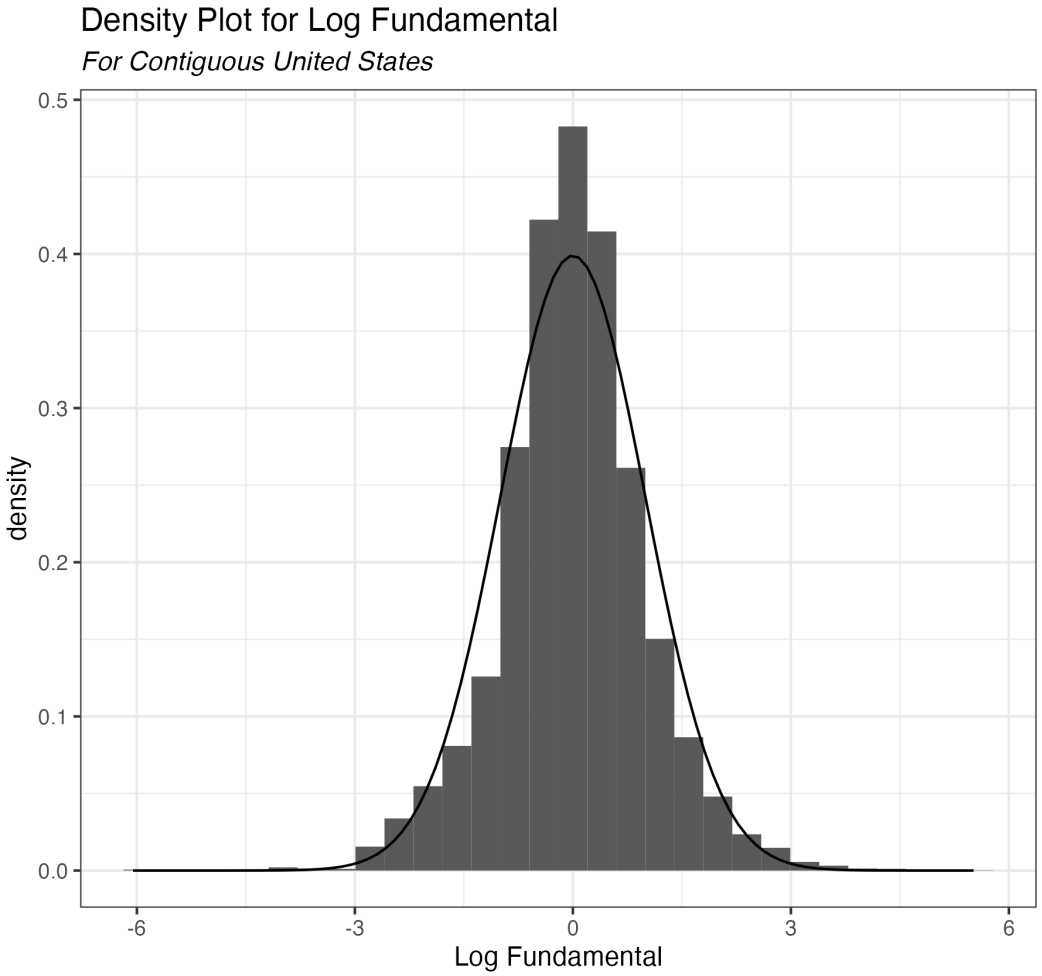


Figure A5: Lognormal distribution of locational fundamentals. All eleven attributes were ordered worst to best in terms of contribution to economic activity, logged, then standardized. The fundamental is calculated as the standardized sum of the standardized, ordered log attributes. The mean and variance are standardized to zero and one and a standard normal curve is overlaid. *Data Source: Authors' calculations using data from Henderson et al. (2018)*

D.4 Recovered fundamentals from Allen and Arkolakis (2014)

We plot the exogenous productivity and amenity values recovered in Allen and Arkolakis (2014). These are at the county-level. County populations are roughly lognormally distributed, as seen in Figure A6. The resulting fundamentals recovered by Allen and Arkolakis (2014), as seen in Figure A7, also appear roughly lognormally distributed. The correlation between A and u is $\approx$0.12, which we use to parameterize our simulations.

