Zipf's Law in Spatial Equilibrium

Matthew Easton and Patrick Farrell*

June 20, 2023

Abstract

The power law distribution of city populations, often called Zipf's law for cities, is a striking empirical regularity observed in most countries and documented across many periods of time as far back as the Bronze Age. City population distributions are also resilient, with the same cities holding roughly the same ranks in the distribution over long periods of time and recovering rapidly from large negative shocks. We propose an explanation for Zipf's law based on geography and the interactions between locations across space within standard quantitative spatial equilibrium (QSE) models that can account for these characteristics of city population distributions. We provide micro-foundations for aggregating spatially correlated observable attributes of locations into productivity and amenity "fundamentals", demonstrating that these fundamentals will also be spatially correlated and lognormally distributed. The equilibrium population will also follow a lognormal distribution as a result of this spatial correlation within a broad class of QSE models. For highly populated locations, i.e. cities, the population distribution will appear to follow a power law as a result of this inherited lognormality.

*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 Donald Davis, Thibault Fally, 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 financial support.

The distribution of city populations in nearly all countries tends to approximate a power law, a striking phenomenon often called Zipf's law for cities. Specifically, Zipf's law for cities states that the relationship between the natural log of city populations and the natural log of the rank of city populations is nearly linear with a slope of -1, or that the largest city within a given country tends to be twice the size of the second largest, three times the size of the third largest, and n times the size of the n^{th} largest. The existence of this distinctive statistical regularity has been documented in an extensive literature across countries (Soo, 2005; Cristelli et al., 2012), different periods of human history spanning millennia (Davis and Weinstein, 2002; Barjamovic et al., 2019), and varying definitions of cities (Dingel et al., 2021).

We propose an explanation for Zipf's law that highlights key roles for both place and space in shaping the population distribution. Both the observable attributes of each location and the interaction between locations across space necessarily shape human geography. That these two forces should both matter for the size of cities can be supported by a broad literature. We demonstrate how to generate such a distinctive population distribution within a spatial equilibrium model with a realistic geography that exhibits spatial correlation in observable geographic attributes. In doing so, we capture many other aspects of the population distribution and incorporate Zipf's law into the burgeoning literature on quantitative spatial equilibrium (QSE) models.

Several existing theoretical explanations for Zipf's law have been based on the concomitant and 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), and was originally formulated to describe the growth of firms. Importantly, Gibrat specifically noted that random growth processes will generate a lognormal size distribution. To explain Zipf's law for cities as a result of a Gibrat's law growth process, prior work has either introduced additional frictions to generate a true power law distribution to explain cities, such as a

Pareto distribution, or focused on the tail of a lognormal distribution, which can appear roughly Pareto distributed. Influential examples taking the former approach include Gabaix (1999b) and Gabaix (1999a), which introduce a friction by fixing the number of cities in order to generate a power law, and Rossi-Hansberg and Wright (2007) who endogenize the entry and exit of cities within a random growth model. For the latter approach, work by Eeckhout (2004) emphasizes the apparent lognormality of the full population distribution in a size-independent growth framework, noting that the upper tail of the full population distribution, when lognormal, well-approximates a power law relationship. Despite different approaches and distributional assumptions, these papers all take random growth to be the root cause of observed population distributions.

However, the random growth theory is inconsistent with many characteristics of real-world population distributions. City size distributions tend to be stable, with cities generally occupying approximately the same rank in the population distribution over long periods of time. 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, 2008; Johnson et al., 2019). Random growth processes cannot capture this fact. Davis and Weinstein (2002) argue in favor of a location fundamentals-based explanation, as the resilience and persistence of the observed population distribution supports there being properties of places that encourages habitation, and the distribution of these properties may lead to Zipf's law population distribution. 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 some kinds of equilibria, it is not an explanation or force itself which generates the equilibrium.

Other facts about the distribution of cities and their geography offer clues regarding the origin of the power-law distribution of city populations. In particular, cities do not appear to be randomly distributed. 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 resi-

dents. 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.[1] Indeed, a literature on the intuitive importance of natural characteristics for explaining settlement patterns has identified a large role for 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 even individual characteristics in shaping settlement patterns (Rappaport and Sachs, 2003; Nunn and Puga, 2012). Potential links between fundamentals and Zipf's law have been developed, such as in Behrens and Robert-Nicoud (2015) which proposes, but ultimately rejects, a fundamentals-based explanation for Zipf's law based on lognormally distributed locational productivities.[2]

Existing theories of Zipf's law's emergence, regardless of whether the population distribution results from random growth or locational fundamentals, lack a role for space. Yet recent theory in economic geography emphasizes an important role for space in determining the distribution of population and economic activity (Allen and Arkolakis, 2014). Additionally, recent empirical work has supported the importance of the *relative* location of economic activity and settlements in addition to exogenous fundamentals (Bosker and Buringh, 2017). Consider two unpopulated locations with equal high amenity and productivity 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

[1]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.

[2]The model presented in Behrens and Robert-Nicoud (2015) gives no role for space, and their empirical test of a fundamentals-based explanation consists of only a limited number of locational attributes. They found that 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. Yet as Henderson et al. (2018) finds, a larger and different set of geographic and climatic attributes indeed play a large role in explaining observed population distributions and settlements.

negative shocks, likewise, could be a result of the influence of nearby surviving cities rather than solely the persistent fundamentals of the impacted location. Even absent shocks, New York's large population is likely influenced by the many other large cities that surround it, while Lost Spring's location far from any settlement of large size could contribute to its small population.

