

# Zipf's Law in Spatial Equilibrium

Matthew Easton and Patrick W. Farrell\*

August 14, 2023

## Abstract

The appearance of a power law distribution for city populations is a distinctive and recurring feature of human geography. City population distributions are also resilient, recovering rapidly from shocks and consisting of the same cities at similar ranks over long periods, even as the growth of cities often appears to follow a random process. We propose an explanation for these patterns based on the spatial correlation of observable geographic attributes and the interactions between locations across space. We micro-found the aggregation of spatially correlated geographic attributes into terms reflecting the productivity and amenity “fundamentals” of place, demonstrating that these terms will be spatially correlated and lognormally distributed. Given these fundamentals, we show that within many quantitative spatial equilibrium (QSE) models populations will also be lognormally distributed, appearing to follow a power law for highly populated locations (i.e., cities). Standard parameter values for agglomeration benefits, congestion externalities, and transportation costs result in city distributions consistent with Zipf's law, a commonly observed power law for cities. Viewing the distribution of city populations as tail realizations of a lognormal population distribution allows for a better understanding of the observed variation in distributions across countries, time, and levels of geographic aggregations.

---

\*Easton: PhD Candidate in Economics, Columbia University (email: me2713@columbia.edu). Farrell: PhD Candidate in Economics, Columbia University (email: pwf2108@columbia.edu). The authors thank Arslan Ali, Pierre-Philippe Combes, Donald Davis, Thibault Fally, Madeline Hansen, Eshaan Patel, Andrés Rodríguez-Clare, Conor Walsh, David Weinstein, 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 nearly all countries tends to approximate a power law. The existence of this distinctive empirical regularity has been documented in an extensive literature across countries,<sup>1</sup> different periods of human history spanning millennia,<sup>2</sup> and varying definitions of cities.<sup>3</sup> In particular, a specific power law commonly referred to as Zipf's law is often observed at the national level, wherein the largest city within a given country tends to be twice the size of the second largest, and  $n$  times the size of the  $n^{\text{th}}$  largest. This relationship is characterized by a slope of approximately -1 on a plot of log-rank and log-population, as seen in Figure 1 for American cities in 2020. The conventional theoretical explanation for the appearance of a power law is based on the concomitant observation that in many countries cities appear to follow a size-invariant growth process, often called Gibrat's law. Influential work has demonstrated that such a "random" growth process for cities can generate population distributions consistent with the empirical evidence. However, the random growth explanation is neither consistent with the evidence that cities tend to recover following shocks,<sup>4</sup> nor the absence of Gibrat's law when transitioning to new equilibria.<sup>5</sup> Further, a random growth process fails to capture the influence that the observable characteristics of a place have on the attractiveness of locating production or residing there; moreover, these models do not allow for interactions between locations to shape settlement patterns.

We propose an explanation for the appearance of a power law distribution of city populations that highlights key roles for both place and space in shaping the population distribution. Both the observable attributes of each location (first-nature geography) and the interaction between locations across space (second-nature geography) necessarily shape human geography. That these two forces should both matter for the size of cities can be supported by broad theoretical and empirical literatures. We demonstrate that lognormal population distributions, which appear to follow a power law for tail observations (cities), are the equilibrium outcome within a wide class of spatial equilibrium models when the exogenous fundamentals are modeled realistically to exhibit spatial correlation in observable geographic attributes. Our work is the first to demonstrate how first- and second-nature geographies together, within modern quantitative spatial equilibrium (QSE) models, naturally generate the appearance of a power law-like city distribution, Gibrat's law, and other well-documented tendencies of empirical population distributions.

Our approach begins from a consideration of how best to characterize the distribution of populations. We argue that empirical population distributions appear to be best fit by a lognormal distribution in both the body and the tail, as in Eeckhout (2004). For tail values, a lognormal distribution is often indistinguishable from a true power law. We demonstrate the

---

<sup>1</sup>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.

<sup>2</sup>For instance, Davis and Weinstein (2002) demonstrate that a power-law relationship existed in pre-modern Japan, while Barjamovic et al. (2019) find evidence of such a distribution in Bronze Age Anatolia.

<sup>3</sup>Several papers have used nightlights data to define cities rather than administrative borders and identified power law-like city distributions, such as Jiang et al. (2014) and Dingel et al. (2021)

<sup>4</sup>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 Johnson et al. (2019) following pandemics

<sup>5</sup>Desmet and Rappaport (2017) document the failure of Gibrat's law during the settlement of the American West.

link between lognormal distributions and Pareto-like tail behavior, and provide evidence that real world population distribution show characteristics of a lognormal distribution within the distribution of cities.

We then investigate how heterogeneity in observable geographic characteristics influence the suitability of a place for production and habitation. Using data on a wide variety of geographic attributes at a granular level, we document that observable physical attributes are highly spatially correlated, with a greater degree of similarity to nearby locations than to more distant ones. This is an unsurprising and intuitive result, but one that has not been well incorporated into the spatial economics literature.<sup>6</sup> The existence of spatial correlation in attributes is so critical to the study of geography that it has been called the first law of the field, and any spatial model incorporating a realistic geography must incorporate this fact.<sup>7</sup> We also demonstrate that there is only weak correlation across attributes within locations. While some attributes are naturally correlated within a location (growing days and temperature) others have no correlation (e.g., temperature and ruggedness). We argue that a location's suitability for production or habitation should be determined by all of its attributes in a way that exhibits complementarities between attributes. A simple, multiplicative method of aggregating over attributes consistent with these restriction results in productivity and amenity fundamentals of place following a lognormal distribution with spatial correlation.

Next, we show how introducing spatially correlated lognormal fundamentals into a broad class of quantitative spatial models will result in lognormality of the equilibrium population. We adopt a discretized version of the QSE model presented in Allen and Arkolakis (2014), which nests several other spatial models. The equilibrium population condition for each location in these models takes the form of a weighted summation over all locations, incorporating both first-nature and second-nature geography. We demonstrate that, because of the unique property of sums of correlated lognormals, these models deliver a lognormal population distribution in equilibrium.<sup>8</sup> We show that the model is indeed successful at generating both lognormal full population distributions and power law-like city population distributions using numerical simulation. The model gives a role for local productivity spillovers, intra-city congestion externalities, and inter-city transportation costs to influence the observed city size distribution. Varying the parameter values generates changes in the estimated power law exponent consistent with deviations observed in the literature. For standard model parameters across many geographic scales, the observed power law exponent is consistent with Zipf's law (-1), though we view Zipf's law as primarily an artifact of misinterpreting the true lognormal distribution as a power law.

This article touches on several different topics within the spatial economics literature, particularly those on city size distributions and Zipf's law, the importance of geography and trade for explaining settlement patterns, and on spatial equilibrium models. Existing

---

<sup>6</sup>Early models like Krugman (1991) assumed an undifferentiated geography, while more recent QSE models have often drawn an independent exogenous geography as in Redding and Rossi-Hansberg (2017).

<sup>7</sup>The “first law of geography” was proposed by Tobler (1970), who formulated it as “everything is related to everything else, but near things are more related than distant things.”

<sup>8</sup>The result of a lognormal population distribution also holds if fundamentals are lognormally distributed but not spatially correlated (in this case, correlation in population itself is sufficient for the result), but we view the more realistic spatially correlated geography as our primary case.

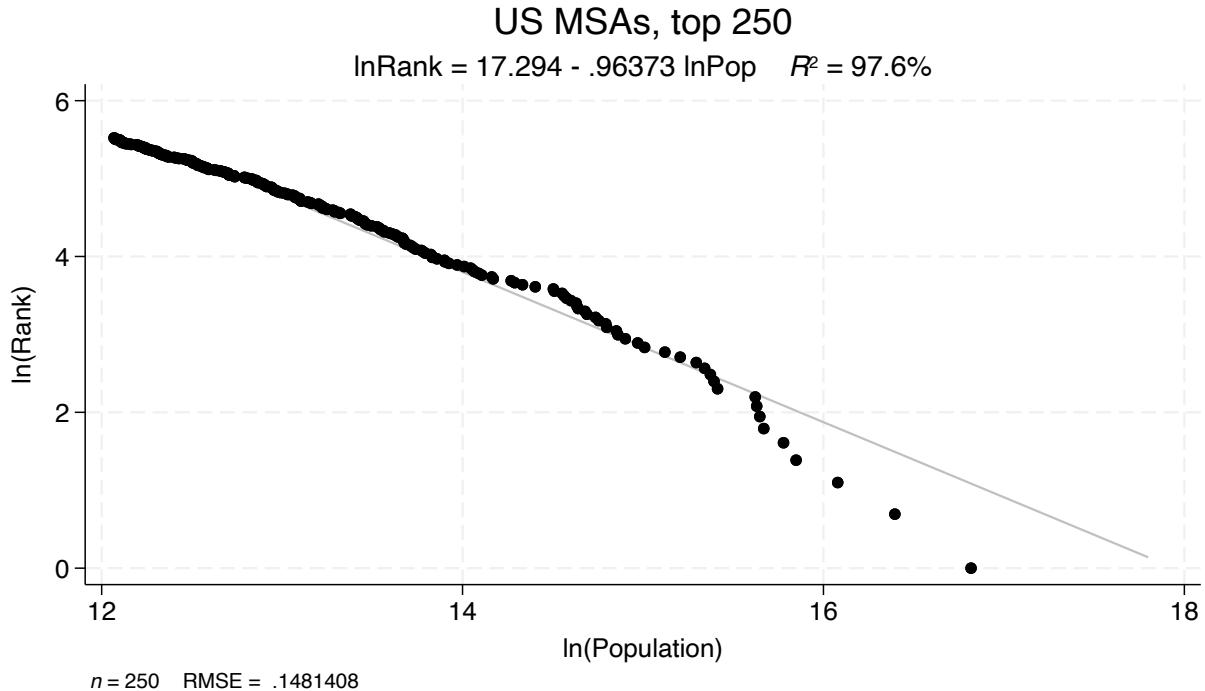


Figure 1: Appearance of a power law for top 250 US metropolitan statistical areas (MSAs) in 2020, with slope near -1.

theoretical explanations for Zipf’s law have often been based on the similarly striking empirical observation that city growth rates are, in many cases, unrelated to city populations. The observation of size-invariant “random growth” for a number of phenomena is referred to in the literature as Gibrat’s law (Gibrat, 1931).<sup>9</sup> Several papers used random growth as the basis for the appearance of a power law-like population distribution, as in Gabaix (1999a), Gabaix (1999b), Blank and Solomon (2000), Rossi-Hansberg and Wright (2007), and Córdoba (2008).<sup>10</sup> In formulating his “law,” Gibrat specifically noted that random growth processes will generate a lognormal size distribution—Eeckhout (2004) emphasizes the apparent lognormality of the full population distribution in a random growth framework, noting that the upper tail of a lognormal distribution can appear to follow a power law. Other work in the random growth tradition has sometimes introduced additional frictions to generate a true power law distribution (a Pareto distribution), rather than a lognormal. Particularly influential examples of the Pareto approach include Gabaix (1999a) and Gabaix (1999b), which introduces a friction in the growth process to prevent cities from becoming too small. This “reflected random process” is also adopted by Córdoba (2008), who extends it to other types of shocks. A related paper including the Pareto as a restrictive special case is Rossi-Hansberg and Wright (2007) who endogenize the entry and exit of cities within a

<sup>9</sup>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.

<sup>10</sup>The review article by Gabaix and Ioannides (2004) emphasizes this approach in the literature and provides several additional citations.

random growth model featuring cities each specialized in one industry.

However, the random growth based theories are inconsistent with characteristics of real-world population distributions. Population distributions tend to be stable, with cities generally occupying approximately the same rank in the population distribution over long periods of time (Madden, 1956) and many modern cities are located on the site of ancient settlements. Even when subject to large negative shocks such as war and disease, cities tend to recover towards their prior position in the population distribution (Davis and Weinstein, 2002; Brakman et al., 2004; Davis and Weinstein, 2008; Johnson et al., 2019).<sup>11</sup> This “reversion” to the prior distribution is incompatible with a random growth process. Further, Desmet and Rappaport (2017) also demonstrate that Gibrat’s law did not hold throughout the historical settlement of the United States. These facts imply that while random growth may hold in equilibrium, it is not the mechanism generating the observed equilibrium.<sup>12</sup> These random growth models both appear to be based on an empirical assumption that does not hold, and (as they are aspatial) fail to incorporate the differences across places and interactions between places that economic geography theory holds to be important.

Other facts about the distribution of cities and their geography offer clues regarding the origin of the power-law distribution of city populations. The largest cities around the world tend to appear in locations that are both good for production and offer quality-of-life benefits (amenity value) to residents, and they tend to be favorably located for trade with other locations. That is, cities tend to have favorable first- and second-nature geography, and the largest cities have the most favorable geographies. For instance, New York City is located on one of the largest natural harbors on Earth and its population relative to Lost Springs, Wyoming, is unlikely to be the result of a purely random process.<sup>13</sup>

Indeed, a literature on the intuitive importance of natural characteristics for explaining settlement patterns has identified a large role for first-nature geographic characteristics (Nordhaus, 2006; Henderson et al., 2018). Related work at a continental or sub-national level has found similarly influential roles for natural characteristics in determining population distributions (Bosker and Buringh, 2017; Alix-Garcia and Sellars, 2020). Studies focused on particular characteristics have identified important roles for individual geographic characteristics in shaping settlement patterns (Rappaport and Sachs, 2003; Nunn and Puga, 2012).

Economic geography theory also emphasizes an important role for second-nature geography in determining the distribution of population and economic activity (Fujita et al., 1999; Allen and Arkolakis, 2014). Indeed, early work in economic geography such as Krugman (1991) primarily focused on second-nature geography and assumed no differentiation in first-nature geography across locations. Empirical work has supported the importance of second-nature geography on economic activity and settlements in addition to first-nature geography (Bosker and Buringh, 2017; Cuberes et al., 2021). The importance of both factors is intuitive. Consider two unpopulated locations with equal high amenity and productiv-

---

<sup>11</sup>Davis and Weinstein (2002) argue in favor of a location fundamentals-based explanation, which we aim to provide a basis for here, as the resilience and persistence of the observed population distribution supports there being properties of places that encourages habitation.

<sup>12</sup>We demonstrate that our model also exhibits Gibrat’s law as a feature of the stable equilibrium.

<sup>13</sup>The 2020 population ratio of the two stands at 8,804,190 : 6. Notable attributes of Lost Springs include its low annual precipitation rates and a coal mine which last operated in the 1930s.

ity values, one of which is located in a desert while the other is in a region composed of other high amenity and productivity values. While the location in the desert may support a charming oasis, the other location will be able to support a larger population given its place in a dense network of productive and livable—and thus populated—places. The recovery of cities from negative shocks, likewise, could be interpreted as the result of the influence of nearby surviving cities rather than solely the persistent fundamentals of the impacted location.

The literature outlined here broadly supports the idea that both first- and second-nature geography must influence the distribution of populations. Our work is the first to generate realistic city size distributions based on both first-nature and second-nature geography. A first-nature only explanation is considered in Gabaix and Ioannides (2004), and a model based on lognormal locational fundamentals (as in our paper) is proposed in Behrens and Robert-Nicoud (2015).<sup>14</sup> These proposals do not allow for second nature geography to play a role in shaping the distribution, and like many random growth models are aspatial despite the spatial nature of economic geography. Brakman et al. (1999) identified a power law-like population distribution within a spatial model through simulation of an economic geography model based on Krugman (1991). They find that a power law-like population distribution emerges without differentiated locations (no differences in productivity, amenities, or position in the network) as a result of interactions across space. They do not identify a mechanism for this result and find that the resulting distributions in their simulations are highly variable. We are able to establish a mechanism that results in realistic city size distributions within a modern spatial model that incorporates both space and a differentiated geography, demonstrate the mechanism behind the result within a broad class of QSE models, and show the stability of this result.

The paper proceeds as follows. Section 1 argues that populations are best described by a lognormal distribution—such a distribution better matches the full distribution, and appears to follow a power law in the tail while naturally generating common deviations from a power law observed real-world population distributions. Section 2 investigates the geographic attributes of place and their correlations across space, establishing facts for integrating a more realistic geography into spatial economics models. We lay out a framework for aggregating the heterogeneous and spatially correlated location attributes into terms that define a location’s exogenous productivity and amenity fundamentals within models of the spatial economy, and argue these will be lognormal and spatially correlated. Section 3 outlines a standard QSE model that nests several canonical spatial models, and demonstrates the lognormality of the resulting population distributions when productivity and amenity fundamentals are spatially correlated and lognormally distributed. This result means that, given a realistic geography, a wide variety of spatial models will exhibit a power law-like distribution for city populations as a result of both first- and second-nature geography. Section 4 uses numerical simulation of the model to demonstrate its ability to capture several results in the empirical literature. Section 5 concludes.

---

<sup>14</sup>Behrens and Robert-Nicoud (2015) found that first-nature, quality-of-life geographic attributes play little role in determining modern-day population distributions by regressing a composite of the attributes on population in the U.S. at the MSA level and ultimately reject a locational fundamentals theory. Yet as Henderson et al. (2018) finds, a larger and different set of geographic and climatic attributes do appear to strongly influence observed population distributions and settlements.

# 1 Seeing a power law in populations

The appearance of a power law-like distribution for city populations is a key puzzle in the spatial economics literature, and its regularity means it can be considered a minimum criterion for a model of the spatial economy (Gabaix, 1999b). Truncating the population distribution to include only the tail distribution of cities, a simple regression of the log rank of a city (in terms of its population) on the log population of the city given by

$$\ln(\text{city rank}) = \beta_0 + \beta_1 \ln(\text{city pop}) + \epsilon \quad (1)$$

will in many countries deliver a very high  $R^2$  (over 0.95), and frequently a slope of -1. An example for the United States can be viewed in Figure 1, where the slope is near -1 for such a regression on the top 250 MSAs with an  $R^2$  of 0.97. This particular slope is characteristic of the appearance of a commonly identified power law called Zipf's law for cities, which can also be formulated as indicating that largest city in a given country is  $n$  times the size of the  $n^{th}$  largest city.

Interpreting the regression as describing the true distribution of city sizes would mean that city populations follow a power law distribution.<sup>15</sup> In particular, given the estimation of  $\beta_1 = -1$ , the particular power law is such that for given size  $S$ , the probability that a city is larger than  $S$  is proportional to  $\frac{1}{S}$ . This can be expressed as a Pareto distribution with shape parameter  $\alpha_P = 1$  (reflecting the estimated  $\beta_1 = -1$ ) and minimum city size  $S_m$ , which gives the aforementioned probability  $P(\text{Size} > S) = \frac{S_m}{S^1}$ .<sup>16</sup>

The Pareto interpretation of the tail city size distribution appears dominant despite its many drawbacks and the strict, non-innocuous assumptions necessary for a power law to hold. In particular, the power law is taken to apply to only a subset of large settlements—this subset can be framed as consisting of either all settlements large enough to be considered cities or only the very largest cities. This requires determining a truncation point in a data series with no obvious truncation point. Early investigations of the city size distribution were limited to only the largest cities or settlements because of the comparative ease of accessing population counts for the largest places, which effectively enforced a truncation point on the distribution.<sup>17</sup> Given more complete data on population distributions, the choice of a truncation point becomes critical and there is no accepted method for determining such a

---

<sup>15</sup>The link between the (log) rank-size plot and the Pareto distribution is established in more detail in Gabaix (2009). Alternatively, plotting the empirical distribution of city populations as a histogram on a logarithmic scale will generate a plot with slope approximately equal to  $\beta_1 - 1$ , where  $\beta_1$  is the estimated coefficient from the (log) rank-size plot (the kernel selected for the empirical distribution will impact the estimated slope). Exponentiation of the implied relationship  $\ln(p(x)) = \beta_0 + \beta_1 \ln(\text{city pop})$  will directly yield the PDF of a Pareto distribution with shape parameter  $\alpha_P$  approximately equal to  $\beta_1$  Newman (2005). While this latter presentation is more direct, the (log) rank-size plot and regression are far more prevalent in the literature.