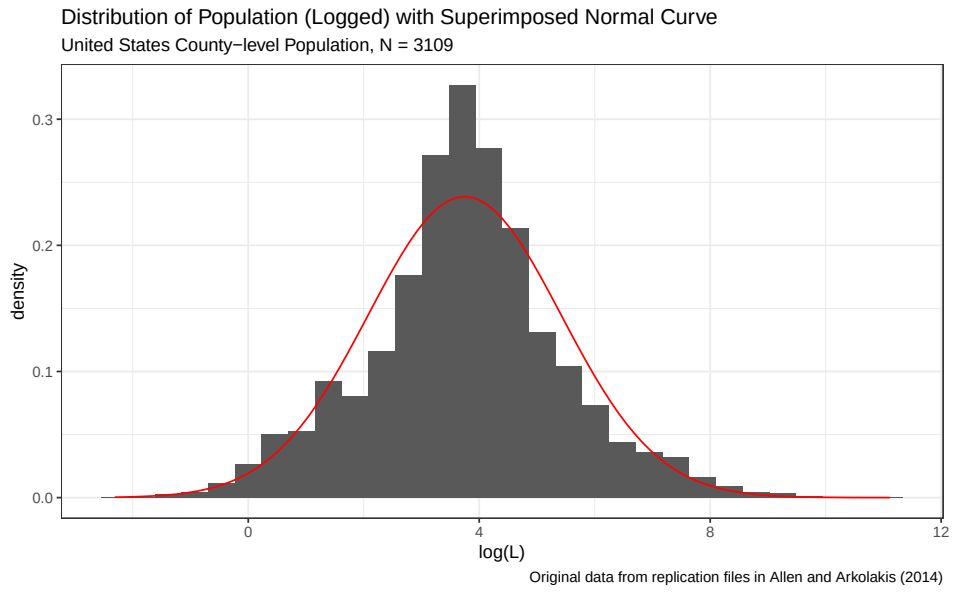


Figure A6: County populations in the United States. *Data Source: online replication package from Allen and Arkolakis (2014).*

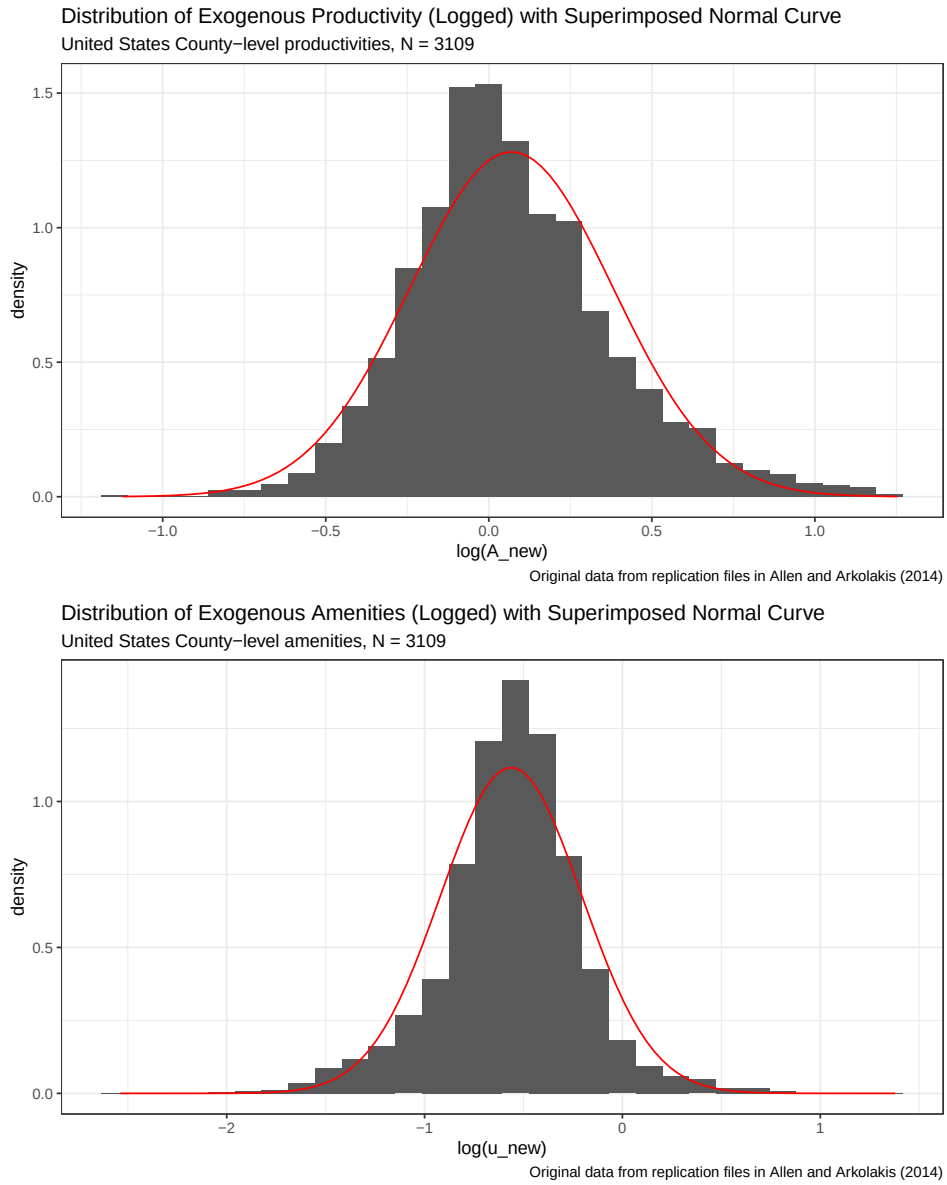


Figure A7: Top panel shows distribution of exogenous productivities at the county-level in the United States; bottom panel for exogenous amenities. *Data Source: online replication package from Allen and Arkolakis (2014).*