In this paper, we propose an explanation for Zipf's law for cities as a product of observable, heterogeneous, and spatially correlated geographic attributes and the economic interaction across locations, generating population distributions which are consistent with many aspects of the city size distribution and economic geography theory. We build on the recent literature concerning quantitative spatial equilibrium (QSE) models, as in Allen and Arkolakis (2014), Redding (2016), and Redding and Rossi-Hansberg (2017). We begin by noting how heterogeneity in the observable physical characteristics of the world influences the unobservable productivity and amenity values of locations. Further, many of these observable physical characteristics are correlated across space: weather and topography are highly correlated, as are soil quality within a valley or access to a river for those upstream and downstream. We argue that a way to aggregate attributes into a fundamental that respects these characteristics of the world is as a product of appropriately ordered attribute values, or a sum of the natural log of the attribute values. The distribution of the resulting fundamentals will be lognormal as a result of the central limit theorem, and given spatial correlation in attributes the lognormal distribution of fundamentals will also exhibit spatial correlation.

We next describe how within a standard QSE model the spatial correlation and lognormality of these fundamentals result in a lognormal population distribution in equilibrium. This matches the lognormal distribution of real-world populations; moreover, for highly populous locations—i.e. cities—the distribution appears to follow a power law distribution. We show both the mechanism that links spatially correlated lognormal fundamentals to lognormal populations and the connection between lognormal population distributions and power

law-like tail behavior characteristic of Zipf's law. Via numerical simulation, we demonstrate that the model indeed generates a city population distribution that follows Zipf's law, exhibits Gibrat's law in equilibrium, and mirrors several other tendencies observed in real-world population distributions.

The paper proceeds as follows. Section 1 discusses the geographic characteristics of place and their correlations across space, laying out a framework for aggregating these heterogeneous and spatially correlated location attributes into terms that define a location's exogenous productivity and amenity fundamentals. Section 2 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. Section 3 highlights the link between lognormal populations and power laws, establishing the ability of the model to generate a power law-like distribution for city populations. Section 4 uses numerical simulation of the model to demonstrate the model's ability to capture several empirical results in the empirical Zipf's law literature. Section 5 concludes.

1 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 habitation 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, dense settlements and 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.

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 a necessary 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 observable "niceness" of place we can measure, but these differences do exist. Even though we cannot observe these fundamentals directly they must at some level be a function of the observable geographic characteristics of the world—they are not magical, mysterious quantities with no relation to the world we inhabit. The broad literature on the importance of geographic attributes for population distribution supports this intuitive assertion.

While it is clear that observable geographic attributes must influence production and amenity fundamentals, it is less clear in which particular way the things we can see impact things we cannot. The high degree of variation in most 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 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 at all (Phoenix). Despite such heterogeneity across individual attributes, these cities must all be attractive for human habitation based on either 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 defines 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 in the

U.S. of a new 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 estimation population of between 25 and 50 thousand people.[3] That is, even if a particular attribute of a certain place is not favorable, other attributes may be favorable instead—places that are populous have "something going for them," but that *something* may be different across different locations.

Given the importance of observable geographic attributes on productivity and amenity fundamentals for locations, we next continue with a deeper investigation of the properties of these fundamentals and the possible ways they contribute to the key fundamental values that can define productivity and quality of life within spatial models.

1.1 Data

To begin, we offer evidence of how geographic attributes are distributed across space.[4] We use data as collected by Henderson et al. (2018), maintaining any adjustments those authors made for augmenting existing data and imputations for missing data.[5] The authors collected data on a wide variety of geographic attributes, of which we borrow 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 stability, and total land area of the node. 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; all variables were logged; and all correlations were calculated to include some data from grid points outside of the U.S., specifically Mexico and Canada. Our sample of grid points for the U.S. is roughly 18,000 points; including Mexico

[3]Indeed, the persistence of human habitation in the region is more evidence for the importance of locational fundamentals for explaining settlement patterns.

[4]See the appendix for a different data source for geographic attributes, as described in Behrens and Robert-Nicoud (2015).

[5]In the appendix, we provide descriptions of all variables used.

Correlation Matrix of North American Geographic and Climate Attributes

Correlations across attributes for all grid points

	Rugged	Elevation	Land Suitability	Dist to River	Dist to Coast	Avg Temp	Avg Mon Precip	Dist to Harbor	Growing Days	Malaria	Land Area
Rugged	1.00	0.56	-0.03	-0.08	-0.04	0.07	0.13	-0.13	-0.08	0.06	0.17
Elevation	0.56	1.00	0.16	0.05	0.43	0.23	-0.10	0.27	-0.24	0.10	0.42
Land Suitability	-0.03	0.16	1.00	-0.15	0.48	0.71	0.26	0.35	0.41	0.16	0.62
Dist to River	-0.08	0.05	-0.15	1.00	-0.09	-0.15	-0.17	-0.01	-0.30	0.02	-0.14
Dist to Coast	-0.04	0.43	0.48	-0.09	1.00	0.29	-0.12	0.87	0.09	0.00	0.42
Avg Temp	0.07	0.23	0.71	-0.15	0.29	1.00	0.45	0.03	0.55	0.34	0.77
Avg Mon Precip	0.13	-0.10	0.26	-0.17	-0.12	0.45	1.00	-0.30	0.67	0.06	0.26
Dist to Harbor	-0.13	0.27	0.35	-0.01	0.87	0.03	-0.30	1.00	-0.13	-0.04	0.18
Growing Days	-0.08	-0.24	0.41	-0.30	0.09	0.55	0.67	-0.13	1.00	0.01	0.39
Malaria	0.06	0.10	0.16	0.02	0.00	0.34	0.06	-0.04	0.01	1.00	0.27
Land Area	0.17	0.42	0.62	-0.14	0.42	0.77	0.26	0.18	0.39	0.27	1.00