<sup>16</sup>This is the counter-cumulative distribution function of a Pareto distribution. The Pareto distribution has a CDF given by  $P(X < s) = 1 - \frac{x_m}{s^{\alpha}}$ , where  $x_m$  is the minimum value taken by draws from the distribution.

<sup>17</sup>This is true of early work documenting this relationship, such as Auerbach (1913) (while Auerbach has data on even small places, a table in his paper includes just the 94 largest; see the recent translation in Auerbach and Ciccone (2023)) and Zipf (1949). Even the more recent investigation of Zipf's law in Krugman (1996), for instance, included just the top 135 cities because the *Statistical Abstract of the United States* included only those cities (Eckhout, 2004).

point. Most researchers rely on a visual test of the data to determine a cutoff (Gabaix, 2009). The implication of a truncation point requires there be some minimal “city” and corresponding population below which the population dynamics behave differently, as the full distribution does not appear Pareto. A rigid cutoff between “cities” and other settlements is arbitrary and unrealistic—further, the many periods and regions in which Zipf’s law appears to hold would necessitate the assumption of several different cutoffs differing across time and varying across different national contexts.

Random growth models generating a Pareto distribution argue that their ability to explain an exact Zipf’s law slope of -1 on the (log) rank-size plot is a strength of this distributional approach. However, it is not clear that this slope is a meaningful feature of the data, and it appears to be in part an artifact of truncating a distribution without any obvious truncation point. Even allowing for this truncation, these models still must rely on further strict assumptions to achieve the Zipf’s law slope. Gabaix (1999b) obtains convergence to a -1 slope, for instance, in the case where the lower bound of standardized city sizes tends to zero.<sup>18</sup> This assumption is in tension with the need to truncate the data to define a minimum city size that excludes a substantial portion of the population. The Pareto case in Rossi-Hansberg and Wright (2007) is similarly restrictive.<sup>19</sup>

Instead of being a true power law, the city population distribution may instead be a feature of a different distribution which behaves like a Pareto for tail observations. With the availability of data capturing the full population distribution, work by Eeckhout (2004) demonstrates that the full distribution appears to be lognormal. The tail of a lognormal distribution can appear similar to a Pareto distribution, which can be seen by considering the PDF of a lognormal...

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

After some algebra (given the appendix), this can be rewritten as...

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

Letting the constant term  $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) = \Gamma$ , the lognormal PDF can be written...

$$f(x) = \Gamma x^{-\alpha(x)-1} \quad , \text{ where } \alpha(x) = \frac{\ln(x) - 2\mu}{2\sigma^2} \quad (3)$$

Contrast this with the PDF of a Pareto distribution with truncation point  $x_m$  defined such that  $\alpha_P x_m^{\alpha_P} = \Gamma$ ...

$$j(x) = \Gamma x^{-\alpha_P-1} \quad (4)$$

---

<sup>18</sup>This restriction is challenged in Blank and Solomon (2000), who argue the conditions in Gabaix (1999b) do not necessarily result in convergence to a Pareto with shape parameter  $\alpha_P = 1$ .

<sup>19</sup>The two cases in which exact Gibrat’s law and Zipf’s law hold exactly in Rossi-Hansberg and Wright (2007) are: (i) there is no physical capital and productivity shocks are permanent, or (ii) production is linear in physical capital, there is no human capital, depreciation is 100%, and productivity shocks are temporary.

The lognormal PDF in equation 3 is similar to that of the Pareto PDF in equation 4, but with a scale-varying shape parameter.<sup>20</sup>

The scale-varying shape parameter of the lognormal tail, when written as in equation 3, demonstrates the key phenomena highlighted by Eeckhout (2004) regarding the sensitivity of the estimated power law coefficient to the truncation point—because the lognormal distribution is not scale-invariant like the Pareto, changing the truncation point changes the estimated power law. The sensitivity to the truncation point, the lack of a reliable rule for truncating the distribution, and the property that observations drawn from many lognormal distributions will appear to take a slope of -1 over some large portion of the distribution at some scale (as the exponent in equation 3 diverges smoothly), suggest that the -1 exponent often estimated in the literature is unlikely to be a meaningful feature of the data. Excluding portions of the distribution (small cities) for which  $\alpha(x)$  is too low, but preserving a sufficient amount of these observations to counteract the portion where  $\alpha(x)$  is too high (very large cities) will be sufficient to estimate a power law coefficient of -1 in most contexts if the true population is lognormal.

Further, beyond the drawbacks of the Pareto interpretation, the lognormal has attractive properties of its own in fitting the data. The tail of a lognormal distribution is not as thick as that of a Pareto distribution, which means the likelihood of very large realizations is lower when the true distribution is lognormal. In the context of city sizes, the divergence of the lognormal from the Pareto will generate fewer very large cities than would be expected. This tendency appears to match the global city distribution (Rossi-Hansberg and Wright, 2007).<sup>21</sup> Proponents of the Pareto interpretation have attempted to integrate these divergences with the interpretation of a true power law by arguing that the forces acting on small cities are different from those acting on large cities, thus generating different power laws for different cities. A lognormal distribution naturally exhibits this deviation, without the need to treat subsets of the distribution differently.

In particular, 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 estimated power law. We first demonstrate this property of the two distributions via Monte Carlo simulation in Figure 2. We calibrate a lognormal distribution to match a Pareto distribution with shape parameter  $\alpha_P = 1$  in the tail.<sup>22</sup> 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

---

<sup>20</sup>The lognormal distribution always is scale-varying, but for large  $\sigma^2$  the exponent can appear stable (Malevergne et al., 2011)

<sup>21</sup>While some work has asserted that large cities tend to be too large (Gabaix, 1999b) and argued for the unique political economy of capital cities as an explanation, empirical work tends to find that large cities (at least in large countries) appear to be substantially smaller than expected based on the rest of the distribution.

<sup>22</sup>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 (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.

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 3, 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 trendline predicted based on the rest of the distribution.<sup>23</sup> 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, substantially more than the entire U.S. population, while panel D indicates a comparatively modest absence of 169 million people, roughly half the U.S. population. Repeating this exercise with other large countries (India, China, and Brazil) using standardized city definitions from Dingel et al. (2021) indicates similarly large divergences in the tail, all in the expected direction (cumulative absences of 135 million, 53 million, and 8 million respectively). These results are presented in the appendix.

The appearance of a power law-like city population distribution is most 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). In the following section, we begin an investigation of a likely source of the observed variance of in human populations in space—the varied geography of Earth.

## 2 Observable Attributes, Spatial Correlation, and Locational Fundamentals

Earth has a diverse geography. Climatic conditions, topography, soil quality and type, and the incidence of extreme natural disasters vary greatly around the world, from the Niger Delta to Siberia, the Australian Outback to the American Midwest. Human settlement patterns similarly vary greatly across space. While the Niger Delta and American Midwest differ tremendously across many observable geographic attributes, both are home to large cities. Siberia and the Outback are both lightly populated despite their own evident differences. There is clear evidence for observable geographic attributes, alone and in combination, playing a role in shaping human habitation patterns.

---

<sup>23</sup>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 3

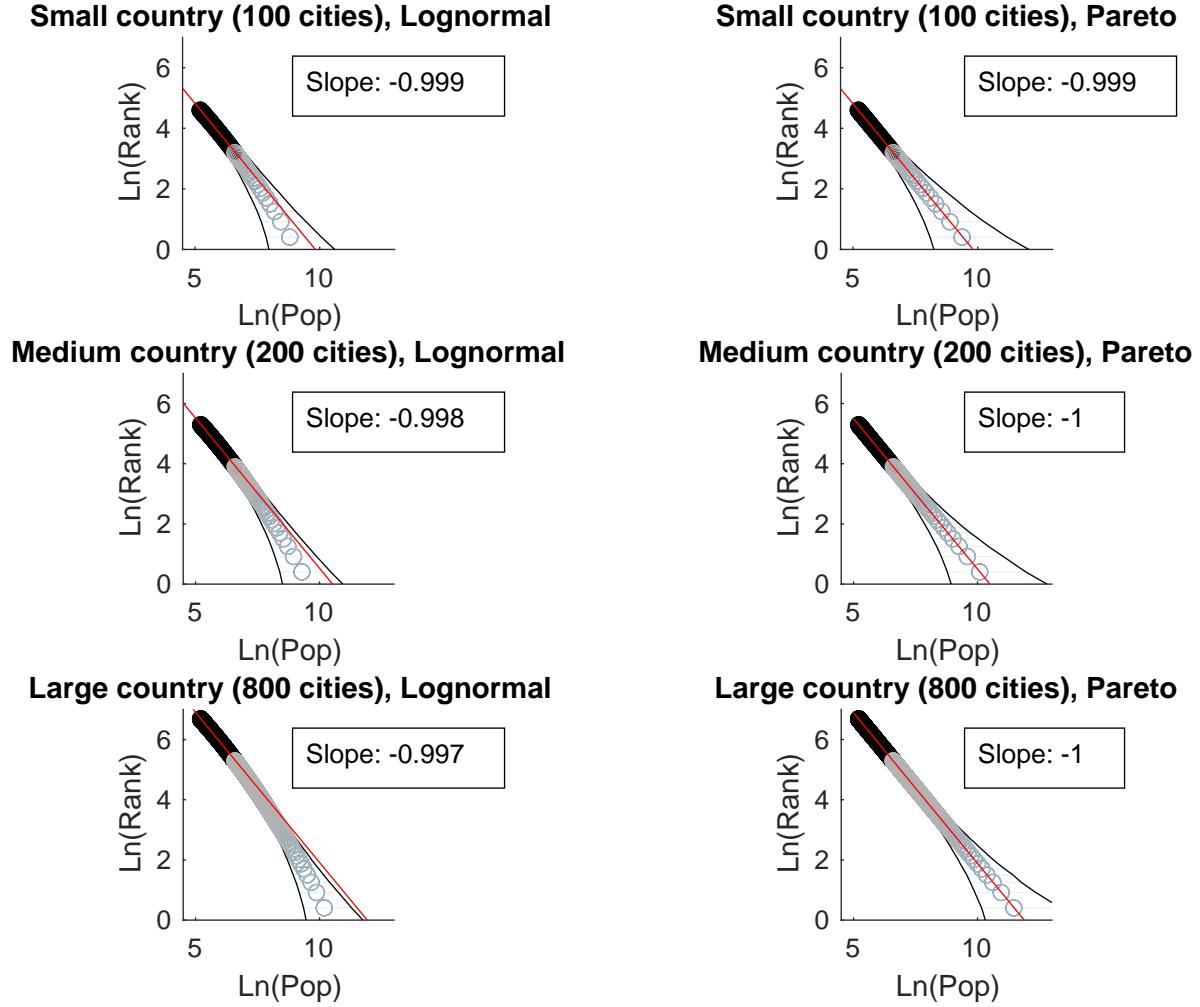


Figure 2: Comparison of Lognormal (left) and Pareto (right) for simulated small, medium, and large “countries.” The LN is truncated for cities 2 standard deviations above  $\mu_{LN}$ , and the Pareto has a minimum value equivalent to this truncation point with shape parameter  $\alpha_P = 1$ . The slope 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 and 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 drawn from a LN distribution the large cities tend to fall below the trendline (with trend above the 95% band) while the Pareto distribution does not diverge.

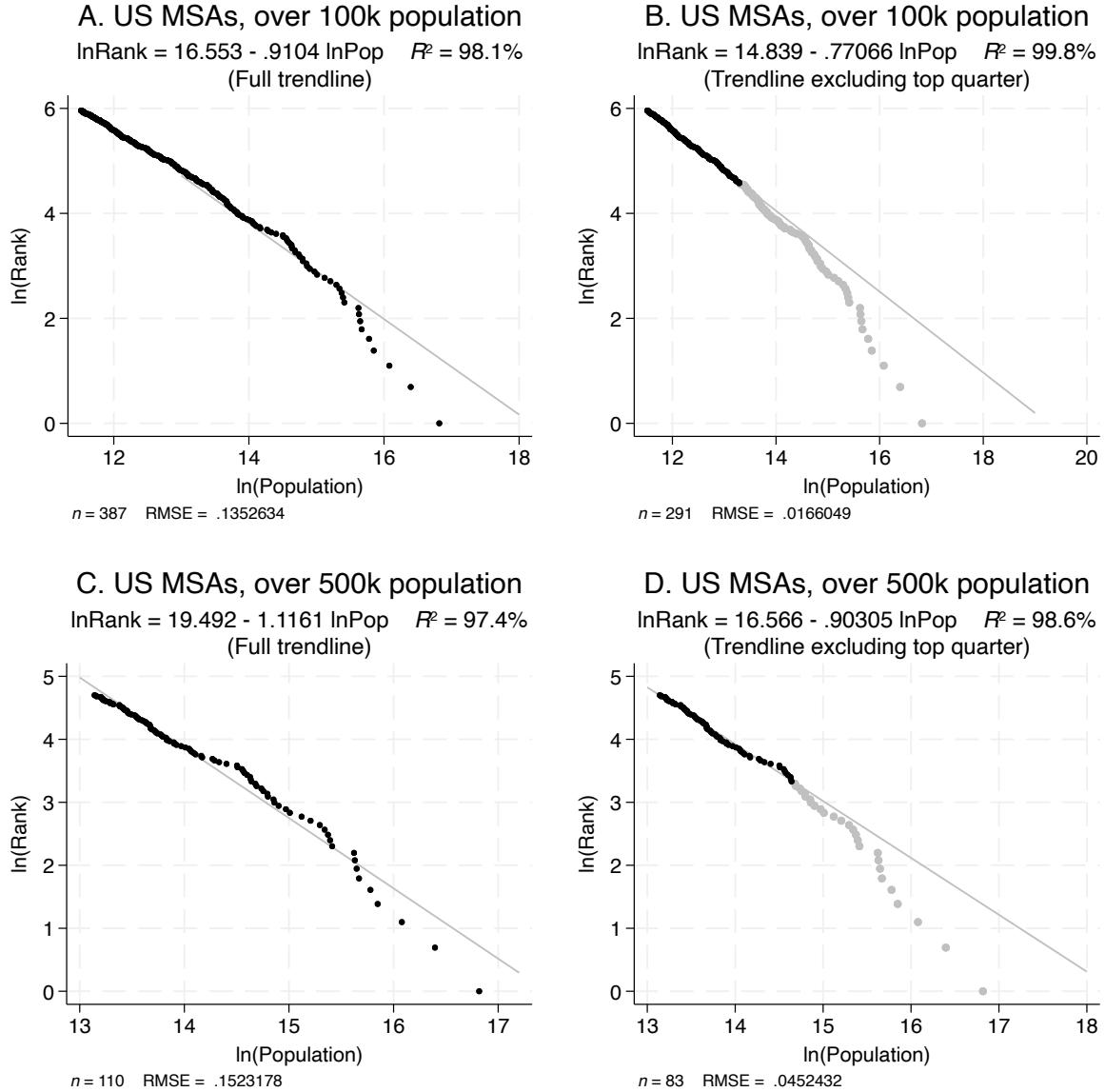


Figure 3: The top panels (A and B) show the 388 U.S. MSAs with a population over 100k in 2020. The bottom panels (C and D) show the 110 MSAs with a population over 500k in 2020. Panels A and C display the trendline for the full distribution and Panels B and D display the trendline excluding the top 25% 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 US population distribution being lognormal.

Within standard spatial models, locations are assumed to have exogenous “productivity” or “amenity” values, which we will refer to as *fundamentals* throughout this paper. Such fundamental values are unobservable and are necessarily an abstraction, but do capture something real about the nature of place. Some places are better for production or more attractive to live in than others—we neither observe the precise productivity benefit or penalty associated with being in one location compared to another, nor is there an measurable “niceness” of place, but these differences do exist. While unobservable, these fundamentals must be a function of the observable geographic characteristics of the world.

While it is reasonable that observable geographic attributes influence production and amenity fundamentals, it is less clear how the attributes we can see impact the things we cannot. The high degree of variation in many attributes across areas of high population strongly suggests that no particular observable attribute alone is a sufficient proxy for these fundamentals. There are highly populous locations which have steep topographies (La Paz) and those with flat topographies (Houston); those which are along the coast (Rio de Janeiro) and those which are far from any coast (Ulaanbataar); and those which have consistent rain (Dublin) and those which have little rain (Phoenix). Despite such heterogeneity across individual attributes these cities must all be attractive for human habitation based on the productivity or amenity benefits their locations offer relative to other steep, flat, coastal, inland, rainy, or arid locations where few people live.

While no single observable attribute determines the productivity or amenity fundamental of a place, places that are highly populous must be favorable in some way across multiple geographic attributes that we can observe. 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 taken 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, constructing complex irrigation canals to support an estimated population of as many as 100,000 people in the Phoenix basin (Doyel, 2001).<sup>24</sup> That is, even if a particular attribute of a certain place is not favorable, other attributes may be—populous places have “something going for them,” but that *something* may be different across different locations.

We continue with a deeper investigation of the characteristics of a realistic geography. We demonstrate that geographic attributes are strongly spatially correlated, but that there is only limited correlation across attributes within a given location. We then discuss how to model the contribution of these attributes to the productivity and amenity fundamental values within spatial models.

---

<sup>24</sup>A brief discussion of the Hohokam people can be found from the [Arizona Museum of Natural History](#). The persistence of human habitation in the region is more evidence for the importance of locational fundamentals for explaining settlement patterns.

## 2.1 Data

We begin by offering evidence on the distribution of geographic attributes. We use gridded geographic data collected by Henderson et al. (2018).<sup>25</sup> These data include a wide variety of geographic attributes, of which we use eleven: 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.<sup>26</sup> Our sample for the U.S. is roughly 18,000 grid cells; including Mexico and Canada, our sample expands to over 47,000 cells.

## 2.2 Attributes and space

We first calculate cross-correlations between our eleven attributes within all grid points in the U.S., Canada, and Mexico. This exercise tells us how correlated each attribute is with all other attributes in each location across all locations, giving an indication of how dependent realizations of geographic attributes may be on one another. High correlations in absolute value would indicate strong dependence while low correlations in absolute value would indicate weak dependence between attributes.

Results are provided in Figure 4. The figure was constructed by ordering the absolute value of each attribute's cross-correlations with other attributes from most correlated (index 1 on the x-axis) to least correlated (index 10 on the x-axis). Within each index, we identify the maximum, median, and minimum correlations across the attributes. In the figure, the set of medians across indices is represented by the black solid line, and the set of maxima and minima across indices are represented by the blue dashed lines above and below with the area in between shaded light blue. The figure tells us that while most attributes in our data set are highly correlated with at least one other attribute—as indicated by a median correlation of 0.6 and minimum correlation of 0.3 for first index attribute—the majority of attributes are only weakly correlated with each other. The median passes below a value of 0.25 by the fifth-most correlated indexed attribute, in conjunction with the minima and maxima trending towards 0 as the index increases, demonstrates that a majority of the attributes exhibit only weak dependence with each other.<sup>27</sup>

Next, we demonstrate that while attributes within each location show little dependence, there is strong correlation within each attribute across space. We construct rings at varying distances with 20-mile buffers (i.e. +/- 10 miles for a given distance) around each grid point (centroid) in the contiguous United States. These rings identify all grid points in the United States, Canada, and Mexico at a given buffered distance around each centroid. The buffer is to ensure there are eligible points at roughly the desired distance, for all grid points in the U.S.<sup>28</sup> We then select a random point from within each ring for every centroid, in order to

---

<sup>25</sup>In the appendix, we provide descriptions of all variables used. We also perform similar analyses in the appendix with a different data source for geographic attributes used in Behrens and Robert-Nicoud (2015).

<sup>26</sup>The data is grid cells at the quarter-degree latitude and longitude size; all data which are averages come from data collected between 1960 and 1990; and all variables were logged, with variables having measured values at or below zero given an affine transformation to shift the minimum value to be positive.