D.5 Regression of Economic Activity on Attributes

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

Table A1: Grid-cell radiant lights on attributes, contiguous United States. (Inv) indicates the attribute data was inverted. The R^2 from this regression is comparable to the main regression from Henderson et al. (2018). *Data source: Authors' calculations based on data from Henderson et al. (2018)*

E Correlation Calculations

Cross-Correlations For every attribute type k, we calculate $corr(\boldsymbol{a_i}, \boldsymbol{a_j}),\ \forall i, j \in K,\ j \neq k$, where $\boldsymbol{a_k}$ is a vector for each attribute type comprised of attribute values a_{ik} 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.[55]

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.[56] We then select a random point, called $i^*(d,k)$, 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 k for every centroid. Mathematically, our calculation for the correlation within an attribute type k between our set of centroids and our set of points at distance d takes the form $corr(\boldsymbol{a_k}, \boldsymbol{a_{dk}}),\ \forall k \in K,\ \forall d$, where $\boldsymbol{a_k}$ is a vector of attribute values a_{ik} for attribute type k for all centroid locations i in our sample, and $\boldsymbol{a_{dk}}$ is a vector of all attribute values $a_{i^*(d,k)k}$, the randomly-selected points for each centroid i at distance d for each attribute type k.

[55] 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).

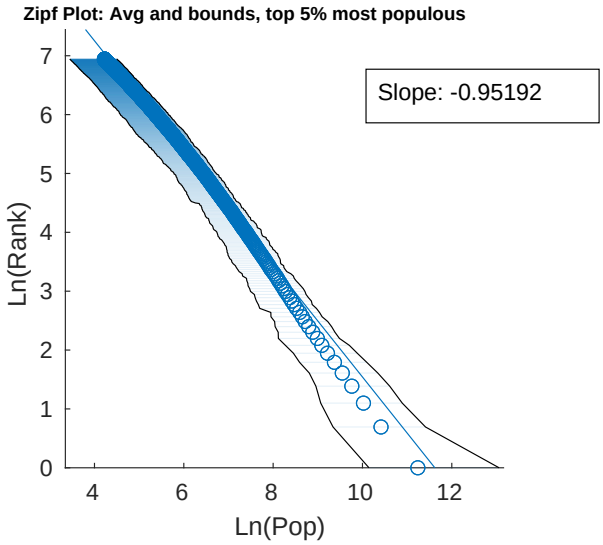
[56] The spatial correlation in attributes between points at distance $d = 100$ miles should be interpreted as "the correlation between a point and a randomly-selected point 90–110 miles away". The buffer is to ensure there are eligible points at roughly the desired distance.

F Additional Simulations

We verify the robustness of the power law coefficient estimate through Monte Carlo simulations. As seen in Figure A8, the average estimated power law coefficient from the (log) rank-size regression across the 1000 simulations is -0.95, with a standard deviation of 0.03. 90% of estimated coefficients are between -0.900 and -1.004. Performing the (log) rank-size regression on the smoothed distribution delivers a slope of -0.95. The parameter values used here are consistent with the literature, as is the truncation point, and estimates come very near Zipf's law.

Even though our model, using standard parameters from the literature, closely approximates the Zipf's law coefficient of -1, we maintain our argument that the estimated coefficient should not be taken as support for cities being Pareto distributed. Changes in scale and the truncation point can influence the estimate, as discussed in Section 1. Nonetheless, it is interesting to note that the estimated power law exponent appears consistent with Zipf's law for typical parameter values in the literature.

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. The sign of change in the estimated exponent associated with changing a parameter, holding the other parameters fixed, is reported in Table 2 in the main text.



100-draw MC	Value
Coef of Avg.	-0.952
Avg. over Coefs	-0.950
Std. Dev.	(0.031)
Min	-1.034
5pct	-1.004
25pct	-0.971
Median	-0.951
75pct	-0.930
95pct	-0.900
Max	-0.875

Figure A8: Monte Carlo average over model output (Left), and statistics over model simulations (Right). The slope on the left represents the slope taken over the average of log(pop) at each rank over 100 simulations, and the bounds contain 95% of the log populations at each rank of the distribution. The table displays statistics over the 100 estimated power law coefficients from the simulations.

Appendix References

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

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, pages 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review*, 99(4):1676–83.

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.

Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Gibrat's law for cities: Uniformly most powerful unbiased test of the pareto against the lognormal. *Physical Review*.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.