Description: All data at quarter-point latitude and longitude level; variables which are means were averaged over the time period 1960-1990. Values shown are for correlations of all grid points in the United States, Canada, and Mexico. All values were logged.

Source: Data as collected by Henderson et al. (2018)

Table 1: Correlation Matrix of North American Geographic and Climate Attributes

and Canada, our sample expands to over 47,000 points.

1.2 Attributes and space

We first calculate cross-correlations between our eleven attributes for all grid points; the results are provided in Table 1. We observe that a location's underlying attributes show little dependence between attributes within that location. An obvious exception is for variables that are, by their very nature, correlated, such as average temperature and land suitability or growing days.

Next, we demonstrate that while attributes within each location show little dependence, there is positive correlation within attributes across space; the results are in Table 2, with an illustration of the computational process behind correlation calculations provided in the appendix.[6] This table describes spatial correlations for each attribute at various distances. We do this calculation by taking a randomly-selected point from a 20-mile wide ring at

[6]The illustration in the appendix uses data and cells at the U.S. county level from a separate data source; the process, however, for grid points was the same.

Spatial Correlation Structure of Geographic and Climate Attributes in the U.S.

Correlations between attributes of centroid grid point and random point in X mile-ring from centroid

	50	100	200	300	400	500	600	1000
Ruggedness	0.46	0.35	0.27	0.23	0.20	0.18	0.18	0.03
Elevation	0.89	0.81	0.69	0.60	0.52	0.44	0.35	0.03
Land Suitability	0.87	0.79	0.70	0.63	0.57	0.53	0.49	0.31
Dist to River	0.87	0.76	0.57	0.40	0.26	0.18	0.14	−0.01
Dist to Coast	0.99	0.97	0.89	0.79	0.66	0.53	0.43	0.09
Avg Temp	0.98	0.96	0.93	0.90	0.86	0.81	0.77	0.67
Avg Mon Precip	0.93	0.86	0.73	0.64	0.57	0.50	0.42	0.10
Dist to Harbor	0.93	0.86	0.72	0.57	0.44	0.33	0.24	−0.02
Growing Days	0.92	0.84	0.73	0.66	0.59	0.52	0.44	0.11
Malaria	0.76	0.63	0.44	0.26	0.20	0.16	0.13	0.03
Land Area	0.47	0.40	0.34	0.35	0.33	0.32	0.31	0.24

Description: All data at quarter-point latitude and longitude level; variables which are means were averaged over the time period 1960-1990. Values shown are for correlations of grid points in the United States with all grid points in the United States, Canada, and Mexico. All values were logged.

Source: Data as collected by Henderson et al. (2018)

Table 2: Spatial Correlation Structure of U.S. Geographic and Climate Attributes

varying distances around each grid point, to ensure there are eligible points at roughly the desired distance, for all grid points in the U.S.[7] To illustrate, the column labeled "400" is choosing a random point 390 to 410 miles away from each centroid for comparison; a new point is drawn from a ring 490 to 510 miles away for the column "500".

Table 2 illustrates two main points. First, spatial correlation in attributes does not necessarily fall immediately over distance; any given grid point is, reasonably, somewhat correlated with all points around it. Second, we find that there is correlation above 0.50 at 600 miles away from the centroid point for just one of the eleven attributes; at only 100 miles away, however, nine out of eleven attribute have correlations at or above 0.50, indicating strong similarities for attributes in nearby areas. For nearly all variables, the correlations

[7]The appendix contains total correlations for each grid point, as well.

exhibit quick decay towards 0.[8] In summary, these results suggest clear spatial correlation of nearby attributes, followed by a reasonable lack of correlation at larger distances, in the real world.

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 neighborhoods 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 are substantially similar across a wide variety of attributes, as the data suggests.

1.3 Aggregating attributes

Given the rationale outlining and empirical support for spatial correlation of geographic attributes across space, we will formalize how to incorporate these features into spatial economy models. As stated earlier, models in regional, spatial, and urban economics assume that locations have exogenous productivity and amenity values. 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 of as defining the quality of life of living in a given location.

Rather than assuming these unobservable fundamental terms have an arbitrary distribution, we attempt to provide a microfoundation for these fundamentals based on those

[8]Robustness checks with Monte Carlo simulations for random-point selection are being worked on for future drafts.

elements of place we can observe. Different observable locational *attributes* can 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 1 for an illustration). Importantly, these observable attributes are also highly spatially correlated, as shown in the previous section using data from the United States.