<sup>27</sup>In the appendix, we include a table including all of the correlations and a plot of all the individual correlations (rather than only the minimum, median, and maximum values).

<sup>28</sup>For clarity, the spatial correlation in attributes between points 100 miles away should be interpreted as

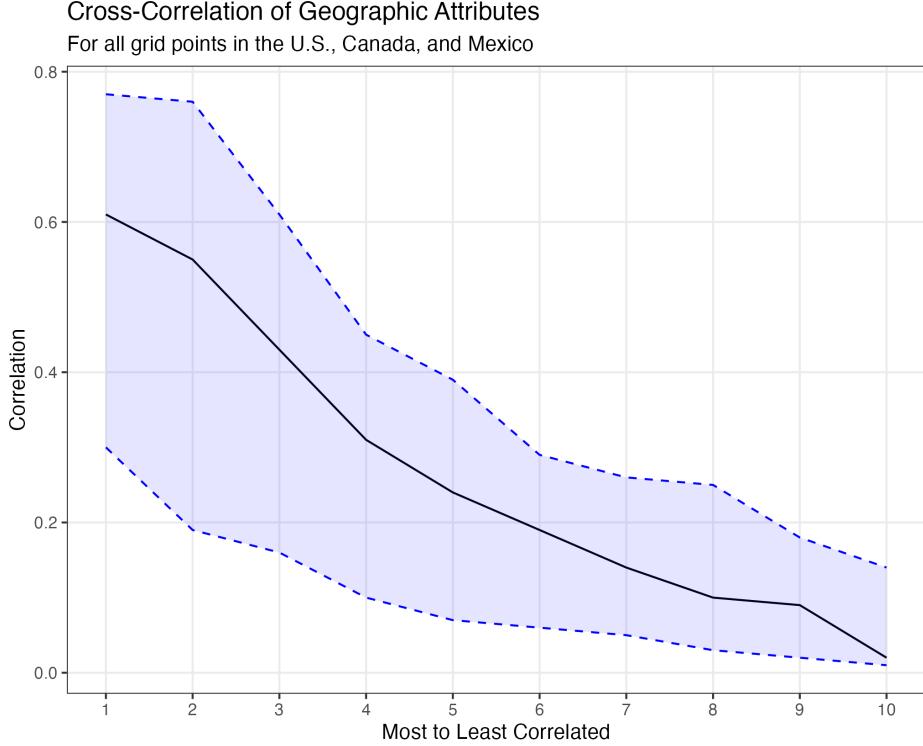


Figure 4: Structure of cross-correlations for geographic attributes in the U.S., Canada, and Mexico. For all eleven geographic attributes, the absolute values of the cross-correlations were calculated and ranked in descending order from 1 to 10 (1st being most correlated, 10th being least; we omit self-correlation). The solid black line represents the median cross-correlation across all attributes at each index; the blue dashed lines represent the maximum and minimum cross-correlations across all attributes at each index.

construct our sets of points to calculate the correlations, drawing a randomly-selected point for each attribute. The results are in Figure 5.<sup>29</sup>

Figure 5 illustrates two main points. First, spatial correlation in attributes is high over short distances and does not immediately decline to zero—the median correlation at 50 miles is over 0.8, and at 250 miles it is still over 0.5. Second, while spatial correlations are strong for nearby points, spatial correlation does fall to near 0 at great distances.<sup>30</sup> In summary, these results suggest clear spatial correlation of geographic attributes.

These results are intuitive. Further they demonstrate what has been called the “first law of geography”: everything is related to everything else, but near things are more related than distant things (Tobler, 1970). If there was no spatial correlation and a location’s attributes were drawn independently, traveling even short distances would result in exposure to a dizzying array of conditions. Take New York City—imagine a trip around the neigh-

<sup>29</sup>the correlation between a point and a randomly-selected point 90 to 110 miles away”.

<sup>30</sup>An illustration of the computational process behind correlation calculations is provided in the appendix. The illustration in the appendix uses data and cells at the U.S. county level from a separate data source; the general process applied to grid points, however, was the same.

<sup>30</sup>In the appendix, we provide a scatter plot version of this figure to further illustrate this point.

Spatial Correlation of Geographic Attributes  
For all grid points in the contiguous United States

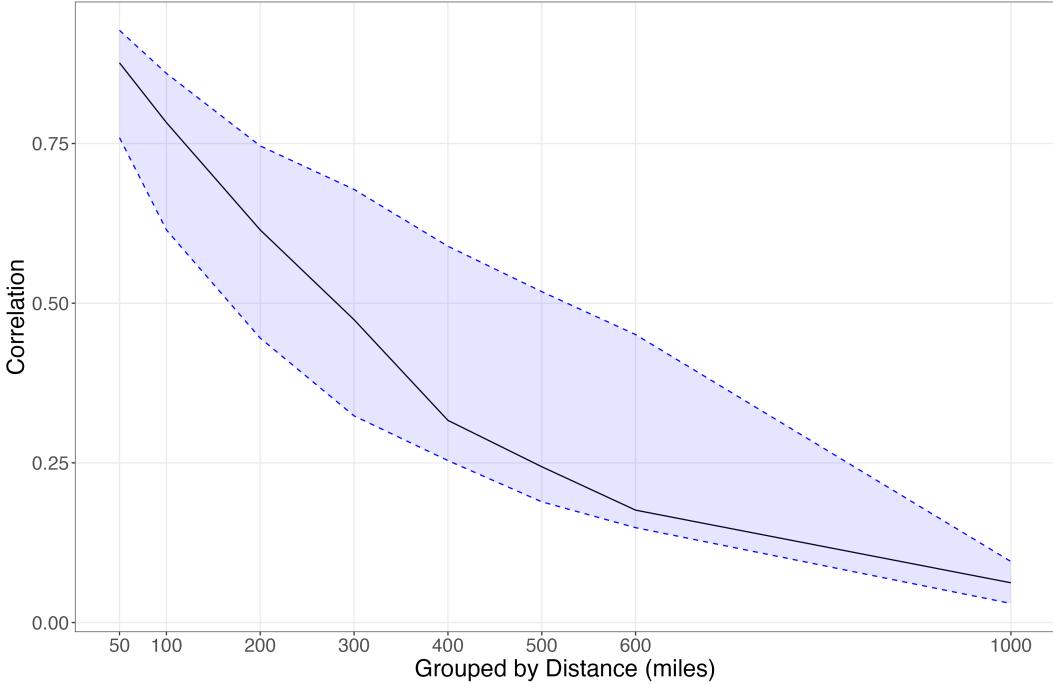


Figure 5: Spatial Correlation Structure of U.S. Geographic Attributes. Values shown are for correlations between grid points in the United States and randomly-selected grid points in the United States, Canada, and Mexico. Absolute values of correlations were used. The solid black line represents the median correlation across various distances across all attributes; the blue dashed lines represent the 25th and 75th percentiles bands at each distance.

borhoods of Manhattan in a world without any spatial correlation in local attributes. One might appreciate a parka while traversing the tundra that could characterize SoHo, quickly abandoned as one enters the arid, hot desert of Greenwich Village, before switching to waders to traverse the swamps of Chelsea. While the scale of this example is quite small, and within the framework of our paper we will tend to think of locations as larger areas than New York City neighborhoods, it should be clear that places in the American mid-Atlantic region (or any region of Earth, or any known planet) are substantially similar across a wide variety of attributes precisely as the data suggests.

### 2.3 Aggregating attributes

Given empirical support for assuming spatial correlation of geographic attributes across locations and only weak dependence across attributes in a given location, we will formalize how to incorporate these features into spatial economy models. Models in regional, spatial, and urban economics assume that locations have exogenous productivity and amenity values rather than define locations directly in terms of their geographic attributes. A *locational productivity fundamental* is an unobservable term that defines the productivity of a given location, and a *locational amenity fundamental* is an unobservable term that can be thought

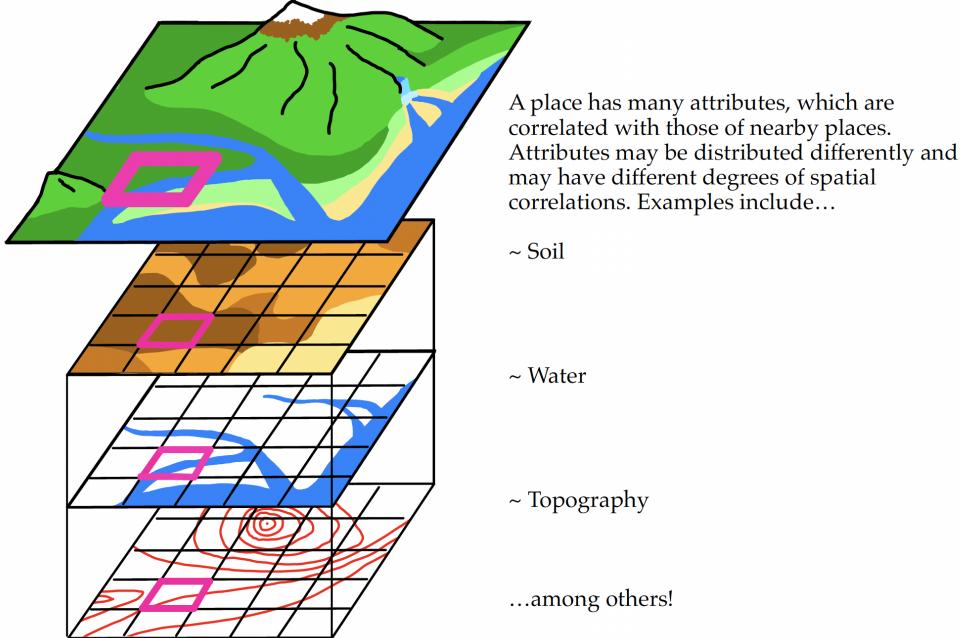


Figure 6: Attributes and their spatial correlation.

of as directly influencing the quality of life of living in a given location.

We now provide a microfoundation for the distribution of these fundamentals, based on those attributes of place we can observe. Observable locational *attributes* must impact the productivity and general attractiveness of a location. Such attributes for a productive and populated location could be fertile soil, regular and mild weather patterns, and favorable topography improving the suitability of a place for agricultural or manufacturing production (See Figure 6 for an illustration). Importantly, these observable attributes are also highly spatially correlated, as shown in the previous section.

Attributes  $a_{ik}$  are associated with a location  $i$  with the type of attribute indexed by  $k \in K$ . There are a large number of attributes. Each attribute may differ in its degree of spatial correlation and in the distribution from which it is drawn. We allow for some attributes to be correlated with each other—availability of water and quality of soil tend to be correlated, for example—while other attributes are uncorrelated. We thus assume that the attributes associated with each location  $i$  are only *weakly* dependent. This is consistent with the empirical evidence.

We assume all such attributes are drawn from distributions with finite variances. We also assume all attributes are strictly positive in value—no place has less than zero access to water, or completely zero access—and for any  $k$  a higher values of  $a_{ik}$  reflects a *better* draw.<sup>31</sup> That is, viewing the locational productivity fundamental for a location  $i$ ,<sup>32</sup> denoted  $A_i$ , as a function of its many attributes  $a_{ik}$ ...

<sup>31</sup>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 taking values from worst to best.

<sup>32</sup>We will define productivity and amenity fundamentals in a similar way, and so will focus first on the productivity fundamental.

$$A_i = F(a_{i1}, a_{i2}, \dots, a_{iK})$$

...a location's overall fundamental  $A_i$  should be increasing in each  $a_{ik}$ :  $\frac{\partial F}{\partial a_{ik}} > 0$  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:  $\frac{\partial^2 F(\cdot)}{\partial a_{ij} \partial a_{ig}} > 0$ , for  $j, g \in K, j \neq g$ . Indeed, complementarities should exist across all arbitrary combinations of attributes.

A functional form for an aggregating function over location attributes that delivers these properties could be multiplicative over all attributes, such as a Cobb-Douglas aggregator. An attractive property of using a Cobb-Douglas aggregator is that we can interpret the powers as weights, allowing us to capture the relative importance of attributes changing over time (to capture structural transformation or changing production technologies, potentially). Allowing the weight at time  $t$  for attribute  $k$  to be denoted  $\xi_{kt}$ ...

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

For simplicity, suppress the  $t$  subscript as we will not be considering change over time in this paper. Taking the natural log yields the following expression for the productivity fundamental:

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

The amenity fundamental is defined similarly, but we allow for different weights to capture that the important attributes for determining quality of life may differ from those influencing productivity. The log of the amenity fundamental, which has different weights given by  $\iota_k$ , is...

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

Given our assumptions that attributes have finite second moments, strictly positive attribute values, and only weak dependence between attributes within each location (demonstrated in Figure 4), applying the central limit theorem to the above expressions will result in both  $\ln(A_i)$  and  $\ln(U_i)$  being normally distributed. Exponentiating,  $A_i$  and  $U_i$  will be lognormally distributed. We also note that the spatial correlation across attributes (demonstrated in Figure 5) will be inherited by the resulting fundamental terms.

As a simple demonstration that fundamentals, constructed as described, will exhibit lognormality, we again use the gridded attributes data from Henderson et al. (2018). We transform the variables so that higher values are associated with larger populations as in the regression results of that paper and scale all logged variables to have a variance of 1. As can be seen in Figure 7, the log of the aggregated values is closely fit by a normal distribution, and so the fundamental value will be lognormal. For more details on the construction of the empirical log “fundamental,” please see the appendix.

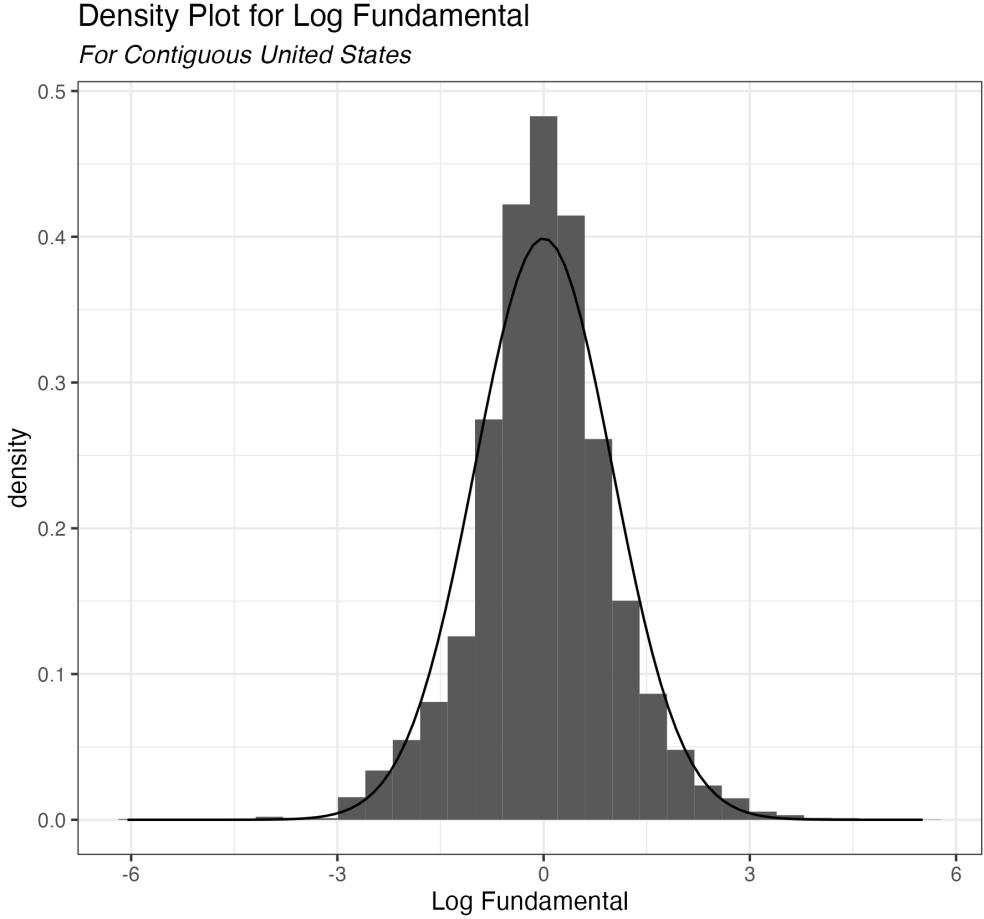


Figure 7: Lognormal distribution of locational fundamentals. Includes eleven attributes from Henderson et al. (2018). All variables were ordered worst to best in terms of contribution to economic activity à la Henderson et al. (2018), logged, then standardized. The fundamental is calculated as the standardized sum of the standardized, ordered attributes. The mean and variance are standardized to zero and one and a standard normal curve is overlaid.

The emergence of a lognormal distribution for fundamentals is unsurprising. The lognormal distribution appears frequently in nature because random variables which are the result of a multiplicative process, rather additive, will tend to be lognormally distributed (Roy, 1950). Consider, for instance, a random variable that might be one of the attributes: annual rainfall, measured in inches. Such a measure of total rainfall will depend on the frequency (number of days), duration (minutes per day), and intensity (amount per minute) of rainfall multiplicatively, and should follow a lognormal distribution regardless of the distribution of those driving factors.<sup>33</sup> A number of attributes—often quite unexpected ones, moreover—tend to lognormality, including the concentration of many elements (gold, cobalt, uranium), the latency period of diseases (chicken pox, food poisoning), air pollution, the abundance of species in an environment, incomes, and the number of ice crystals present in ice cream

---

<sup>33</sup>Provided frequency, intensity, and duration have sufficiently weak correlation. Indeed, there is strong evidence that total rainfall measures are lognormally distributed (Kedem and Chiou, 1987).

(Limpert et al., 2001). Even absent a microfoundation as provided here, that the fundamentals would be lognormal would be unsurprising given the distribution’s ubiquity in nature.

### 3 QSE With Correlated Lognormal Fundamentals

In this section, we describe a quantitative spatial equilibrium model with spatially correlated and lognormally distributed productivity and amenity fundamentals. The model presented here is a discretized version of the model in Allen and Arkolakis (2014), which nests several other recent QSE models such as Redding (2016) and Redding and Rossi-Hansberg (2017). These models are based on the two location model presented in Helpman (1998), but generalized to an arbitrary number of locations. We characterize the equilibrium population distribution within this model, showing that populations will be lognormally distributed if fundamentals are lognormally distributed and spatially correlated.

It is important to note that this is not simply due to populations reflecting the lognormal first-nature geography in each location. As the equilibrium condition in these models takes the form of a weighted summation over all locations, second-nature geography plays a key role. The result relies on the unique properties of sums of correlated lognormal random variables which we describe in more detail in the proof below.

The world consists of  $N$  discrete locations with distances defined by a Euclidean distance metric. Locations are indexed by  $n \in N$ . Trade between locations is costly. The trade cost function is given by  $\tau : N \times N \rightarrow [1, \infty)$ , where  $\tau_{n,n} = 1$  and  $\tau_{n,i} > 1$  for  $n \neq i$ .

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. These attributes are spatially correlated across locations and only weakly dependent within a location. Aggregating over the attributes results in spatially correlated, lognormally distributed exogenous productivity and amenity fundamentals as motivated in the prior section. A place’s effective productivity and amenity value may also be impacted by negative or positive population externalities. In particular, amenities can suffer from overcrowding while production can benefit from local spillovers. We define these “composite fundamentals” as

$$\tilde{U}_i = U_i L_i^\beta \quad (5)$$

$$\tilde{A}_i = A_i L_i^\alpha \quad (6)$$

where the typical case will consist of  $\beta < 0$  and  $\alpha > 0$ .

Geography within this model is represented by the set of functions defining the locational fundamentals,  $\tilde{A}$  and  $\tilde{U}$ , along with the function  $\tau$  defining the spatial relationship between points 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  $\tau_{n,i}$  are similarly bounded above and below by strictly positive numbers between all locations. We will use the term “regular equilibrium” to describe an equilibrium in which all locations are inhabited.<sup>34</sup>

---

<sup>34</sup>Allen and Arkolakis (2014) also include continuity in their definition of a regular equilibrium in a continuous geography. Our geography is discrete.