We will now describe our aggregation of attributes to form the fundamentals. Attributes a_{ik} are associated with a location i with the type of attribute indexed by $k \in K$. Each attribute may differ in its degree of spatial correlation and in the distribution from which it is drawn. 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 higher values of $a_{(\cdot)k}$ reflect *better* draws. That is, these should not be thought of as being measured in the familiar units for each attribute.[9] Further, 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 to be uncorrelated with each other. We thus assume that all the attributes associated with each location i are, if not independent, only *weakly* dependent. This is consistent with the empirical evidence.

Locational fundamentals for location i are a combination of all of the attributes of that place. We will define productivity and amenity fundamentals in a similar way, and so will focus first on the productivity fundamental. We think of the locational productivity fundamental for a location i, denoted A_i, as a function of these attributes a_{ik}:

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

A location's overall fundamental A_i is increasing in each a_{ik}: $\frac{\partial F(\cdot)}{\partial a_{ik}} > 0$ for all $k \in K$. There are complementarities between each of the fundamental attributes—the benefit of having reliable rainfall for production is increased when there is better arable land in a

[9]Rainfall in inches has a nonlinear relationship with agricultural output, for instance.

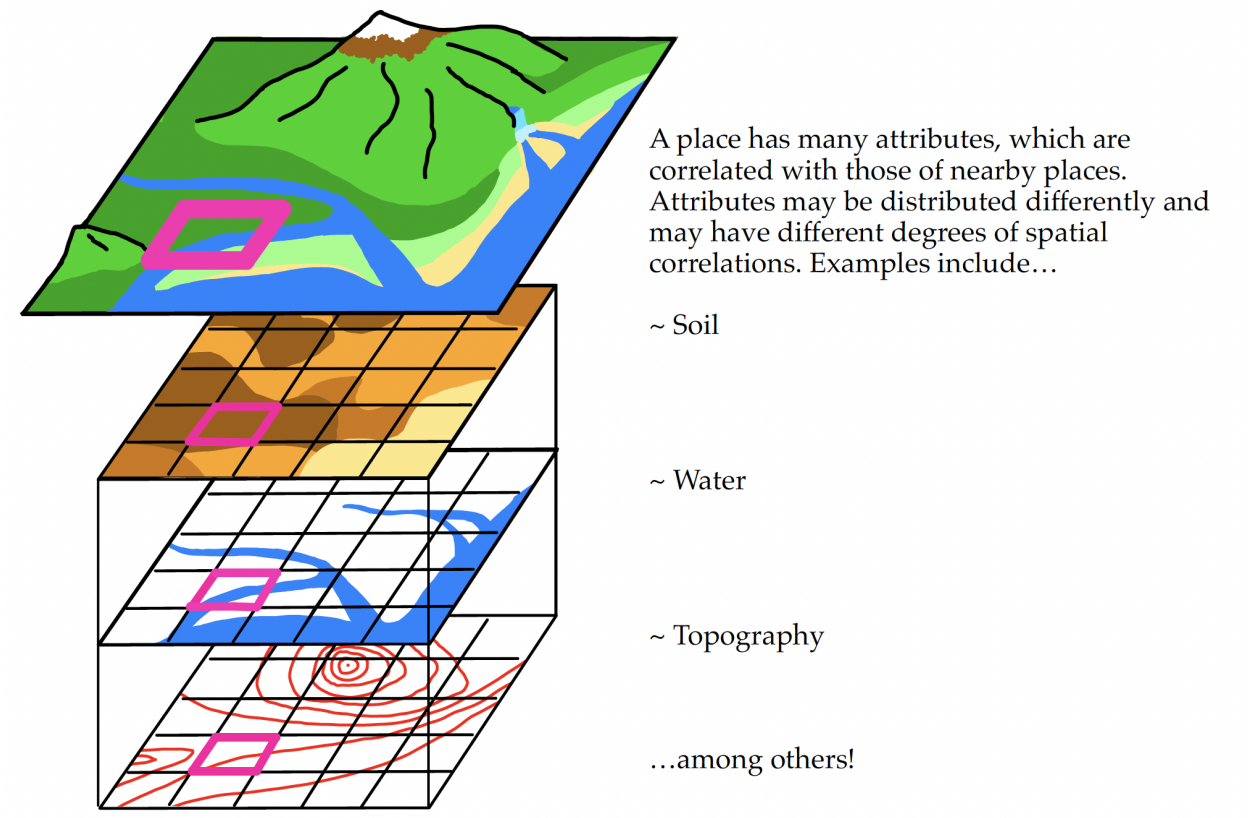


Figure 1: Attributes and their spatial correlation.

location. 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.[10]

A functional form for the aggregating function over location attributes delivering these properties will be multiplicative over all attributes, such as a Cobb-Douglas aggregator. We can also assume that the relative importance of attributes differs, and that these weights can vary over time (to capture structural transformation or changing production technologies, potentially) and are different for determining production and amenity fundamental values. Allowing the weight at time t for attribute k to be denoted ξ_{kt}, such that $\sum_{k \in K} \xi_{kt} = 1$ for all k, t:

[10] $\frac{\partial^n F(\cdot)}{\prod_{j \in J} a_{ik}} > 0$, $\forall J$ such that $J \subset K$ and $|J| = n$, and $\forall n$ such that $n \leq |K|$.

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

For simplicity, assume the weights are the same across time, and so suppress the t subscript. Taking the natural log yields the following expression for the productivity fundamental:

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

Similarly, the log of the amenity fundamental, which has different weights given by ι_k, is

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

Given further assumptions on the attributes of finite second moments, strictly positive attribute values, and weak dependence between attributes within each location (demonstrated in Table 1), 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 Table 2) will be inherited by the resulting fundamental terms. A simple simulation confirms the resulting lognormality and spatial correlation of the Figure 2 for an illustration of a random draw of 100 arbitrarily distributed, spatially correlated attributes leading to a lognormal distribution of fundamental that exhibits spatial correlation.

The emergence of a lognormal distribution 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

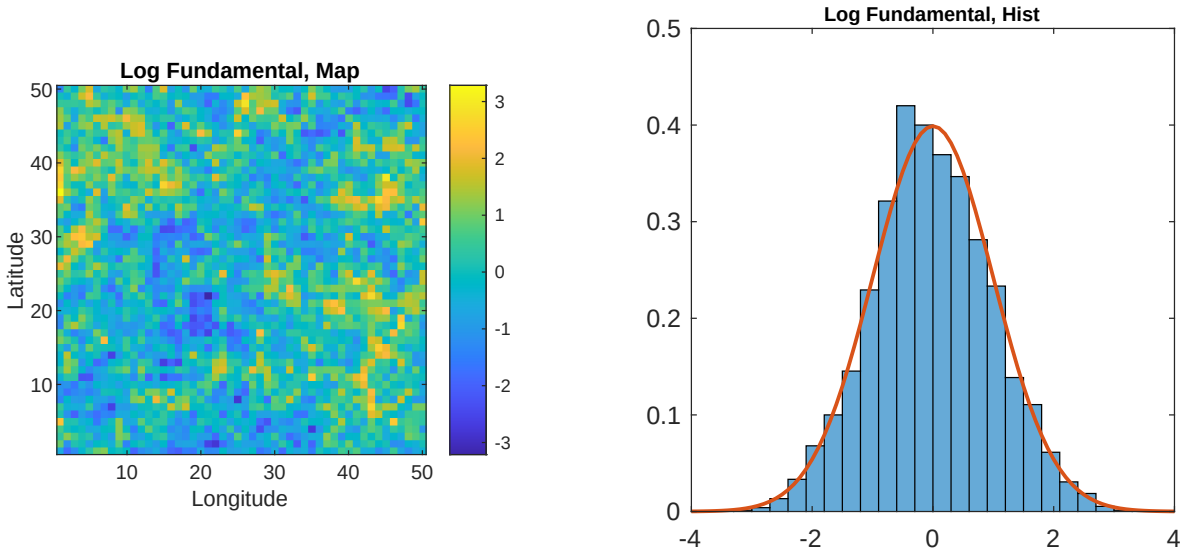


Figure 2: Spatially-Correlated Lognormal Fundamental Draw

driving factors.[11] Indeed, there is strong evidence that total rainfall measures are lognormally distributed (Kedem and Chiu, 1987). 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 (Limpert et al., 2001).

2 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). We characterize the equilibrium population distribution within this model, showing that populations will be lognormally distributed.

[11] Provided frequency, intensity, and duration have sufficiently weak correlation.

The world consists of N discrete locations within a discrete grid with a Euclidean distance metric. Locations are indexed by $n \in N$. Trade between locations is costly. The trade cost function is given by $\tau : N \times N \to [1, \infty)$, 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 of each location. Aggregating over these attributes and allowing for spatial correlation of 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 the composite fundamentals as

$$\tilde{U}_i = U_i L_i^{\beta} \tag{1}$$

$$\tilde{A}_i = A_i L_i^{\alpha} \tag{2}$$

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 τ 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.[12]

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

[12] Allen and Arkolakis (2014) also include continuity in their definition of a regular equilibrium in a continuous geography. Our geography is discrete.

$$W_i = \left(\sum_{n \in N} q_{n,i}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tilde{U}_i \tag{3}$$

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.[13] 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 $L : N \to R_+$ and $w : N \to R_{++}$.

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

$$q_{i,n} = Q_n \left(\frac{p_{i,n}}{P_n} \right)^{-\sigma} \tag{4}$$

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

$$P_n = \left(\sum_{i \in N} p_{i,n}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{5}$$

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

$$p_{i,n} = \frac{\tau_{i,n} w_i}{\tilde{A}_i} \tag{6}$$

Combining the quantity (equation 5) and price (equation 6) expressions, we can write the value of the good produced in i consumed by n as

$$X_{i,n} = \left(\frac{\tau_{i,n} w_i}{\tilde{A}_i P_n} \right)^{1-\sigma} w_n L_n \tag{7}$$

We can re-write the price index as

[13] For discussion, see the online appendix to Allen and Arkolakis (2014). As they demonstrate, the model presented here nests several other QSE models.

$$P_n = \left(\sum_{i \in N} \tau_{i,n}^{1-\sigma} w_i^{1-\sigma} \tilde{A}_i^{\sigma-1} \right)^{\frac{1}{1-\sigma}} \tag{8}$$