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 locate at any point in the grid. Workers have common constant elasticity of substitution preferences given by

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

where  $\tilde{U}_i$  is the composite amenity fundamental of location  $i$ .

Production is perfectly competitive, although the model does nest cases of monopolistic competition.<sup>35</sup> 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 densities of workers and wages are given by  $L : N \rightarrow R_+$  and  $w : N \rightarrow R_{++}$ .

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

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

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

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

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

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

Combining the quantity (equation 9) and price (equation 10) expressions, we can write the value of the good produced in  $i$  consumed by  $n$  as

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

We can re-write the price index as

$$P_n = \left( \sum_{i \in N} \tau_{i,n}^{1-\sigma} w_i^{1-\sigma} \tilde{A}_i^{\sigma-1} \right)^{\frac{1}{1-\sigma}} \quad (12)$$

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

$$W_i = \frac{w_i}{P_i} \tilde{U}_i \quad (13)$$

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

---

<sup>35</sup>The online appendix to Allen and Arkolakis (2014) demonstrates that the model presented here nests several other QSE models.

$$w_i L_i = \sum_{n \in N} X_{i,n} \quad (14)$$

Additionally, the labor market clears as given by

$$\sum_{n \in N} L_n = \bar{L} \quad (15)$$

We can then combine the welfare expression (equation 13), value of consumption expression (equation 11), and income expression (equation 14) to get

$$L_i w_i^\sigma = \sum_{n \in N} W_n^{1-\sigma} \tau_{i,n}^{1-\sigma} \tilde{A}_i^{\sigma-1} \tilde{U}_n^{\sigma-1} L_n w_n^\sigma \quad (16)$$

The welfare expression combined with the price index yields

$$w_i^{1-\sigma} = \sum_{n \in N} W_i^{1-\sigma} \tau_{n,i}^{1-\sigma} \tilde{A}_n^{\sigma-1} \tilde{U}_i^{\sigma-1} w_n^{1-\sigma} \quad (17)$$

We assume transportation costs are symmetric, so that  $\tau_{n,i} = \tau_{i,n}$ .

We first consider the case with no productivity or amenity spillovers and demonstrate how to characterize the resulting equilibrium population distribution. While the method we use in the case with spillovers is largely identical, the result in this simpler case relies primarily on standard linear algebra results rather than results from the study of nonlinear systems and is helpful for understanding why the equilibrium population follows a lognormal distribution.

With no productivity or amenity spillovers,  $\tilde{U}_i = U_i$  and  $\tilde{A}_i = A_i$ . In the case where welfare is equalized across locations, which will hold given homogeneous labor with free mobility,  $W_i^{1-\sigma} = \bar{W}^{1-\sigma}$ . This gives the following two equations:

$$\bar{W}^{\sigma-1} L_i w_i^\sigma = \sum_{n \in N} \tau_{i,n}^{1-\sigma} A_i^{\sigma-1} U_n^{\sigma-1} L_n w_n^\sigma \quad (18)$$

$$\bar{W}^{\sigma-1} w_i^{1-\sigma} = \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} w_n^{1-\sigma} \quad (19)$$

As spatial correlations demonstrated the first law of geography, these equilibrium conditions reflect second-nature geography and the related “second law of geography”: the phenomenon external to an area of interest affects what goes on inside (Tobler, 2004). Wages and population in any given location depend on the wages and populations of neighboring locations.

In matrix form, we can write  $f_i = L_i w_i^\sigma$  and  $g_i = w_i^{1-\sigma}$ , which are the dominant eigenvectors of the respective systems, and  $\lambda = \bar{W}^{\sigma-1}$  denotes the common eigenvalue. These systems are

$$\lambda \mathbf{f} = \mathbf{K} \mathbf{f}$$

$$\lambda \mathbf{g} = \mathbf{K}' \mathbf{g}$$

where  $\mathbf{K}$  is given by

$$\underbrace{\begin{bmatrix} \tau_{1,1}^{1-\sigma} A_1^{\sigma-1} U_1^{\sigma-1} & \tau_{1,2}^{1-\sigma} A_1^{\sigma-1} U_2^{\sigma-1} & \dots & \tau_{1,N}^{1-\sigma} A_1^{\sigma-1} U_N^{\sigma-1} \\ \tau_{2,1}^{1-\sigma} A_2^{\sigma-1} U_1^{\sigma-1} & \tau_{2,2}^{1-\sigma} A_2^{\sigma-1} U_2^{\sigma-1} & \dots & \tau_{2,N}^{1-\sigma} A_2^{\sigma-1} U_N^{\sigma-1} \\ \dots & \dots & \dots & \dots \\ \tau_{N,1}^{1-\sigma} A_N^{\sigma-1} U_1^{\sigma-1} & \tau_{N,2}^{1-\sigma} A_N^{\sigma-1} U_2^{\sigma-1} & \dots & \tau_{N,N}^{1-\sigma} A_N^{\sigma-1} U_N^{\sigma-1} \end{bmatrix}}_{\mathbf{K}}$$

We now characterize the resulting population distribution, which for each location depends on both local fundamentals as well as those of its neighbors.

**Theorem 1:** Consider a regular geography with exogenous productivity and amenity fundamentals which are spatially correlated and lognormally distributed. Then:

- i. There exists a unique, regular spatial equilibrium
- ii. This equilibrium can be computed as the uniform limit of a simple iterative procedure
- iii. The equilibrium population distribution is approximately lognormal

This Theorem is similar to Theorem 1 of Allen and Arkolakis (2014), with the additions of the explicit assumptions regarding the form and spatial correlation of the fundamentals and point (iii) on the resulting population distribution. The proofs of part (i) and (ii) are the discrete case of the continuous-space model considered in Allen and Arkolakis (2014) and are included in the appendix.

To demonstrate (iii), consider the simple iterative procedure for solving equation 19 established in the proof of part (ii) in the appendix. This system can be solved by power iteration:

$$\mathbf{g}_{k+1} = \frac{\mathbf{K}\mathbf{g}_k}{\|\mathbf{K}\mathbf{g}_k\|}$$

where the full vector of guesses at stage  $k$  of iteration is given by  $\mathbf{g}_k$  and the subsequent guess given by  $\mathbf{g}_{k+1}$ . The guess for place  $n$  at step  $k$  is  $g_{n,k} = w_n^{1-\sigma}$ .

For any valid initial guess  $\mathbf{g}_0$  (wages are solved up to a constant, so the only restriction is that all  $g_{n,0} \in \mathbb{R}_{++}$ ), iterating over this procedure will result in convergence to the true  $\mathbf{g}^*$ . Writing out the value of  $g_{i,k+1}$  for some  $i \in N$ , we have

$$g_{i,k+1} = \frac{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} g_{n,k}}{\sum_{l \in N} \sum_{n \in N} \tau_{n,l}^{1-\sigma} A_n^{\sigma-1} U_l^{\sigma-1} g_{n,k}} \quad (20)$$

where the bottom term is a constant determined by the normalization. Considering a particular initial guess will allow us to characterize the resulting population distribution, as any starting point will arrive at the same final eigenvector (up to scale). Consider an initial guess such that  $g_{n,0} = \bar{g} > 0, \forall n \in N$ . This will result in  $g_{i,1}$  being expressible as

$$g_{i,1} = \frac{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} \bar{g}}{\sum_{l \in N} \sum_{n \in N} \tau_{n,l}^{1-\sigma} A_n^{\sigma-1} U_l^{\sigma-1} \bar{g}}$$

We can cancel  $\bar{g}$  from the top and bottom of the expression, and as the denominator is a constant the second guess is thus proportional to the numerator

$$g_{i,1} \propto \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} \quad (21)$$

This numerator will determine the distribution of  $g_{i,1}$ , and consists of constants  $\tau_{n,i}$  (capturing the distance between locations) while  $A_n$  and  $U_i$  are random variables drawn from spatially correlated lognormal distributions. Lognormality is maintained under exponentiation, and so both  $A_n^{\sigma-1}$  and  $U_i^{\sigma-1}$  remain lognormal, and the product of two lognormal distributions is also lognormal given bivariate normality of the two distributions.<sup>36</sup> Multiplying each lognormal term resulting from the exponentiation and multiplication of  $A_n$  and  $U_i$  by the corresponding constant  $\tau_{n,i}$  will likewise maintain the lognormality of each term in the summation, and by construction these will be spatially correlated. Each element of  $g_{i,1}$  is distributed like the sum of correlated lognormal random variables.

**Key property of the sum of correlated lognormals:** *The sums of correlated lognormal random variables are well-approximated by a lognormal distribution.* Sums of correlated lognormal random variables are well-studied given the appearance of such sums across a wide range of fields such as electrical engineering, finance, and communications. No closed-form solution for the resulting distribution exists. However, a well-known and highly successful result is that sums of correlated lognormal random variables have an approximately lognormal distribution.<sup>37</sup> Several successful approximations of the resulting lognormal distribution, with moments calculated from the moments of the random variables entering the sum, have been developed over nearly a century (for a review, see Dufresne (2009)), with increasing refinements to the method of constructing the approximation (Fenton, 1960; Schwartz and Yeh, 1982; Mehta et al., 2007). Recent progress in the study of sum of correlated lognormals has developed lognormal approximations with only minor deviations from the true distribution of the resulting sum (Lo, 2012, 2013).<sup>38</sup>

We apply this approximation to characterize the equilibrium population distribution. Given the initial guess of  $\bar{g}$ , as shown in equation 21 the resulting subsequent guess is proportional to a sum of correlated lognormal random variables. Applying the approximation, the distribution of  $g_{i,1}$  will be well-approximated by a lognormal distribution.

As the iterative process continues, given the near-lognormality of  $g_{i,1}$ , the distribution of  $g_{i,2}$  will also be well-described by a lognormal distribution. Again noting that the denominator is a constant term

$$g_{i,2} \propto \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} g_{n,1}$$

where the resulting  $g_{i,2}$  will be approximately lognormal as a result of the lognormality of the productivity and amenity fundamentals, their correlation, and the near-lognormality of  $g_{i,1}$ ,

---

<sup>36</sup>This holds by assumption—we assume that attributes and weights are such that  $A_i$  and  $U_i$  are bivariate normal.

<sup>37</sup>Even absent correlation, it has long been noted that finite sums of lognormals often appear to be well described by a lognormal distribution (Mitchel, 1968).

<sup>38</sup>For example of an application of the Lo (2012) approximation, see the documentation here [implementing the approximation](#).

and so on for all subsequent iterations. Any admissible initial guess will converge to the same vector over iteration. The particular guess of a constant  $g_{i,0} = \bar{g}$  for all  $i \in N$  allows us to see that the subsequent guess is near-lognormal, and this near-lognormality will be maintained in all subsequent iterations. This implies that when convergence is achieved the distribution of elements of the true vector  $\mathbf{g}^*$  will be well-described by a lognormal distribution, and as iteration does not depend on the initial guess this will hold regardless of the initial guess of the population distribution.

Wages can be solved from this distribution by exponentiating the elements of  $\mathbf{g}^*$  as  $w_i = g_i^{\frac{1}{1-\sigma}}$ , which will result in wages following a near-lognormal distribution. Likewise, populations can be solved for by iterating for the corresponding vector  $\mathbf{f}$  where  $f_i = L_i w_i^\sigma$ , where each element will also follow a near-lognormal distribution given the same iterative procedure. Solving for population can be done by combining the two vectors, preserving approximate lognormality for the population distribution as well.<sup>39</sup> This proves part (iii) of Theorem 1.

Next, we extend the model by considering the case with spillovers and externalities. Continuing to adapt from Allen and Arkolakis (2014), we now allow for externalities resulting from population agglomeration. We assume these externalities are such that productivity is increasing in agglomeration ( $\alpha > 0$ ) while amenities are decreasing ( $\beta < 0$ ), as is typical in the literature

$$L_i^{1-\alpha(\sigma-1)} w_i^\sigma = \bar{W}^{1-\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} A_i^{\sigma-1} U_n^{\sigma-1} L_n^{1+\beta(\sigma-1)} w_n^\sigma \quad (22)$$

$$w_i^{1-\sigma} L_i^{\beta(1-\sigma)} = \bar{W}^{1-\sigma} \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} L_n^{\alpha(\sigma-1)} U_i^{\sigma-1} w_n^{1-\sigma} \quad (23)$$

When bilateral trade costs are symmetric ( $\tau_{n,i} = \tau_{i,n}$  for all combinations), the system can be re-written such that the equilibrium can be characterized by a single equation. First, suppose that the relationship below holds, where  $\phi > 0$  is some scalar. Then

$$L_i^{1-\alpha(\sigma-1)} A_i^{1-\sigma} w_i^\sigma = \phi w_i^{1-\sigma} U_i^{1-\sigma} L_i^{\beta(1-\sigma)} \quad (24)$$

If this equation holds for all  $i \in N$ , then any sequences of  $L_i$  and  $w_i$  that satisfy one of the two prior equations will also satisfy the other. The proof of this unique relationship is in the appendix.

Substituting this expression into either of the two preceding terms, we can write the equilibrium as a single expression given by

$$\bar{W}^{\sigma-1} L_i^{\tilde{\sigma}\gamma_1} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} (L_n^{\tilde{\sigma}\gamma_1})^{\frac{\gamma_2}{\gamma_1}} \quad (25)$$

where

---

<sup>39</sup>The population vector can be recovered by iterating for  $\mathbf{f}^*$  and  $\mathbf{g}^*$ , and solving  $L_i = \bar{L} \frac{f_i g_i^{\frac{1}{1-\sigma}}}{\sum_{n \in N} f_n g_n^{\frac{1}{1-\sigma}}}$ .

$$\begin{aligned}\tilde{\sigma} &= \frac{\sigma - 1}{2\sigma - 1} \\ \gamma_1 &= 1 - \alpha(\sigma - 1) - \beta\sigma \\ \gamma_2 &= 1 + \alpha\sigma + (\sigma - 1)\beta\end{aligned}$$

This equation can also be expressed as:

$$\theta \mathbf{h} = \mathbf{J}[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}} \quad (26)$$

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  $J$  is given in the appendix.

We now establish the existence and uniqueness of the equilibrium, describe a mechanism for finding it, and characterize the population distribution.

**Theorem 2:** Consider a regular geography with overall productivity and amenity functions specified in equations (1) and (2), respectively, assume that iceberg trade costs are symmetric, and parameters are such that  $\gamma_1 > 0$  and  $\frac{\gamma_2}{\gamma_1} \in (-1, 1]$ .<sup>40</sup> Then:

- i. There exists a unique, regular spatial equilibrium;
- ii. This spatial equilibrium can be computed as the uniform limit of a simple iterative procedure;
- iii. The equilibrium population distribution will be approximately lognormal.

This Theorem is similar to Theorem 2 of Allen and Arkolakis (2014), with the addition of the explicit assumptions regarding the form and spatial correlation of the fundamentals and point (iii) on the resulting population distribution. The proofs of part (i) and (ii) are based on the discrete case of the continuous-space model considered in Allen and Arkolakis (2014), drawing on the results of Fujimoto and Krause (1985) and Karlin and Nirenberg (1967), and are included in the appendix.

As in the case without spillovers, the resulting population distribution can be understood by considering the iterative procedure described in point (ii). The iterative procedure with spillovers takes a similar form to standard power iteration and is given by

$$\mathbf{h}'_{k+1} = \frac{\mathbf{J}[\mathbf{h}_k]^{\frac{\gamma_2}{\gamma_1}}}{\|\mathbf{J}[\mathbf{h}_k]^{\frac{\gamma_2}{\gamma_1}}\|}$$

where the notation  $\mathbf{h}'_{k+1}$  indicates that the resulting vector may not satisfy the population condition when the vector  $\mathbf{L}'_k$  is recovered. As such, the resulting  $\mathbf{L}'_{k+1}$  will need to be rescaled by some scalar  $c$  such that  $\sum_{i \in N} c \cdot L'_{i,k} = \sum_{i \in N} L_{i,k} = \bar{L}$  before iterating over the subsequent guess  $\mathbf{h}_{k+1}$ .

Again, consider an initial guess such that  $L_i = \frac{\bar{L}}{N}$ , which gives  $h_{n,0} = \left(\frac{\bar{L}}{N}\right)^{\tilde{\sigma}\gamma_1} = \bar{h}$  for all  $i \in N$ . Then for each location  $i$ , we have

---

<sup>40</sup>Our preferred parameter values ( $\alpha = 0.03, \beta = -0.429, \sigma = 5$ ) for in the simulations in section 4 result in  $\frac{\gamma_2}{\gamma_1} = -0.187$ .

$$h_{i,k+1} = \frac{A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \bar{h}^{\frac{\gamma_2}{\gamma_1}}}{\sum_{i \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \bar{h}^{\frac{\gamma_2}{\gamma_1}}}$$

This can again be re-written as

$$h_{i,1} \propto A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma}$$

As before, the summation will take the form of a sum of correlated lognormals, given the preservation of the lognormality of  $U_n$  and  $A_n$  under exponentiation, multiplication by constants, and multiplication of lognormals. Applying the approximation, the summation term will be well-described by a lognormal distribution. The resulting vector  $h'_{i,1}$  will need to be re-scaled such that  $h_{i,1}$  satisfies the population condition. This scaling will not change the distribution.

Continued iteration given the near-lognormal  $h_{i,1}$  will take the form:

$$h_{i,k+1} \propto A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}} \quad (27)$$

As each  $h_{i,1}$  will be near-lognormal, this will be preserved under exponentiation by  $\frac{\gamma_2}{\gamma_1}$ , and so the same approximation of sum of correlated lognormal terms applies. As a result, the subsequent vector  $h'_{i,2}$  will likewise be near-lognormal and can be exponentiated and rescaled such that the population condition holds to construct  $h_{i,2}$ . The same logic will hold for all subsequent iterations ensuring lognormality of all appropriately scaled  $h_{i,k}$ , for  $k > 1$ . As the procedure will converge to the true vector  $\mathbf{h}^*$ , and all steps  $k > 1$  of the procedure are near-lognormal, the eventual population distribution will be approximately lognormal. This final distribution will hold given any initial guess, proving part (iii) of Theorem 2.

## 4 Results: Numerical Simulation of the Model

In this section we demonstrate that the model is successful at generating near-lognormal population distributions. We then show that it can reliably generate city sizes distributions for standard parameter values and truncation points that closely mirror those in the empirical literature. Next, we demonstrate that Gibrat's law holds, showing that size-invariant growth is a feature of the equilibrium distribution as changing the total population of the country proportionally scales local populations. We provide comparative statics to document how changing the parameters for congestion, spillover, and trade costs influences the observed power law. We show that the deviations are in the direction implied by the comparative empirical literature.

### 4.1 Numerical simulation of the population distribution

To test the reliability of our spatially-correlated QSE model in producing realistic population distributions, we simulate the model to demonstrate that the resulting populations are indeed lognormally distributed and that the city distribution appears to follow a power law.