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

$$W_i = \frac{w_i}{P_i} \tilde{U}_i \tag{9}$$

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

$$w_i L_i = \sum_{n \in N} X_{i,n} \tag{10}$$

Additionally, the labor market clears as given by

$$\sum_{n \in N} L_n = \bar{L} \tag{11}$$

We can then combine the welfare expression (equation 9, value of consumption expression (equation 7), and income expression (equation 10) to get

$$L_i w_i^{\sigma} = \sum_{n \in N} W_n^{1-\sigma} \tau_{i,n}^{1-\sigma} \tilde{A}_i^{\sigma-1} \tilde{U}_n^{\sigma-1} L_n w_n^{\sigma} \tag{12}$$

The welfare expression combined with the price index yields

$$w_i^{1-\sigma} = \sum_{n \in N} W_i^{1-\sigma} \tau_{n,i}^{1-\sigma} \tilde{A}_n^{\sigma-1} \tilde{U}_i^{\sigma-1} w_n^{1-\sigma} \tag{13}$$

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:

$$\bar{W}^{\sigma-1} L_i w_i^{\sigma} = \sum_{n \in N} \tau_{i,n}^{1-\sigma} A_i^{\sigma-1} U_n^{\sigma-1} L_n w_n^{\sigma} \tag{14}$$

$$\bar{W}^{\sigma-1} w_i^{1-\sigma} = \sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} w_n^{1-\sigma} \tag{15}$$

which can be represented in matrix form.[14] As spatial correlations demonstrated the first law of geography, these equilibrium conditions reflect the "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. We now characterize the resulting population distribution, which for each location depends on both local fundamentals as well as those of its neighbors.

Proposition 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 proposition 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 15 established in part (ii) given by

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

where $g_{n,k} = w_n^{1-\sigma}$ is the guess for place n at step k, and the full vector of guesses at stage k of the iterative mechanism is given by $\mathbf{g}_k$, and $\mathbf{K}$ is a matrix given by

[14] Representation provided in the appendix.

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

For any valid initial guess $\mathbf{g}_0$ (all $g_{n,0} > 0$)), iterating over this procedure will result in convergence to the true $\mathbf{g}^*$. This procedure for any individual $i \in N$ is given by

$$g_{i,k+1} = \frac{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} g_{n,k}}{\sum_{l \in N} \sum_{n \in N} \tau_{n,l}^{1-\sigma} A_n^{\sigma-1} U_l^{\sigma-1} g_{n,k}} \tag{16}$$

The convergence of the model to approximate lognormality, given spatially correlated lognormal draws for productivity and amenity fundamentals, can be seen through this iterative procedure. Consider a particular guess, where $\mathbf{g}_0$ is defined such that $g_{n,0} = \bar{g} > 0, \forall n \in N$. This will result in $g_{i,k}$ being expressible as

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

Now, consider the nature of the remaining exogenous terms, in particular the random variables $A_{(\cdot)}$ and $U_{(\cdot)}$. We have τ as a series of constants capturing the distance between locations, while $A_{(\cdot)}$ and $U_{(\cdot)}$ are drawn from spatially correlated lognormal distributions. Lognormality is maintained under exponentiation and multiplication by constants (τ and $\bar{g}$), so the numerator consists of a sum of correlated lognormal terms. The denominator consists of a sum over sums of correlated lognormal terms.

The distribution of the sum of correlated lognormal random variables is an important and well-studied distribution, given the appearance of sums of correlated lognormal random variables across a wide range of fields such as electrical engineering, finance, and communications. No closed-form solution for this distribution exists. However, a well-known and highly successful approximation for the sum of correlated, lognormally distributed random variables

is with another lognormally distributed random variable, with moments calculated from the moments of the random variables that enter the sum. This approximation has been used in the mathematics and statistics literature for nearly 100 years (for a review, see Dufresne (2009)), with several refinements to methods of constructing the approximation over this period (Fenton, 1960; Schwartz and Yeh, 1982; Mehta et al., 2007).[15] Recent progress in the study of sum of correlated lognormals has developed approximations with only minor deviations from the true distribution of the resulting sum (Lo, 2012, 2013).[16]

We apply this approximation to describe the equilibrium population distribution. When applying this approximation, note that the numerator of the iterative procedure consists of a series of correlated lognormal random variables and will be well-approximated by a lognormal distribution. The denominator, likewise, will also be well-approximated by a lognormal distribution after applying this approximation twice. The quotient of two lognormal distributions will also be lognormally distributed.[17] That is, for all $i \in N$, the distribution of $g_{i,k+1}$ will be well-described by a lognormal distribution.

As the iterative process continues, given the near-lognormality of each $g_{i,k+1}$, the distribution of $g_{i,k+2}$ will also be well-described by a lognormal distribution as lognormality of each term in the summation will be maintained across exponentiation and multiplication. Spatial correlation in the fundamental terms, which will also be inherited by the wage distributions, will ensure the appropriateness of the approximation across iterations

$$g_{i,k+2} = \frac{\sum_{n \in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} U_i^{\sigma-1} g_{i,k+1}}{\sum_{l \in N} \sum_{n \in N} \tau_{n,l}^{1-\sigma} A_n^{\sigma-1} U_l^{\sigma-1} g_{i,k+1}}$$

where $g_{i,k+2}$ will be approximately lognormal as a result of the lognormality of the productivity and amenity fundamentals, their correlation, and the near-lognormality of $g_{i,k+1}$, and so on for subsequent iterations.