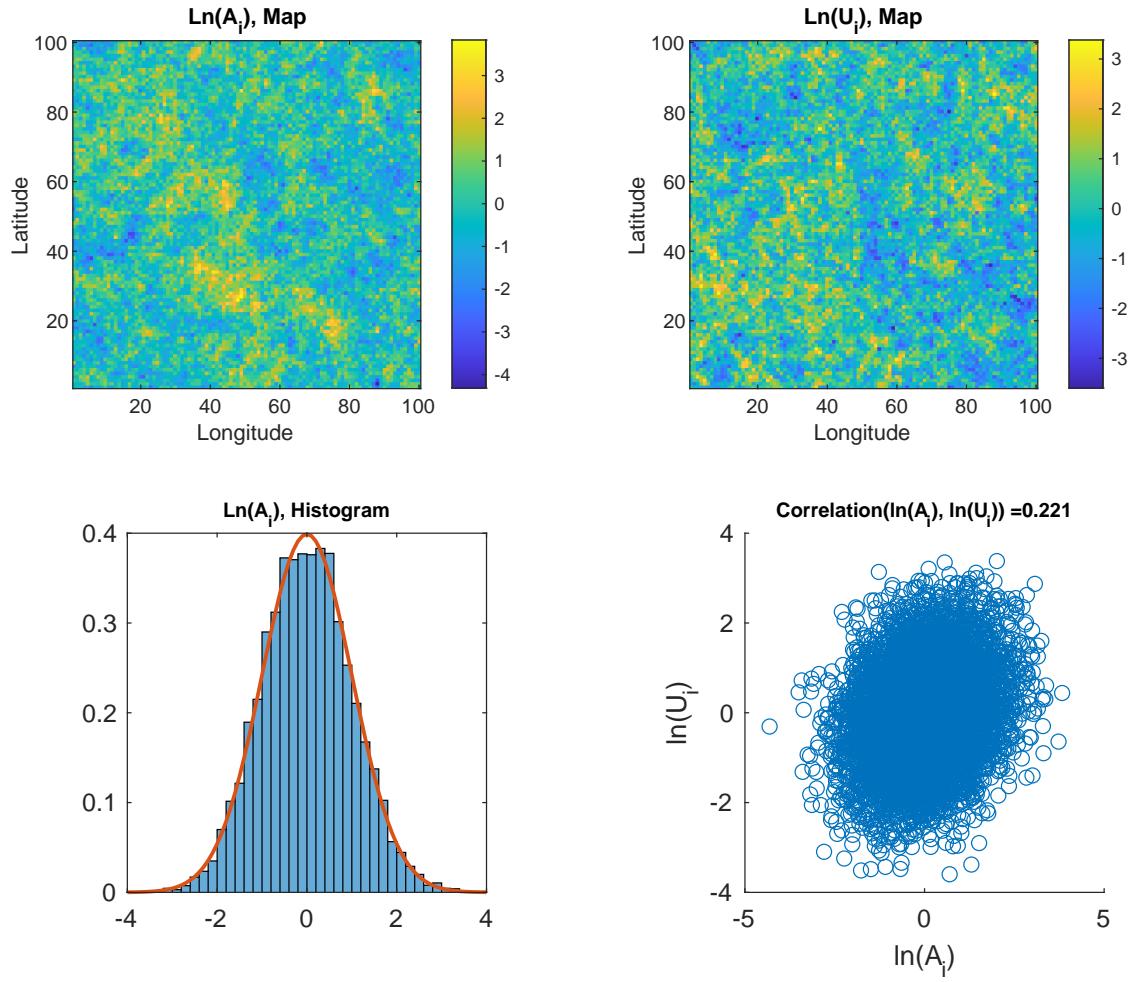


Figure 8: Example of the randomly generated model input. The top two panels display a map of the productivity and amenity fundamentals on a log scale. Both of these distributions are drawn from a lognormal distribution (the lower left panel shows that the log of  $A_i$  follows a normal curve). The maps show clear spatial correlation — places which have a high productivity or amenity value tend to be near other high value locations. We allow for correlation between the productivity and amenity fundamental, as demonstrated in the lower right panel.

Each location in our model should be interpreted as a region, which can potentially hold one settlement of any scale. “Cities” are thus tail realizations of the complete population distribution. This interpretation differs from some work in the QSE literature which considers all locations within discrete model as cities, but matches that in Redding and Rossi-Hansberg (2017). We define the most populous 5% of locations as “cities” within the model.

To simulate the model, we require randomly generated draws of exogenous productivity and amenity fundamentals. We draw these from lognormal distributions with parameters  $\sigma_{LN} = 1$  and  $\mu_{LN} = 0$  and induce spatial correlation using a Choleski decomposition for a chosen correlation structure.<sup>41</sup> We allow productivity and amenity fundamentals to be correlated, as while they are constructed using different weights of the underlying attributes as discussed in Section 2 there should be some degree of correspondence between the two fundamentals.<sup>42</sup> One such draw for both attributes can be viewed in Figure 8 on a 100-by-100 grid, where both show clusters of high and low values consistent with spatial correlation.

We use standard values in the literature to parameterize the other key parameters in the model. The benefits of increased density are given by  $\alpha = 0.03$ , which is in line with the estimates in Combes et al. (2008) and those surveyed in Rosenthal and Strange (2004) and Combes and Gobillon (2015). The model contains an isomorphism which we use to parameterize congestion costs. As discussed in Allen and Arkolakis (2014), the model is isomorphic to one with a fixed quantity of housing where spending on housing is  $\delta$  and  $\beta = -\frac{\delta}{1-\delta}$ . Congestion costs are parameterized to match a level of spending on housing of 30% of income, consistent with the midpoint of the estimates in Combes et al. (2018) and Davis and Ortalo-Magné (2011). This gives a congestion parameter of  $\beta = -0.429$  ( $\frac{3}{1-3} = 0.429$ ). The elasticity of trade costs to distance is given by  $\tau_\epsilon = 1$ , as is standard in the literature (Chaney, 2018). We vary these parameters in section 4.3.

Figure 9 shows the equilibrium population distribution associated with the exogenous productivity and amenity fundamentals in Figure 8. The log of the population distribution very closely matches the overlaid normal distribution, implying approximate lognormality. The upper right panel shows QQ plot of goodness of fit against the quantiles of a normal distribution, demonstrating very good fit throughout the full distribution. Concentrating only on the most populated 5% of locations within the model, we find that the model is very successful at generating Zipf’s law-like population distributions. Given the near-lognormality of the population distributions, the most populated locations in our model will appear to follow a power law distribution as demonstrated in equation 3. Performing the standard (log) rank-size regression on this simulated data, we find a slope of -1.009. The plot is visually similar to those for many countries using real data. Further, the most populous locations are not clustered in one place, with the population map indicating substantial dispersion of the largest cities as in real-world population distributions.

We demonstrate the robustness of this result by performing 1000 Monte Carlo simulations of the model, each time drawing a new randomly generated geography on a 100-by-100 grid.

---

<sup>41</sup>We assume the degree of spatial correlation of the log-scale fundamental declines in  $\frac{1}{dist_{i,j}+1}$ . For  $j = i$ , this gives the necessary  $\rho = 1$ , and for neighboring cells horizontally and vertical from  $j$  which are at a distance of 1 correlation is  $\rho = \frac{1}{2}$ , and so on. It is important to note that while this decay appears more rapid than that motivated by the empirical work on attribute correlations in Section 1, a direct comparison is difficult because measures of distance are not unit-less.

<sup>42</sup>In the below, we allow for correlation between productivity and amenity fundamentals of  $\rho_{AU} = 0.25$ .

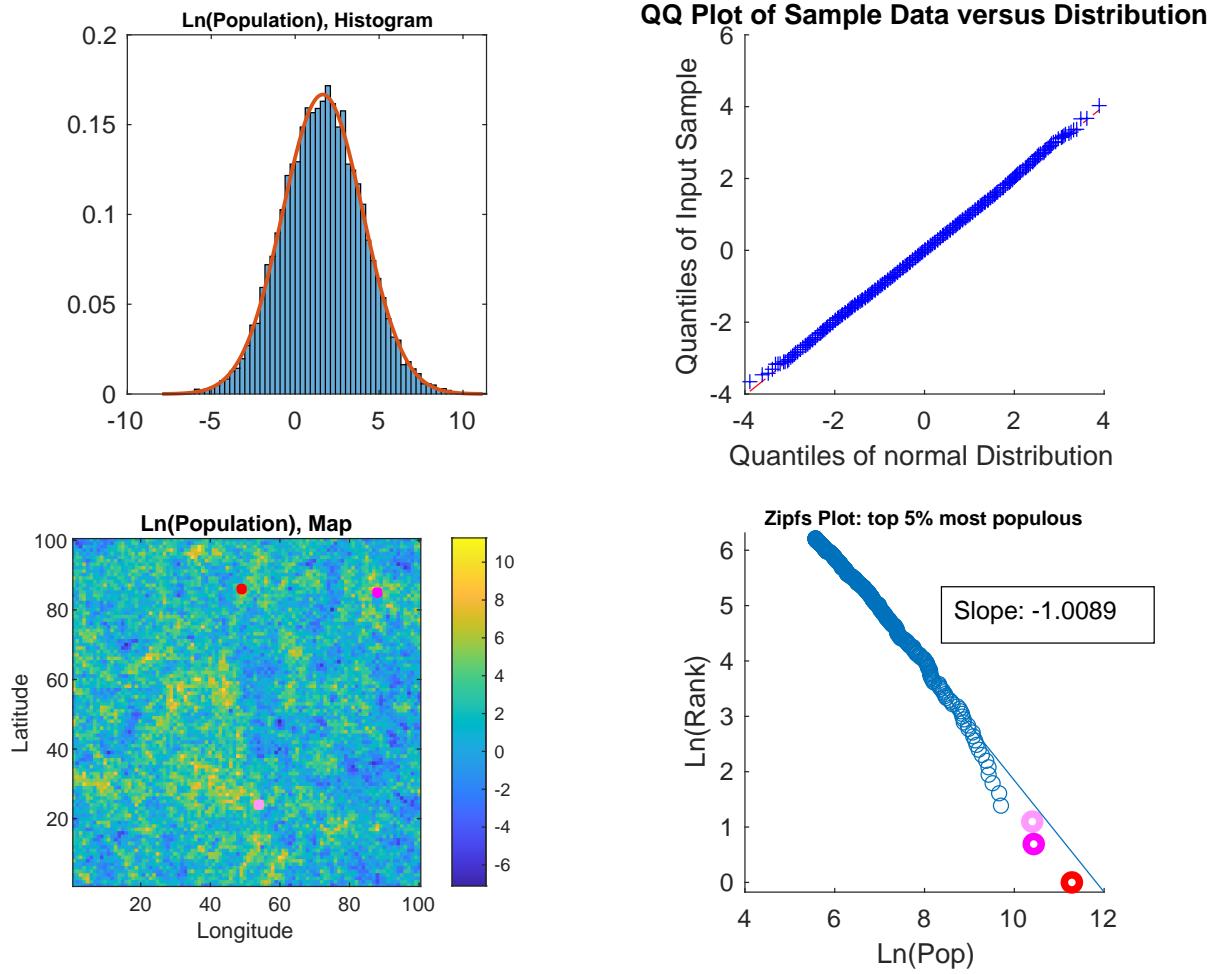


Figure 9: Example of the equilibrium population distribution. following simulation. The top left panel demonstrates that the model's log population appears to follow a normal distribution. The top right panel contains a QQ plot of the model's log population distribution, indicating that it very closely matches a normal distribution. The lower left panel shows the population distribution on a log scale, with the top three cities highlighted (to demonstrate these are not clustered in one location). The power-law plot on the lower right shows a good fit to Zipf's law, along with the characteristic divergence of the largest locations below the trendline.

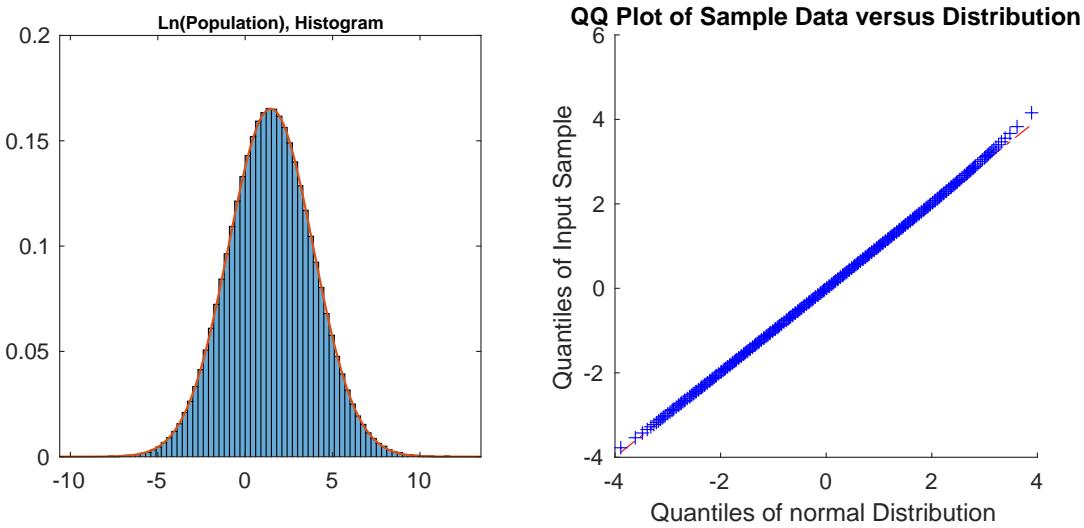


Figure 10: Results from smoothing the output over 1000 Monte Carlo simulations of the model. The population distribution resulting from numerical simulation of model is in the right panel, and the resulting QQ plot is on the left. Both show that the equilibrium population distribution appears lognormal.

	Kolmogorov–Smirnov	Lilliefors	Jarque–Bera
Rejected at 1%	0.000	0.068	0.273
Rejected at 5%	0.010	0.172	0.426

Table 1: Table shows the share of normality tests rejected for log equilibrium population over of 1000 Monte-Carlo simulations.

Figure 11 displays smoothed output over 1000 simulations of the model.<sup>43</sup> The QQ plot also demonstrates lognormality of the expected log population over these simulations. We further test each of the 1000 simulated population distributions against the null hypothesis that the logged population distribution is normally distributed using three standard tests of normality. In Table 1, we report share of times that normality of the log population is rejected for the Kolmogorov-Smirnov, Lilliefors, and Jarque-Bera tests at the 1% and 5% confidence level for the 1000 distributions. Normality of the log population is not reliably rejected by any of these tests. 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 the normal distribution.

The average estimated power law coefficient from the (log) rank-size regression across the 1000 simulations is 1.019, with a standard deviation of 0.073. 90% of estimated coefficients are between 1.137 and 0.901. Performing the (log) rank-size regression on the smoothed distribution delivers a slope of -1.02. The parameter values used here are consistent with the literature, as is the truncation point, and estimates come very near Zipf's law. However it is

<sup>43</sup>The log of population is averaged at each rank of the distribution over the 1000 simulations. Results are similar when averaging the population and taking the log.

important to note that, as discussed in Section 1, given an underlying lognormal distribution the specific power law exponent observed is not meaningful as changes in scale and truncation can influence the estimate. Further, the apparent power law distribution is also a result of the unknown variance of the unobserved exogenous fundamental terms. However, it is interesting to note how, for typical parameter values, the estimated power law exponent appears consistent with Zipf’s law. Varying these parameters to take other values typical in the literature also delivers estimates near -1. We discuss varying the parameters further in Section 4.3 below, with a particular focus on the comparative statics rather than the point estimate.

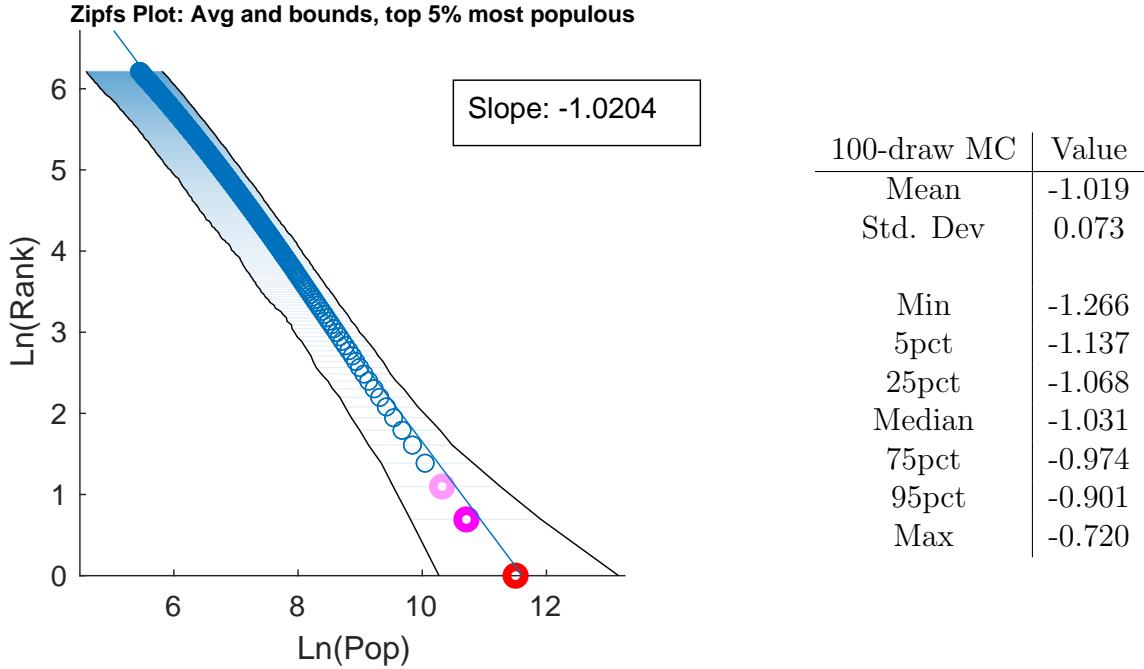


Figure 11: Monte Carlo average over model output (Left), and statistics over model simulations (Right). The slope on the left represents the slope taken over the average of  $\log(\text{pop})$  at each rank over 1000 simulations, and the bounds contain 95% of the log populations at each rank of the distribution. The table displays statistics over the 1000 estimated power law coefficients from the simulations.

## 4.2 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 population distribution demonstrates proportional growth and satisfies Gibrat’s law in equilibrium in response to increases in the aggregate population  $\bar{L}$ . We can write equation 26 as

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

where  $\tilde{h}_i = h_i \theta^{\frac{1}{1-\frac{\gamma_2}{\gamma_1}}} = h_i \bar{W}^{\frac{\sigma-1}{1-\frac{\gamma_2}{\gamma_1}}}$ . This expression is equivalent to equation 26, and must hold for any level of  $\bar{L}$ . As a result, changing  $\bar{L}$  does not impact the resulting population distribution even as it does impact welfare ( $\bar{W}$ , which is the same across all locations). This means that a percentage increase in overall population will result in each location experiencing population growth of the same amount. That is, population growth rates are unrelated to population size and Gibrat's law holds within the equilibrium of this model.

This model is static and not suited for the analysis of transition dynamics. To say a word on possible dynamics, however, while Gibrat's law is a feature of this equilibrium it would not be expected to hold if (assuming frictions prevent immediate movement) the equilibrium was disrupted, or if new unpopulated regions were added to the model. This could explain why instances where spatial equilibrium is disrupted, as in post-war Japan and Germany after bombing (Davis and Weinstein, 2002; Brakman et al., 2004), or in an economy in transition to a new spatial equilibrium as regions are settled, such as the Manifest Destiny-era United States (Desmet and Rappaport, 2017), fail to exhibit Gibrat's law.

### 4.3 Comparative statics

The variance of the resulting population is determined by the parameters of the model and the underlying variance of the unobserved exogenous fundamentals. Changing these parameters alters the estimated coefficient of the (log) rank-size regression, which we denote  $\beta_1$  as in equation 1. Over a set of reasonable parameter values, we perform 100 Monte Carlo simulations for each of 150 combination of parameters on a 50-by-50.<sup>44</sup> We find that increasing the benefits of agglomeration ( $\alpha$ ) results in a flatter slope (larger cities) and increasing local congestion costs results in a steeper slope (smaller cities). Increasing trade costs (by increasing how rapidly trade costs increase with distance) likewise results in a flatter slope and larger large cities. The flatter slopes for developing world countries (documented in Duben and Krause (2021)) may be attributable to high transportation costs, while the flattening slope in the US in recent decades (documented in Gabaix and Ioannides (2004)) could be a result of increased agglomeration benefits.

	Sign of change
$\frac{\partial \beta_1}{\partial \alpha}$	+
$\frac{\partial \beta_1}{\partial \beta}$	-
$\frac{\partial \beta_1}{\partial \beta_1}$	+
$\frac{\partial \tau_\epsilon}{\partial \beta_1}$	-
$\frac{\partial N}{\partial \beta_1}$	-

Table 2: Direction of the change in coefficient in the (log) rank-size regression for changes in  $\alpha, \beta, \tau_\epsilon$ , and  $N$ , holding other parameters constant. “+” means the slope has become flatter (larger large cities), while “-” means the distribution has become steeper (smaller large cities).

As the lognormal distribution is not scale invariant, another key parameter is the size of the simulated geography, which is given by the dimension of our  $N$  by  $N$  grid. Taking

---

<sup>44</sup>We simulate for 5 values of  $\alpha$ , 6 values of  $\beta$ , and 5 values of  $\tau_\epsilon$ .

more draws means that, in expectation, more very favorable locations are observed. For very large observations the deviation of the lognormal means that these appear to follow a different, steeper power law than the rest of the distribution. As a result, increasing  $N$  also results in a steepening of the distribution, as more large observations appear but these tend to fall below the trendline. In large countries, this deviation is particularly evident in the data as discussed in Section 1. Increasing the truncation point (selecting a larger minimum city) also increases the estimated power law coefficient by placing more weight on the observations drawn from a region of the lognormal distribution that approximate a power law with a higher exponent, as in equation 3 and explained in Eeckhout (2004). If the true distribution were Pareto, this relationship between  $N$ , the truncation point, and the slope would not hold as the Pareto distribution is scale invariant.

	$\beta = 0.25$	$\beta = 0.30$	$\beta = 0.35$	$\beta = 0.40$	$\beta = 0.45$	$\beta = 0.50$
$\alpha = 0.02$	-0.657 (0.081)	-0.735 (0.086)	-0.841 (0.098)	-0.935 (0.112)	-1.021 (0.112)	-1.096 (0.127)
$\alpha = 0.04$	-0.633 (0.073)	-0.714 (0.086)	-0.815 (0.093)	-0.886 (0.119)	-0.988 (0.108)	-1.071 (0.151)
$\alpha = 0.06$	-0.594 (0.076)	-0.690 (0.089)	-0.763 (0.101)	-0.88 (0.122)	-0.949 (0.114)	-1.028 (0.108)
$\alpha = 0.08$	-0.579 (0.066)	-0.665 (0.075)	-0.745 (0.095)	-0.854 (0.104)	-0.933 (0.115)	-0.999 (0.103)
$\alpha = 0.10$	-0.55 (0.062)	-0.629 (0.079)	-0.711 (0.087)	-0.795 (0.104)	-0.885 (0.108)	-0.981 (0.114)

Table 3: Table reports the average estimated coefficients from 100 simulations of a 50 by 50 grid using the combination of agglomeration benefits ( $\alpha$ ) and congestion cost ( $\beta$ ) parameters indicated.

The full table of mean estimated power law coefficients for 30 permutations, varying  $\alpha$  and  $\beta$  to cover the range of reasonable values in the literature, is in Table 3. In all of these simulations  $\tau_e = 1$ . The values for  $\alpha$  cover the estimates in Combes and Gobillon (2015) and Rosenthal and Strange (2004). The parameter values for  $\beta$  cover a share of spending on housing ranging from  $\delta = \frac{1}{5}$  (below those in Davis and Ortalo-Magné (2011)) to  $\delta = \frac{1}{3}$  (above those in Combes et al. (2018)), given the isomorphism of the model. Many of the estimated coefficients are not statistically different from -1, particularly in the portion of the parameter space near our preferred parameters of  $\alpha = 0.03$  and  $\beta = -0.429$  (corresponding to  $\delta = 0.3$ ). While the -1 exponent is likely not meaningful, it is interesting that it appears frequently for standard parameter values and reasonable truncation points.

## 5 Conclusion

The power law distribution of city populations within countries is a striking empirical regularity that holds across continents and millennia. In this paper, we have shown how a broad class of economic geography models generate these characteristic population distributions when modeled with a realistic, spatially-correlated geography. By demonstrating how both

first- and second-nature geography generate a lognormal population distribution, we integrate insights from economic geography theory about the importance of both place and space into the literature on Zipf's law and power law-like population distributions. This conception of the population distribution's emergence better and more reliably captures features of the real world than random growth-based theories that are prevalent in the literature. Viewing population distributions as arising naturally in response to favorable geography provides a simple explanation for the persistence of human settlements, the recovery of cities from disasters, and is consistent with the random growth phenomenon 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.
- Auerbach, F. (1913). Das gesetz der bevölkerungskonzentration. *Petermann's Geographische Mitteilungen*.
- Auerbach, F. and Ciccone, A. (2023). The law of population concentration. *Environment and Planning B: Urban Analytics and City Science*, 50(2):290–298.
- 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*, Chapter 5.
- 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.
- Chaney, T. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy*, pages 150–177.
- 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. (2018). 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 Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 247–348. Elsevier.

- Cuberes, D., Desmet, K., and Rappaport, J. (2021). Urban growth shadows. *Journal of Urban Economics*, 123:103334.
- Córdoba, J.-C. (2008). On the distribution of city sizes. *Journal of Urban Economics*, 63(1):177–197.
- Davis, D. R. and 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. Urbanization in Developing Countries: Past and Present.
- Dingel, J. I., Mischio, 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.
- Dufresne, D. (2009). Sums of lognormals. *Actuarial Research Conference*.
- EEckhout, J. (2004). Gibrat’s law for (all) cities. *American Economic Review*, pages 1429–1451.
- Fenton, L. F. (1960). The sum of lognormal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, 8.
- Fujimoto, T. and Krause, U. (1985). Strong ergodicity for strictly increasing nonlinear operators. *Linear Algebra and its Applications*, 71:101–112.
- 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. *In: Pines, D., Sadka E. and Zilcha, I., Eds, 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.
- 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*.
- Karlin, S. and Nirenberg, L. (1967). On a theorem of p. nowosad. *Journal of Mathematical Analysis and Applications*, 17(1):61–67.
- Kedem, B. and Chiu, L. S. (1987). On the lognormality of rain rate. *Proceedings of the National Academy of Sciences*, 84(4):901–905.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.
- Krugman, P. (1996). *The Self-Organizing Economy*. Blackwell.
- Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience*, 51(5):341–352.
- Lo, C. F. (2012). The sum and difference of two lognormal random variables. *Journal of Applied Mathematics*.
- Lo, C. F. (2013). Wkb approximation for the sum of two correlated lognormal random variables. *Applied Mathematical Sciences*.
- Madden, C. H. (1956). On some indications of stability in the growth of cities in the united states. *Economic Development and Cultural Change*, 4(3):236–252.
- 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*.
- Mehta, N. B., Wu, J., Molisch, A. F., and Zhang, J. (2007). Approximating a sum of random variables with a lognormal. *IEEE Transactions on Wireless Communications*, 6.

- Mitchel, R. (1968). Permanence of the log-normal distribution. *Journal of the Optical Society of America*.
- Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351.
- Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National 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. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics*.
- Redding, S. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*.
- 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.
- Schwartz, S. and Yeh, Y. (1982). On the distribution function and moments of power sums with lognormal components. *The Bell System Technical Journal*, 61.
- 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.
- Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers*, 94.
- Zipf, G. K. (1949). Human behavior and the principle of least effort. *Addison-Wesley Press*.

## A Proofs for main text

### 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  $\Gamma$ , the lognormal distribution can be written...

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

Which is the same as equation 3 in the main text.

### A.2 Proof of Theorem 1, (i) and (ii)

In the case with no spillovers, the two equations, expressed for all  $i \in N$ , can be written as the following vectors and matrices:

$$\underbrace{\bar{W}^{\sigma-1} \begin{bmatrix} L_1 w_1^\sigma \\ L_2 w_2^\sigma \\ \vdots \\ L_N w_N^\sigma \end{bmatrix}}_{\lambda \ f} = \underbrace{\begin{bmatrix} \tau_{1,1}^{1-\sigma} A_1^{\sigma-1} U_1^{\sigma-1} & \tau_{1,2}^{1-\sigma} A_1^{\sigma-1} U_2^{\sigma-1} & \cdots & \tau_{1,N}^{1-\sigma} A_1^{\sigma-1} U_N^{\sigma-1} \\ \tau_{2,1}^{1-\sigma} A_2^{\sigma-1} U_1^{\sigma-1} & \tau_{2,2}^{1-\sigma} A_2^{\sigma-1} U_2^{\sigma-1} & \cdots & \tau_{2,N}^{1-\sigma} A_2^{\sigma-1} U_N^{\sigma-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N,1}^{1-\sigma} A_N^{\sigma-1} U_1^{\sigma-1} & \tau_{N,2}^{1-\sigma} A_N^{\sigma-1} U_2^{\sigma-1} & \cdots & \tau_{N,N}^{1-\sigma} A_N^{\sigma-1} U_N^{\sigma-1} \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} L_1 w_1^\sigma \\ L_2 w_2^\sigma \\ \vdots \\ L_N w_N^\sigma \end{bmatrix}}_{f}$$
  

$$\underbrace{\bar{W}^{\sigma-1} \begin{bmatrix} w_1^{1-\sigma} \\ w_2^{1-\sigma} \\ \vdots \\ w_N^{1-\sigma} \end{bmatrix}}_{\lambda \ g} = \underbrace{\begin{bmatrix} \tau_{1,1}^{1-\sigma} A_1^{\sigma-1} U_1^{\sigma-1} & \tau_{2,1}^{1-\sigma} A_1^{\sigma-1} U_2^{\sigma-1} & \cdots & \tau_{N,1}^{1-\sigma} A_N^{\sigma-1} U_1^{\sigma-1} \\ \tau_{1,2}^{1-\sigma} A_1^{\sigma-1} U_2^{\sigma-1} & \tau_{2,2}^{1-\sigma} A_2^{\sigma-1} U_2^{\sigma-1} & \cdots & \tau_{N,2}^{1-\sigma} A_N^{\sigma-1} U_2^{\sigma-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1,N}^{1-\sigma} A_1^{\sigma-1} U_N^{\sigma-1} & \tau_{2,N}^{1-\sigma} A_2^{\sigma-1} U_N^{\sigma-1} & \cdots & \tau_{N,N}^{1-\sigma} A_N^{\sigma-1} U_N^{\sigma-1} \end{bmatrix}}_{\mathbf{K}'} \underbrace{\begin{bmatrix} w_1^{1-\sigma} \\ w_2^{1-\sigma} \\ \vdots \\ w_N^{1-\sigma} \end{bmatrix}}_{g}$$

where both take the form of eigenvectors of the matrices  $\mathbf{K}$  and  $\mathbf{K}'$ , with eigenvalues  $\lambda$ :

$$\lambda f = \mathbf{K}f$$

$$\lambda g = \mathbf{K}'g$$

That these eigenvalues should be common to both systems,  $\lambda = \bar{W}^{\sigma-1}$ , can be seen by noting that the corresponding matrices,  $\mathbf{K}$  and  $\mathbf{K}'$ , are square—the transposes of each other—and consist of all positive terms. Thus, by the Perron–Frobenius theorem, there is an eigenvector  $\mathbf{v}$  of the matrix  $\mathbf{K}$  consisting of all positive terms, and all other positive eigenvectors of  $\mathbf{K}$  are multiples of  $\mathbf{v}$  and there is one corresponding real eigenvalue, and the associated eigenvalue of a square matrix and its transpose are the same. This establishes the existence, uniqueness, and regularity of the equilibrium (as all elements of this vectors  $\mathbf{f}$  and  $\mathbf{g}$  are positive).

Calculating the equilibrium can be done straightforwardly using power iteration. The eigenvalue associated with the eigenvectors  $\mathbf{f}$  and  $\mathbf{g}$  is the Perron root of the matrix  $\mathbf{K}$ , and as such is the maximal eigenvalue. Power iteration of  $\mathbf{K}$  based on an arbitrary guess for  $\mathbf{g}$  or  $\mathbf{f}$  will result in convergence to either eigenvector.

Begin with any valid initial guess for  $\mathbf{g}_k = \mathbf{g}_0$ . Standard power iteration of the form below will converge to the true  $\mathbf{g}$  given by:

$$\mathbf{g}_{k+1} = \frac{\mathbf{K}\mathbf{g}_k}{\|\mathbf{K}\mathbf{g}_k\|}$$

For each row of the vector  $\mathbf{g}_k$ , corresponding to the value of  $g$  for a particular location, this initial guess and resulting value can be written

$$g_{i,k+1} = \frac{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} g_{n,k}}{\sum_{l \in N} \sum_{n \in N} \tau_{n,l}^{1-\sigma} A_n^{\sigma-1} U_l^{\sigma-1} g_{n,k}}$$

### A.3 Proof of Theorem 2, (i) and (ii)

First, we demonstrate that if there is a regular spatial equilibrium then equation 24 is the unique relationship between  $w_i$  and  $L_i$ . Re-write equation 28 as...

$$\phi_i = \frac{L_i^{1-\alpha(\sigma-1)} A_i^{1-\sigma} w_i^\sigma}{w_i^{1-\sigma} U_i^{1-\sigma} L_i^{\beta(1-\sigma)}} \quad (28)$$

...where we do not yet assume that  $\phi_i$  is the same across  $i \in N$ . Under the assumption of identical trade costs, we can substitute in equations 22 and 23 into the numerator and denominator of this expression

$$\phi_i = \frac{\sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\sigma-1} L_n^{1+\beta(\sigma-1)} w_n^\sigma}{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} L_n^{\alpha(\sigma-1)} w_n^{1-\sigma}}$$

Assume that the trade cost function  $\tau_{i,n}$  is symmetric such that  $\tau_{i,n} = \tau_{n,i}$  for all  $i, n \in N$ . Now, use equation 28 to re-write the above as

$$\begin{aligned}
\phi_i &= \frac{\sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{(1-\beta)(\sigma-1)} L_n^{1+\beta(\sigma-1)+\beta((\alpha-\beta)(\sigma-1)-1)} w_n^{\sigma+\beta(1-2\sigma)} \phi_n^\beta}{\sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{(1-\beta)(\sigma-1)} L_n^{1+\beta(\sigma-1)+\beta((\alpha-\beta)(\sigma-1)-1)} w_n^{\sigma+\beta(1-2\sigma)} \phi_n^{\beta-1}} \\
\implies \phi_i &= \frac{\sum_{n \in N} F_{i,n} \phi_n^\beta}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}}
\end{aligned}$$

where  $F_{i,n} = \tau_{i,n}^{1-\sigma} U_n^{(1-\beta)(\sigma-1)} L_n^{1+\beta(\sigma-1)+\beta((\alpha-\beta)(\sigma-1)-1)} w_n^{\sigma+\beta(1-2\sigma)}$ , using symmetry of trade costs. Each element of  $F_{i,n}$  is positive and bounded above by some finite number. Re-write the prior equation as

$$\begin{aligned}
\frac{\phi_i^\beta}{\phi_i^{\beta-1}} &= \frac{\sum_{n \in N} F_{i,n} \phi_n^\beta}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}} \\
\implies \frac{\phi_i^\beta}{\sum_{n \in N} F_{i,n} \phi_n^\beta} &= \frac{\phi_i^{\beta-1}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}}
\end{aligned} \tag{29}$$

and define  $\gamma_i = \frac{\phi_i^\beta}{\sum_{n \in N} F_{i,n} \phi_n^\beta}$ . Substituting this term in to equation 29 and defining  $j_{1,i} = \phi_i^\beta$  and  $j_{2,i} = \phi_i^{\beta-1}$ , we get

$$j_{1,i} = \sum_{n \in N} \gamma_i F_{i,n} j_{1,n}$$

$$j_{2,i} = \sum_{n \in N} \gamma_i F_{i,n} j_{2,n}$$

Both of these can be written in matrix form as

$$\begin{aligned}
\underbrace{\begin{bmatrix} j_{1,1} \\ j_{1,2} \\ \dots \\ j_{1,N} \end{bmatrix}}_{\mathbf{j}_1} &= \underbrace{\begin{bmatrix} \gamma_1 F_{1,1} & \gamma_1 F_{1,2} & \dots & \gamma_1 F_{1,N} \\ \gamma_2 F_{2,1} & \gamma_2 F_{2,2} & \dots & \gamma_2 F_{2,N} \\ \dots & \dots & \dots & \dots \\ \gamma_N F_{N,1} & \gamma_N F_{N,2} & \dots & \gamma_N F_{N,N} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} j_{1,1} \\ j_{1,2} \\ \dots \\ j_{1,N} \end{bmatrix}}_{\mathbf{j}_1} \\
\underbrace{\begin{bmatrix} j_{2,1} \\ j_{2,2} \\ \dots \\ j_{2,N} \end{bmatrix}}_{\mathbf{j}_2} &= \underbrace{\begin{bmatrix} \gamma_1 F_{1,1} & \gamma_1 F_{1,2} & \dots & \gamma_1 F_{1,N} \\ \gamma_2 F_{2,1} & \gamma_2 F_{2,2} & \dots & \gamma_2 F_{2,N} \\ \dots & \dots & \dots & \dots \\ \gamma_N F_{N,1} & \gamma_N F_{N,2} & \dots & \gamma_N F_{N,N} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} j_{2,1} \\ j_{2,2} \\ \dots \\ j_{2,N} \end{bmatrix}}_{\mathbf{j}_2}
\end{aligned}$$

The matrix  $\mathbf{F}$  is positive and is common each expression. Both  $\mathbf{j}_1$  and  $\mathbf{j}_2$  are eigenvectors of this matrix. By the Perron–Frobenius theorem, there is an eigenvector  $\mathbf{v}$  of the matrix  $\mathbf{F}$  consisting of all positive terms, and all other positive eigenvectors of  $\mathbf{K}$  are multiples of  $\mathbf{v}$ . Thus, both  $\mathbf{j}_1$  and  $\mathbf{j}_2$  must be multiples of  $\mathbf{v}$ , and there will be some constant  $\phi$  such that  $\mathbf{j}_1 = \phi \mathbf{j}_2$ . Using the definition of  $\mathbf{j}_1$  and  $\mathbf{j}_2$ , this means for all  $i$ ,  $\phi_i^\beta = \phi \phi_i^{\beta-1}$  and so  $\phi_i = \phi$ .

This establishes that equation 24 is the unique relationship between  $w_i$  and  $L_i$  in a regular spatial equilibrium.

Given the validity of equation 24, the equilibrium condition is as in equation 25 and can be expressed as

$$\bar{W}^{\sigma-1} L_i^{\tilde{\sigma}\gamma_1} = \sum_{n \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(\sigma-1)} (L_n^{\tilde{\sigma}\gamma_1})^{\frac{\gamma_2}{\gamma_1}} \quad (30)$$

We can write this more simply, with  $K_{i,n} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(\sigma-1)}$  and let  $h_i = L_i^{\tilde{\sigma}\gamma_1}$  and  $\theta = \bar{W}^{\sigma-1}$ , and thus

$$\theta \underbrace{\begin{bmatrix} L_1^{\tilde{\sigma}\gamma_1} \\ L_2^{\tilde{\sigma}\gamma_1} \\ \vdots \\ L_N^{\tilde{\sigma}\gamma_1} \end{bmatrix}}_{\mathbf{h}} = \underbrace{\begin{bmatrix} K_{1,1} & K_{1,2} & \dots & K_{1,N} \\ K_{2,1} & K_{2,2} & \dots & K_{2,N} \\ \vdots & \dots & \dots & \dots \\ K_{N,1} & K_{N,2} & \dots & K_{N,N} \end{bmatrix}}_{\mathbf{J}} \underbrace{\begin{bmatrix} (L_1^{\tilde{\sigma}\gamma_1})^{\frac{\gamma_2}{\gamma_1}} \\ (L_2^{\tilde{\sigma}\gamma_1})^{\frac{\gamma_2}{\gamma_1}} \\ \vdots \\ (L_N^{\tilde{\sigma}\gamma_1})^{\frac{\gamma_2}{\gamma_1}} \end{bmatrix}}_{[\mathbf{h}]^{\frac{\gamma_2}{\gamma_1}}}$$

Note that this takes much the same form as in the prior section, but the vector on the left-hand side consists of each element of the right-hand vector raised to  $\frac{\gamma_2}{\gamma_1}$ . We can then write this as...