[15] Even absent correlation, it has long been noted that sums of lognormals often appear to be well described by a lognormal distribution (Mitchel, 1968).

[16] For example of an application of the Lo (2012) approximation, see the documentation here implementing the approximation.

[17] Given bivariate normality of the logged distribution.

When convergence is achieved, the distribution of elements of the true vector $\mathbf{g}^*$ will be well-described by a lognormal distribution. Wages can be solved from this distribution by exponentiating $\mathbf{g}^*$ as $w_i = g_i^{\frac{1}{1-\sigma}}$), which will preserve lognormality and results in wages following an approximately lognormal distribution. Likewise, populations can be solved for by iterating for the corresponding sequences $f_i = L_i w_i^\sigma$ by solving for $L_i = \bar{L}\frac{f_i g_i^{\frac{\sigma}{1-\sigma}}}{\sum_{n\in N} f_n g_n^{\frac{\sigma}{1-\sigma}}}$. These operations will preserve approximate lognormality for the population distribution as well.

Next, we extend the model by considering the case with spillovers and externalities. Continuing to adapt from Allen and Arkolakis (2014), we now allow 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

$$L_i^{1-\alpha(\sigma-1)} w_i^\sigma = \bar{W}^{1-\sigma} \sum_{n\in N} \tau_{i,n}^{1-\sigma} A_i^{\sigma-1} U_n^{\sigma-1} L_n^{1+\beta(\sigma-1)} w_n^\sigma \tag{17}$$

$$w_i^{1-\sigma} L_i^{\beta(1-\sigma)} = \bar{W}^{1-\sigma} \sum_{n\in N} \tau_{n,i}^{1-\sigma} A_n^{\sigma-1} L_n^{\alpha(\sigma-1)} U_i^{\sigma-1} w_n^{1-\sigma} \tag{18}$$

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

$$L_i^{1-\alpha(\sigma-1)} A_i^{1-\sigma} w_i^\sigma = \phi w_i^{1-\sigma} U_i^{1-\sigma} L_i^{\beta(1-\sigma)} \tag{19}$$

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

$$L_i^{\tilde{\sigma}\gamma_1} = A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \bar{W}^{1-\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} \left(L_n^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \tag{20}$$

where

$$\tilde{\sigma} = \frac{\sigma - 1}{2\sigma - 1}$$

$$\gamma_1 = 1 - \alpha(\sigma - 1) - \beta\sigma$$

$$\gamma_2 = 1 + \alpha\sigma + (\sigma - 1)\beta$$

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

Proposition 2*: Consider a regular geography with overall productivity and amenity functions specified in equations (1) and (2), respectively, and assume that iceberg trade costs are symmetric and parameters are such that* $\gamma_1 > 0$ *and* $\frac{\gamma_1}{\gamma_2} \in (0, 1)$*. 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 proposition 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, a restriction on the parameter space such that congestion forces dominate agglomeration forces, 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 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 and is given by

$$h_{i,k+1} = \frac{A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}}{\sum_{i \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}} \tag{21}$$

where (excluding the iteration subscript k) $h_i = L_i^{\gamma_1}$ and $\lambda = \bar{W}^{\sigma-1}$. Again, consider an initial guess such that $h_{n,0} = \bar{h} > 0$ for all $n \in N$. Then

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

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 and multiplication by $\lambda A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma}$ will result in lognormality of the resulting $h_{i,1}$ for all $i \in N$. In all subsequent iterations, lognormality will again by transmitted until the procedure arrives at the true vector $\mathbf{h}^*$ which will consist of lognormally distributed h_i for all $i \in N$. Solving for the final population distribution as $L_i = h_i^{\frac{1}{\gamma_1}}$ will again result in the population distribution being well described by a lognormal distribution. In Section 4 we demonstrate this result via numerical simulation of the model.

3 The lognormal distribution and Zipf's law

Zipf's law, in its strict formulation as the largest city in a given country being n times the size of the n^{th} largest city, can be represented by a power-law distribution: for a given size S, the probability that a city is larger than S is proportional to $\frac{1}{S}$. This latter formulation implies that cities follow a Pareto distribution with shape parameter $\alpha = 1$ and minimum city size S_m, which gives the aforementioned probability

$$P(Size > S) = \frac{S_m}{S^1} \tag{22}$$

However, this formulation requires there be some minimal "city" and corresponding population below which the population dynamics behave differently.[18] This rigid cutoff 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. Instead, the characteristic power law distribution of Zipf's law for large cities may be a feature of a similar distribution, such as a lognormal, which can be visually indistinguishable from a Pareto in the tail. The lognormal formulation is attractive as it well matches the full population distribution rather than just a small part of it, can be motivated by data and explained using observable characteristics of the world, and also generates key empirical deviations from Zipf's law which are observed in the data that the assumption of a Pareto distribution cannot match.

Having demonstrated that the population distribution inherits an approximately lognormal distribution from spatially correlated locational productivity and amenity fundamentals within a QSE, we now demonstrate that lognormal distributions have Pareto-like tail behavior. The density function of a lognormal distribution is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \tag{23}$$

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},

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

[18] That is, not necessarily by a Pareto distribution.