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

as in the main text.

Fujimoto and Krause (1985) provide a generalization of the Perron-Frobenius theorem which extends to non-linear operators of this kind when  $\frac{\gamma_2}{\gamma_1} \in (0, 1]$ , and Karlin and Nirenberg (1967) provide a generalization that applies for  $\frac{\gamma_2}{\gamma_1} \in (-1, 0)$ . The operation that transforms each element of the vector  $\mathbf{h}$  in the above equation is a continuous operator which is strictly increasing and (positively) weakly homogeneous on the set of possible input values.

Call this operator  $\mathbf{T}$ . This operator satisfies the conditions for the main theorem of Fujimoto and Krause (1985) and Karlin and Nirenberg (1967), establishing the existence and uniqueness of the equilibrium. As the operator is defined such that it is bounded below and all  $h_i > 0$ , there must be a unique eigenvector  $\mathbf{h}^*$  and eigenvalue  $\theta$  which solve the equilibrium condition. This establishes existence and uniqueness of the equilibrium for  $\frac{\gamma_2}{\gamma_1} \in (-1, 0) \cup (0, 1]$

This leaves the cases of  $\frac{\gamma_2}{\gamma_1} = 0$ . In this case, the system is equivalent to  $\theta \mathbf{h} = \mathbf{J} \cdot \mathbf{1}^T$ , which has a unique solution given the matrix  $J$  and the population restriction. This establishes existence and uniqueness of the equilibrium for  $\frac{\gamma_2}{\gamma_1} \in (-1, 1]$ .

To establish that each location is inhabited if  $\gamma_1 > 0$  (establishing regularity), note that combining equations describing the value of trade (equation 11) and total income (equation 14) into the welfare expression (equation 13):

$$W_i = \frac{(\sum_{n \in N} \tau_{i,n}^{1-\sigma} P_n^{\sigma-1} w_n L_n)^{\frac{1}{\sigma}}}{P_i} A_i^{\frac{\sigma-1}{\sigma}} U_i L_i^{-\frac{\gamma_1}{\sigma}}$$

If a location was uninhabited, then if  $\gamma_1 > 0$  the marginal utility from moving to that location would be infinite. As there is free mobility, all locations must be inhabited in equilibrium. This completes the proof of part (i).

To show part (ii), as demonstrated in Allen and Arkolakis (2014), Fujimoto and Krause (1985) and Karlin and Nirenberg (1967) establish that the iterative procedure takes much the same form as that in Theorem 1...

$$\mathbf{h}'_{k+1} = \frac{\mathbf{T}(\mathbf{h}_k)}{\|\mathbf{T}(\mathbf{h}_k)\|}$$

...which will converge to the unique solution regardless of the initial guess of  $\mathbf{h}_0$ , up to a constant. In the main text this is written...

$$\mathbf{h}'_{k+1} = \frac{\mathbf{J}[\mathbf{h}_k]^{\frac{\gamma_2}{\gamma_1}}}{\|\mathbf{J}[\mathbf{h}_k]^{\frac{\gamma_2}{\gamma_1}}\|}$$

...to demonstrate the parallelism with the proof of Theorem 1.

The scaling is important in this case as each guess must be based on a vector of  $L_i$  which sums to the total population. We add a step to the iterative procedure to enforce the population condition. Iteration delivers  $\mathbf{h}'_{k+1}$  from which we recover the vector  $\mathbf{L}'_{k+1}$  that may not satisfy the population constraint. We thus re-scaled this vector by some constant  $\lambda$  so that  $\sum_{i \in N} \lambda L'_{i,k+1} = \sum_{i \in N} L_{i,k+1} = \bar{L}$ , and then exponentiate each element of the vector  $\mathbf{L}_{k+1}$  to find the our subsequent guess of  $\mathbf{h}_{k+1}$ .

For each row of the vector  $\mathbf{h}$ , corresponding to the value of  $h$  for a particular location, the iterative mechanism can be written in full as

$$h_{i,k+1} = \frac{A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}}{\sum_{i \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}} \quad (31)$$

## B Additional large country Zipf's plots

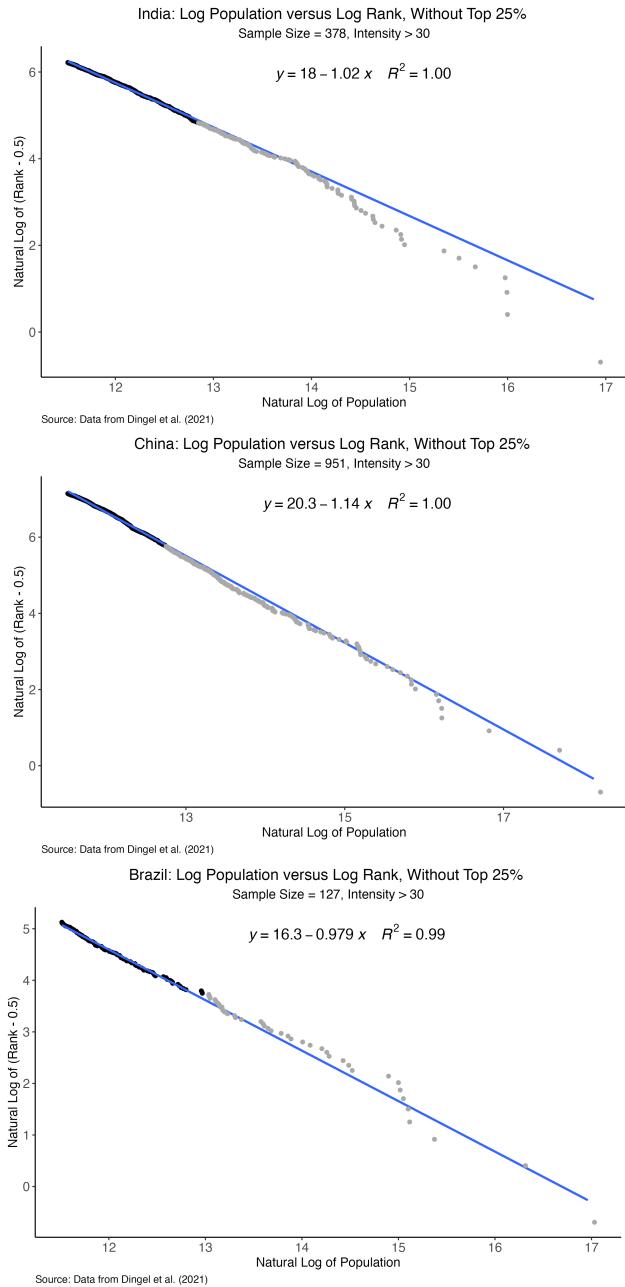


Figure A1: Plot of data from Dingel et al. (2021) based on a standardized night lights-based measure of population. Plots indicate that, as in the US, cities tend to deviate below the trendline excluding the top 25% of cities in India, China, and (to a lesser extent) Brazil . The cumulative missing population is 135 million (India), 53 million (China) and 8 million (Brazil).

## C Henderson et al. (2018) Data Description and Additional Table

In this section, we list the variables we used in Section 1 for our correlation matrices and tables. The data come from the publicly-available data associated with Henderson et al. (2018), itself an assemblage of data from various sources.

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 the 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<sup>2</sup>
11. *Growing Days*: Length of agricultural growing period, days/year

For more details on the variables and their original sources, please refer to Henderson et al. (2018).

## D Additional Spatial Correlation Figures

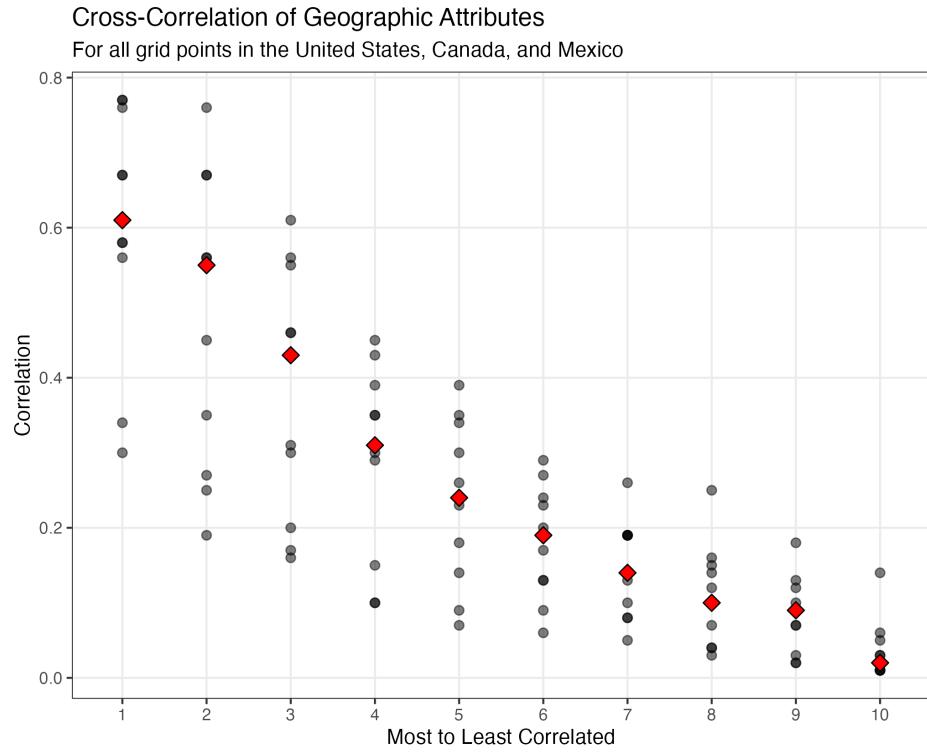


Figure A2: Structure of cross-correlations for geographic attributes in the U.S., Canada, and Mexico. For all eleven geographic attributes, the absolute values of the cross-correlations were calculated and ranked in descending order from 1 to 10 (1st being most correlated, 10th being least; we omit self-correlation). Each grey dot at each index is an attribute's cross-correlation; the red diamonds are the medians at each index.

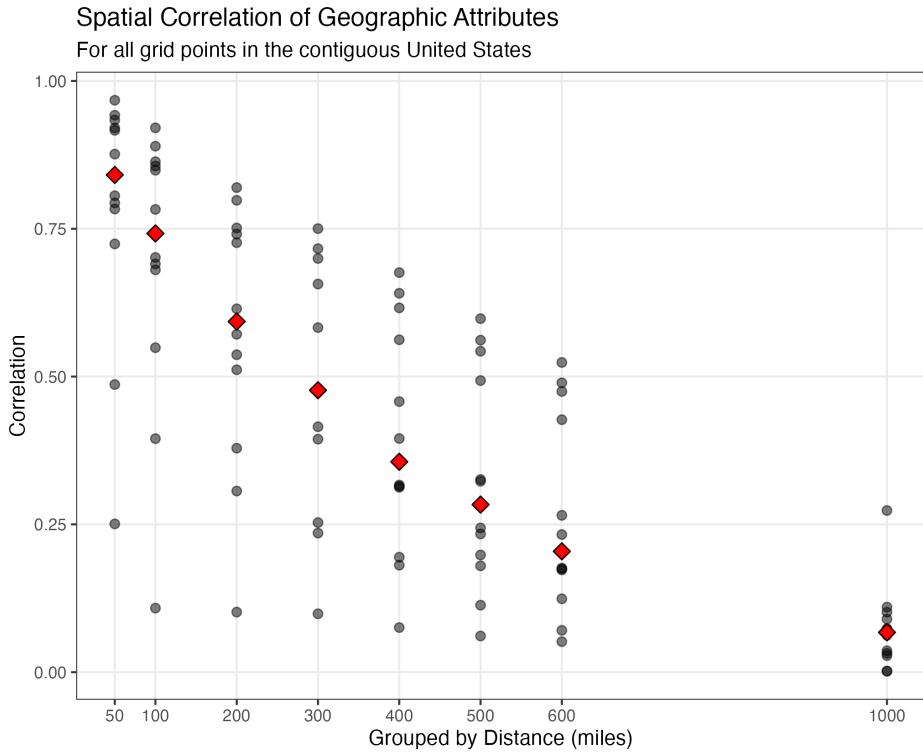


Figure A3: Spatial Correlation Structure of U.S. Geographic Attributes. Values shown are for correlations between grid points in the United States and randomly-selected grid points in the United States, Canada, and Mexico. Each grey dot at each distance is an attribute's correlation; the red diamonds are the medians at each distance. Absolute values of correlations were used.

## E County Attribute Analysis

In this section, we complement our exercise from Section 2.1 with U.S. county-level data, in the spirit of Behrens and Robert-Nicoud (2015).

### E.1 Data

We use data from the U.S. Department of Agriculture on climate and topography to calculate spatial correlation of physical attributes. The U.S.D.A. data is aggregated to the county level, giving us over 3,000 observations for analysis. For each county, the U.S.D.A. data contains information on average January and July temperatures, average hours of sunlight in January, a topography index representing ruggedness and elevation, and the percentage of area in a county which contains bodies of water. All averages come from data collected between 1941 and 1970. The attributes, with the exception of “Pct Water Area”, are logged; variables “July Temp”, “July Humidity”, and “Topography” are inverted to ordinaly structure all county variables from *worst* to *best*.<sup>45</sup> We then merge these six attributes at the county level with U.S. Census Bureau data on population, home ownership rates, and median household incomes at the county-level. Finally, we calculate the geographic centers for all counties in the U.S. in R.

### E.2 Attributes and space

We first calculate cross-correlations between our six attributes for all counties; the results are provided in Table A1. We observe that a location’s underlying attributes show little dependence within that location. An obvious exception is for variables that intuitively opposites, such as mean January and mean July temperatures.

Next, we demonstrate that while attributes in each location show little dependence, there is strong positive correlation within attributes across space; the results are provided in Table A2. This table calculates spatial correlations for each attribute for each county by comparing a particular centroid county’s attribute value to the attribute value of counties within a certain distance from the centroid. We calculate the correlations between centroid counties all counties within a certain radius from centroid counties, and between centroid counties and a randomly-selected county in a 20-mile wide ring from each centroid county; the latter is the same exercise as done for gridded data in the main text. In the latter case, the column labeled “400” is randomly-selecting a county 390-410 miles away from the centroid county, as before.

As a brief aside, extending radii further to 1000 and beyond necessarily introduces inaccuracies in measurement of correlations because of sample selection issues for counties near oceans or borders with Canada or Mexico. As a result, the point estimates in the final column for the county data should be taken with caution.

---

<sup>45</sup>These transformations follow Behrens and Robert-Nicoud (2015).

---

### Correlation Matrix of U.S. Geographic and Climate Attributes

*Correlations across attributes for all counties*

	Jan Temp	Jan Hrs Sun	July Temp	July Humidity	Topography	Pct Area Water
January Temp	1.00	0.20	-0.65	-0.18	0.08	0.00
January Hrs Sun	0.20	1.00	-0.48	0.11	0.23	-0.21
July Temp	-0.65	-0.48	1.00	0.19	-0.36	0.07
July Humidity	-0.18	0.11	0.19	1.00	-0.28	-0.27
Topography	0.08	0.23	-0.36	-0.28	1.00	0.20
Pct Area Water	0.00	-0.21	0.07	-0.27	0.20	1.00

Description: All data at county level; variables which are means were averaged over the time period 1941-1970. 'January Temp' is mean temperature in January. 'January Hrs Sun' is mean hours of sun in January. 'July Temp' is mean temperature in July. 'July Humidity' is mean humidity in July. 'Topography' is a discrete value between 1 and 21 indicating elevation and ruggedness of a location. 'Pct Area Water' is a value between 0 and 100 indicating how much water is included within county limits as a percent of total county area. Variables were transformed following Behrens and Robert-Nicoud (2015): all values, with the exception of 'Pct Area Water', were logged; 'July Temp', 'July Humidity', and 'Topography' were inverted.

Source: U.S. Department of Agriculture and authors' calculations

---

Table A1: Correlation Matrix of U.S. Geographic and Climate Attributes

## Spatial Correlation Structure of U.S. Geographic and Climate Attributes

*Correlations between attributes of centroid county and the attributes of counties in X mile-radius or ring from centroid*

	50	100	200	300	400	500	600	1000
<b>Correlation between centroid and all counties in X mile disk from centroid</b>								
January Temp	0.99	0.99	0.98	0.96	0.95	0.93	0.92	0.70
January Hrs Sun	0.97	0.96	0.93	0.90	0.87	0.83	0.77	0.48
July Temp	0.95	0.93	0.91	0.88	0.85	0.83	0.80	0.66
July Humidity	0.98	0.96	0.94	0.92	0.91	0.89	0.87	0.77
Topography	0.90	0.84	0.74	0.62	0.48	0.39	0.37	0.31
Pct Area Water	0.76	0.69	0.62	0.56	0.52	0.46	0.40	0.25
<b>Correlation between centroid and random county in (X-10, X+10) mile ring from centroid</b>								
January Temp	0.99	0.98	0.96	0.93	0.88	0.80	0.63	-0.59
January Hrs Sun	0.97	0.94	0.87	0.80	0.67	0.43	0.11	-0.62
July Temp	0.94	0.91	0.87	0.83	0.78	0.71	0.58	-0.36
July Humidity	0.97	0.95	0.91	0.88	0.85	0.82	0.76	0.11
Topography	0.87	0.78	0.55	0.24	0.06	0.14	0.26	-0.10
Pct Area Water	0.70	0.60	0.51	0.39	0.27	0.07	0.05	0.03

Description: All data at county level; variables which are means were averaged over the time period 1941-1970. 'January Temp' is mean temperature in January. 'January Hrs Sun' is mean hours of sun in January. 'July Temp' is mean temperature in July. 'July Humidity' is mean humidity in July. 'Topography' is a discrete value between 1 and 21 indicating elevation and ruggedness of a location. 'Pct Area Water' is a value between 0 and 100 indicating how much water is included within county limits as a percent of total county area. Variables were transformed following Behrens and Robert-Nicoud (2015): all values, with the exception of 'Pct Area Water', were logged; 'July Temp', 'July Humidity', and 'Topography' were inverted.

Source: U.S. Department of Agriculture and authors' calculations

Table A2: Spatial Correlation Structure of U.S. Geographic and Climate Attributes

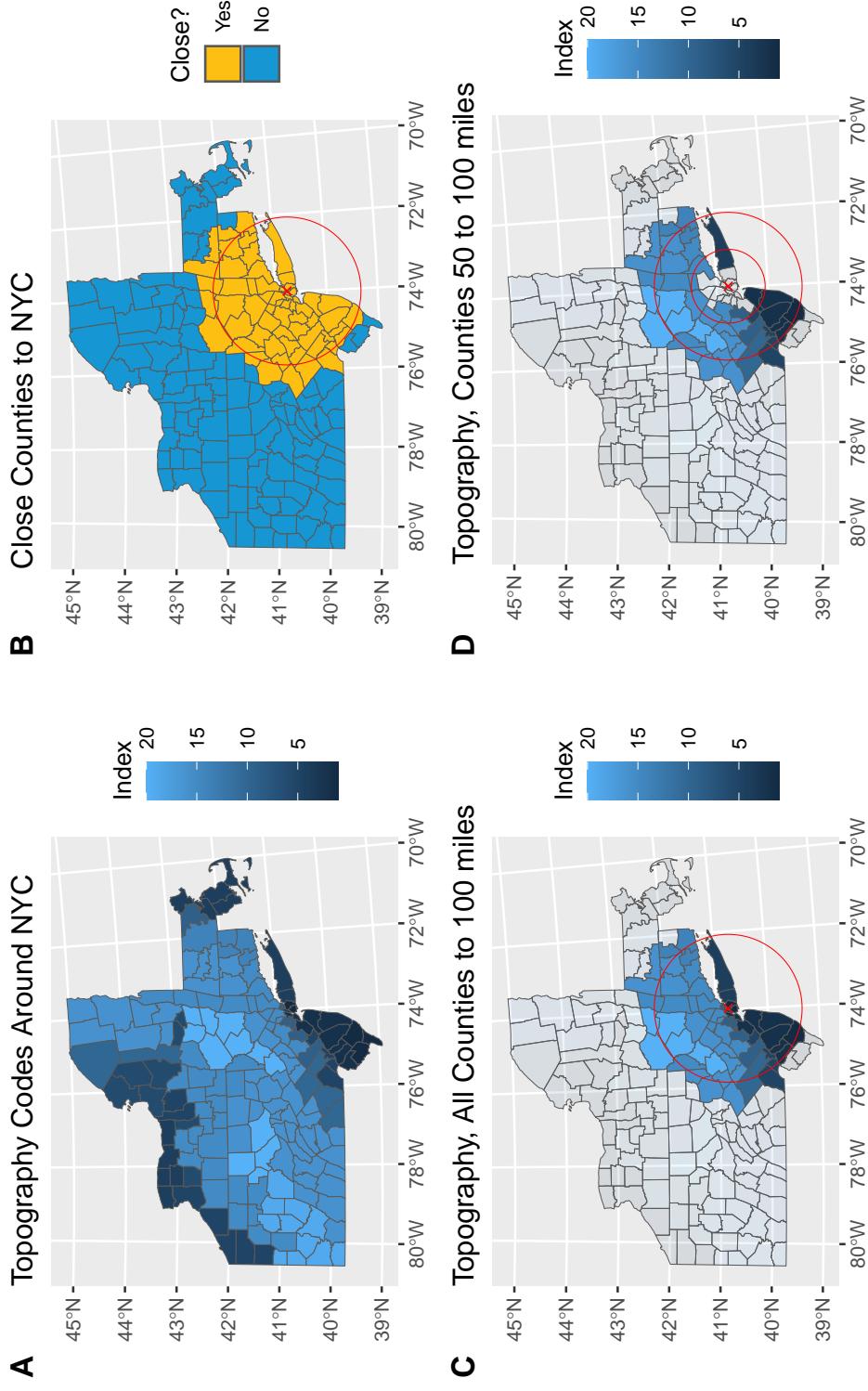


Figure A4: An example of how we calculate spatial correlations. **A:** Topography index for all counties in New York, Pennsylvania, New Jersey, Connecticut, and Massachusetts. **B:** Counties within 100 mile radius of the center (red “X”) of New York county (yellow) and all other counties outside that radius (blue). **C:** Topography scores of all counties within 100 mile radius of the center of New York County. **D:** Topography scores of all counties between 50 and 100 miles of the center of New York County. Topography score given by USDA according to elevation and ruggedness of terrain; score between 1 and 21.

*Source: U.S. Department of Agriculture and authors' calculations*

	<i>Dependent variable:</i>			
	Population			
	(1)	(2)	(3)	(4)
January Temp	0.860*** (0.068)	0.583*** (0.140)	0.269*** (0.091)	0.192** (0.082)
Jan Hrs Sun	−0.405*** (0.119)	−1.023*** (0.218)	−0.550*** (0.143)	−0.588*** (0.130)
(Inv) July Temp	2.940*** (0.521)	−4.109*** (0.692)	−0.728 (0.456)	−0.563 (0.411)
(Inv) July Humidity	−0.460*** (0.095)	−0.754*** (0.179)	−0.297** (0.117)	−0.233** (0.106)
(Inv) Topography	−0.017 (0.030)	−0.011 (0.032)	−0.055*** (0.021)	−0.072*** (0.019)
Pct Area Water	0.263*** (0.014)	0.175*** (0.014)	0.065*** (0.010)	0.051*** (0.009)
Land Area	−0.015 (0.035)	−0.052 (0.037)	0.043* (0.024)	0.070*** (0.022)
Constant	18.172*** (2.056)	−9.190*** (3.216)	7.754*** (2.120)	9.830*** (1.914)
State FE	No	Yes	Yes	Yes
Rural/Urban Code	No	No	No	Yes
Urban Influence Code	No	No	Yes	Yes
Observations	3,105	3,105	3,105	3,105
R <sup>2</sup>	0.212	0.377	0.736	0.786
Adjusted R <sup>2</sup>	0.210	0.366	0.730	0.781

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Description:** All data at county level; variables which are means were averaged over the time period 1941-1970. ‘January Temp’ is mean temperature in January. ‘January Hrs Sun’ is mean hours of sun in January. ‘July Temp’ is mean temperature in July. ‘July Humidity’ is mean humidity in July. ‘Topography’ is a discrete value between 1 and 21 indicating elevation and ruggedness of a location. ‘Pct Area Water’ is a value between 0 and 100 indicating how much water is included within county limits as a percent of total county area. ‘Land Area’ is size of county. Variables, dependent and independent, were transformed following Behrens and Robert-Nicoud (2015): all values were logged; ‘July Temp’, ‘July Humidity’, and ‘Topography’ were inverted.

*Source: U.S. Department of Agriculture*

Table A3: County Population on Attributes

## F Calculating the Fundamental

### F.1 Methodology

For every attribute in our data set which has a minimum value less than or equal to 0, we re-define the variable using an affine transformation to put the minimum around 0.1. This ensures that when we take the natural logarithm of variables we do not drop observations with negative values inside the logarithm.

We then construct the “worst-to-best” ordering of our attribute values in two separate ways:

1. **Naïve ordering:** we order the attribute values across all locations from worst-to-best by transforming the attributes’ underlying data. We decide how to order each attribute across locations according to the sign on each attribute from a regression for economic activity on attributes as performed 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.
2. **Absolute Distance:** we impose a relationship in which attribute values are ordered by the difference between the attribute values of the location (i.e. grid cell) with those of the grid cell with the highest economic activity in our data set. Attribute values closer to the “maximum” grid cell are better; attribute values farther away are worse. Attributes value orders were calculated as  $\frac{1}{|x_i^j - x_i^{max}| + 0.1}$ , where  $x_{ij}$  is the value of attribute  $i$  at location  $j$ , and  $x_i^{max}$  is attribute  $i$ ’s value at the grid cell with the highest radiance in the United States. This grid cell in the data for the contiguous U.S. is associated with Chicago, Illinois.

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

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

where  $\text{mean}(\ln(a_k))$  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

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

## F.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 A5: Summary statistics for geographic attributes for data points in the contiguous United States. Attributes values ordered using Henderson et al. (2018) ordering.

Summary Statistics for Attributes									
Variable	N	Min	Max	Median	Q1	Q3	IQR	Mean	SD
Ruggedness	13,426	-2.85	1.70	0.34	-0.70	0.72	1.42	0.00	1.00
Elevation	13,426	-1.66	5.41	-0.01	-0.91	0.63	1.54	0.00	1.00
Land Suitability	13,426	-1.35	1.72	-0.38	-0.87	1.03	1.90	0.00	1.00
Dist to River	13,426	-1.64	6.72	-0.23	-0.63	0.46	1.08	0.00	1.00
Dist to Coast	13,426	-1.05	7.47	-0.25	-0.77	0.46	1.23	0.00	1.00
Temperature	13,426	-1.63	3.52	-0.22	-0.74	0.56	1.30	0.00	1.00
Precipitation	13,426	-2.25	5.98	-0.33	-0.65	0.33	0.98	0.00	1.00
Dist to Harbor	13,426	-1.06	7.44	-0.25	-0.77	0.45	1.22	0.00	1.00
Growing Days	13,426	-0.97	5.89	-0.53	-0.69	0.53	1.21	0.00	1.00
Malaria Ecology	13,426	-6.14	0.73	0.33	0.01	0.33	0.33	0.00	1.00
Land Area	13,426	-1.02	22.56	-0.16	-0.49	0.23	0.72	0.00	1.00

Figure A6: Summary statistics for geographic attributes for data points in the contiguous United States. Attributes values ordered using Absolute Distance. Attributes value orders were calculated as  $\frac{1}{|x_i^j - x_i^{max}| + 0.1}$ , where  $x_{ij}^j$  is the value of attribute  $i$  at location  $j$ , and  $x_i^{max}$  is attribute  $i$ 's value at the grid cell with the highest radiance in the United States. This grid cell in the data for the contiguous U.S. is associated with Chicago, Illinois.

### F.3 Heatmap of Fundamental in Contiguous United States

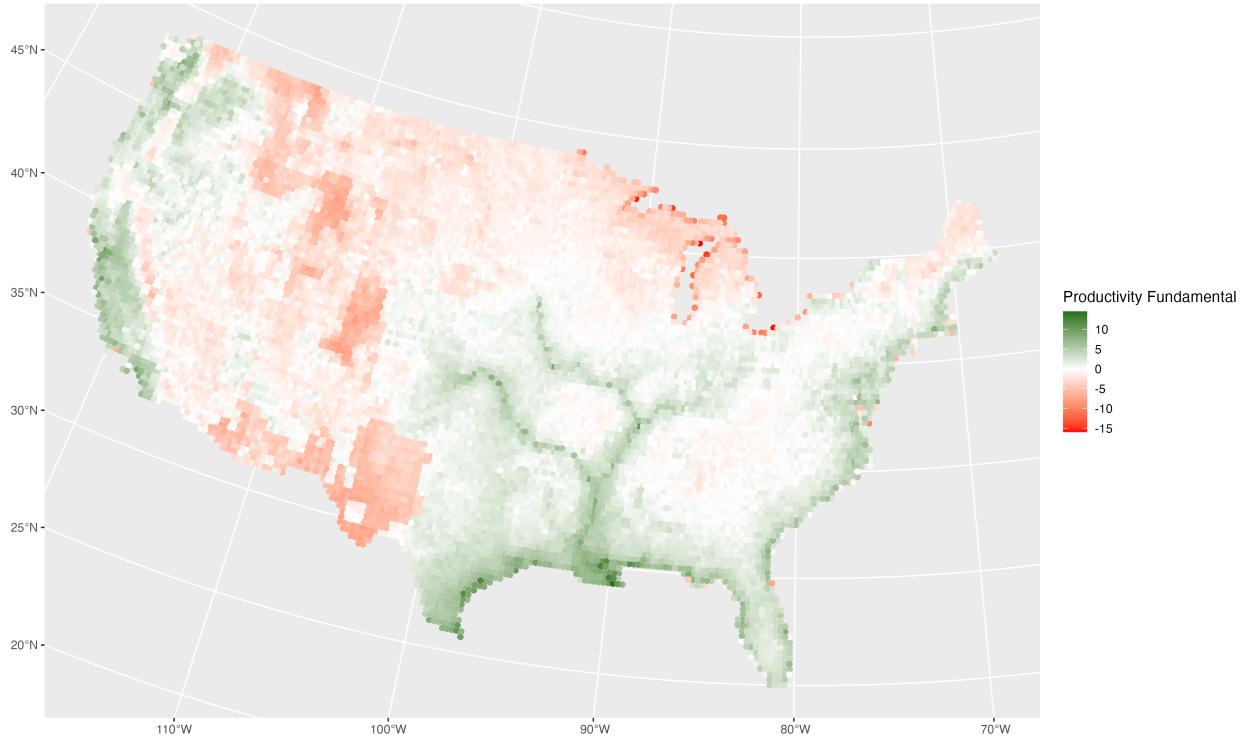


Figure A7: Heatmap for (log) productivity fundamental in the contiguous United States. Includes eleven attributes from Henderson et al. (2018). All variables were ordered worst to best in terms of contribution to economic activity à la Henderson et al. (2018), logged, then standardized. The fundamental is calculated as the standardized sum of the standardized, ordered attributes. The mean and variance are standardized to zero and one and a standard normal curve is overlaid.

## F.4 Regression Results for Henderson et al. (2018) Ordering

<i>Dependent variable:</i>	
(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 <sup>2</sup>	0.431
Adjusted R <sup>2</sup>	0.430
Residual Std. Error	2.294 (df = 13414)
F Statistic	922.395*** (df = 11; 13414)

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A4: Grid-Cell Radiant Lights on Attributes, contiguous United States. (Inv) indicates the attribute data was inverted.

Note that the signs on the coefficients may differ somewhat from Henderson et al. (2018); they used more attributes in their Table 1 regression than were used here. We provide this regression for robustness purposes to demonstrate that these 11 attributes produce an  $R^2 = 0.43$ , close to Henderson et al. (2018)'s 0.47.

## F.5 Regression Results for Absolute Distance Ordering

<i>Dependent variable:</i>	
(Log of) Radiance	
Ruggedness	0.278*** (0.018)
Elevation	0.823*** (0.020)
Land Suitability	0.141*** (0.032)
Dist to River	0.073*** (0.020)
Dist to Coast	0.122*** (0.046)
Temperature	-0.113*** (0.026)
Precipitation	0.106*** (0.030)
Dist to Harbor	-0.261*** (0.046)
Growing Days	0.476*** (0.028)
Malaria	0.118** (0.046)
Land Area	-0.055 (0.061)
Constant	5.542*** (0.416)
Observations	13,426
R <sup>2</sup>	0.376
Adjusted R <sup>2</sup>	0.376
Residual Std. Error	2.401 (df = 13414)
F Statistic	736.190*** (df = 11; 13414)

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A5: Grid-Cell Radiant Lights on Attributes, contiguous United States. Includes eleven attributes from Henderson et al. (2018). Attributes values calculated as  $\frac{1}{|x_i - x_i^{max}| + 0.1}$ , where  $x_i^{max}$  is attribute  $i$ 's value at location  $x^{max}$ , the grid cell with the highest radiance in the United States. Attribute values are thus ordered worst-to-best given their absolute distance from the reference grid cell; in the data set we used, this is Chicago, Illinois.

## F.6 Fundamental Under Alternative Specifications

### F.6.1 Absolute Distance

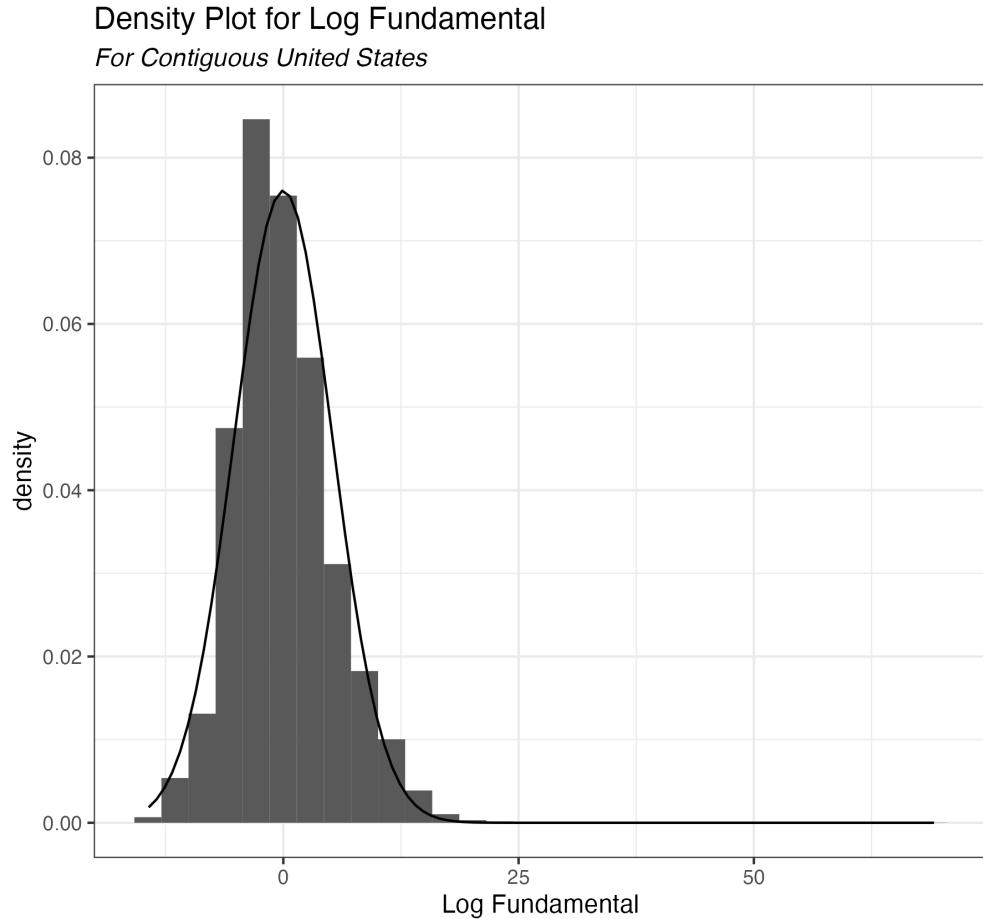


Figure A8: Lognormal distribution of locational fundamentals. Includes eleven attributes from Henderson et al. (2018). Attributes value orders were calculated as  $\frac{1}{|x_i^j - x_i^{max}| + 0.1}$ , where  $x_{ij}$  is the value of attribute  $i$  at location  $j$ , and  $x_i^{max}$  is attribute  $i$ 's value at the grid cell with the highest radiance in the United States. This grid cell in the data for the contiguous U.S. is associated with Chicago, Illinois. The fundamental is calculated as the sum of the standardized, ordered attributes. A normal curve with the same mean and variance as the empirical distribution is overlaid.

## G Gibrat's law plot, comparison with Rossi-Hansberg and Wright (2007)

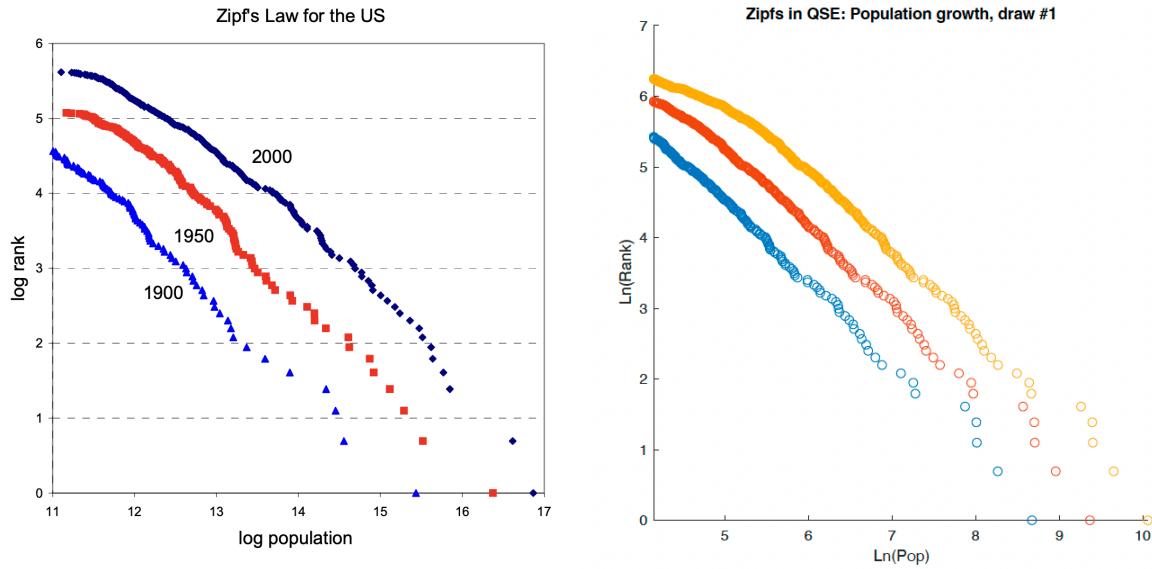


Figure A9: (Left) Rossi-Hansberg and Wright (2007), (Right) Model output

Rossi-Hansberg and Wright (2007) includes a plot of the U.S. city distribution in 1900, 1950, and 2000. Taking our model and simulating it with three values of  $\bar{L}$  such that  $\bar{L}_l < \bar{L}_m < \bar{L}_h$  and differ by the same relative values as the U.S. population in those three years. The purely simulated output on the right very closely resembles the real data on the left in Figure A9. Note, too, the curvature that emerges in both the real and simulated data. This is a result of the lognormality of the full real population distribution, which is evident when the data is not truncated.