Writing the constant term $\frac{1}{\sigma\sqrt{2\pi}}$ as Γ, the lognormal distribution thus has the form of a Pareto density function with a varying exponent

$$f(x) = \Gamma x^{-\alpha(x)-1} \quad \text{, where } \alpha(x) = \frac{\ln(x) - 2\mu}{2\sigma^2} \tag{24}$$

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

$$j(x) = \Gamma x^{-\alpha-1} \tag{25}$$

Note that the exponent does diverge, but for large σ^2 from the underlying lognormal distribution, the Pareto exponent is very stable for high draws of x (Malevergne et al., 2011).[19] The distribution is also fairly stable even for $\sigma = 1$ over the 95 to 99.5 percentile. The divergence of the lognormal from the Pareto will generate fewer very large cities than would be expected, a tendency which appears to match the global city distribution (Rossi-Hansberg and Wright, 2007).

Given the near-lognormality of the population distributions, the most populated locations in our model will be distributed approximately Pareto and will appear to follow a power law distribution. According to our model, cities are thus just the tail realizations of the complete human population distribution across all of space. We need to make no additional restrictions on the model in order to observe what appears to be a Pareto distribution for the upper tail of the overall population distribution.

The popularity of the Pareto distribution is a result of its mathematical fit to the exact Zipf's formulation, as the rank statistics of the Pareto distribution with shape exponent 1 exactly generate a straight line for expected city populations in the log-log plot that has -1 slope. Yet, the Pareto assumption necessarily ignores most locations in any population distribution and relies on arbitrary assumptions of threshold population. The Pareto also

[19]The tail of a lognormal distribution is not as thick as that of a Pareto distribution.

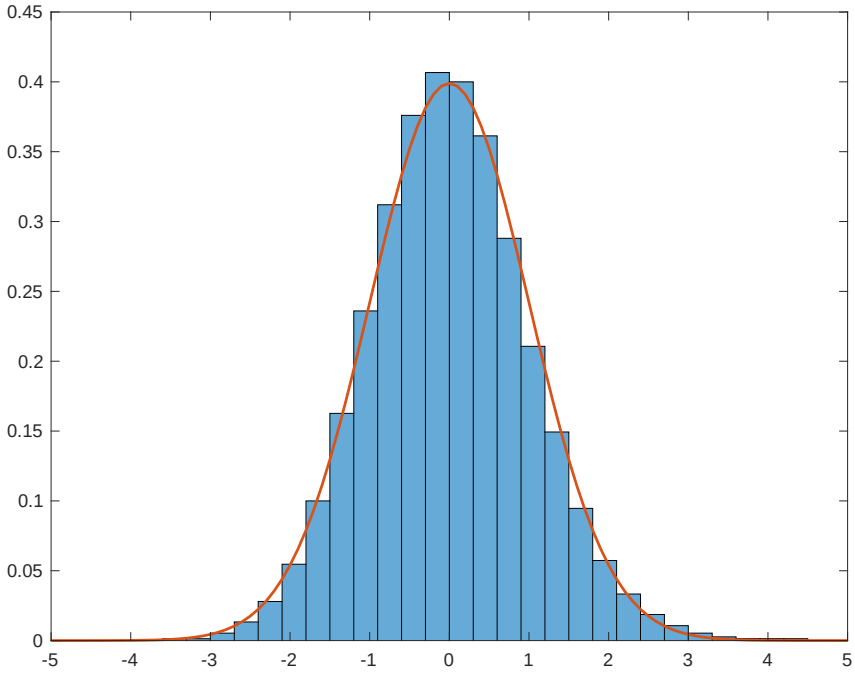


Figure 3: Lognormal Population Distribution (numerical simulation of model)

does not well fit even truncated real-world population distributions, which tend to diverge from Pareto at both the very upper and lower ends of the city size distribution rather than follow a straight line. Viewing Zipf's law as the result of a focusing on the tail of a lognormal distribution requires fewer restrictive or arbitrary assumption and appears to better fit observed data.

4 Results: Zipf's law in QSE and extensions

The simulations in this section were done using code adapted from Redding and Rossi-Hansberg (2017), which is a version of the QSE nested within our current model (proof in the online appendix to Allen and Arkolakis (2014)). We are currently developing simulations based on the updated, more general model and will revise this section ahead of the UEA conference.

4.1 Numerical simulation of the population distribution

To test the reliability of our spatially-correlated QSE model in producing Zipf's Law, we simulate a QSE model based on Redding and Rossi-Hansberg (2017), which is nested by the generalized model described in Section 2, to demonstrate that the resulting population distributions are lognormal. We induced spatial correlation in the productivity fundamental A_i using a Choleski decomposition for a chosen correlation structure.[20] The QSE model simulated here does not have amenity draws (no U_i) but does incorporate a fixed supply of housing, with the model parameterized using consumption share (parameter α in Redding and Rossi-Hansberg (2017)) of $\frac{2}{3}$ and elasticity of substitution $\sigma = 5$, which are standard in literature. Figure 3 shows the equilibrium population distribution, which very closely matches the overlaid normal distribution. The equilibrium population distribution appears lognormal.

Concentrating only on the most populated 10% of locations within the model, we find that the model is very successful at generating Zipf's law-like population distributions. Figure 4 displays smoothed output over 100 simulations of the model. The average slope across the 100 simulations is roughly -1 (-1.074), with a standard deviation of 0.166.

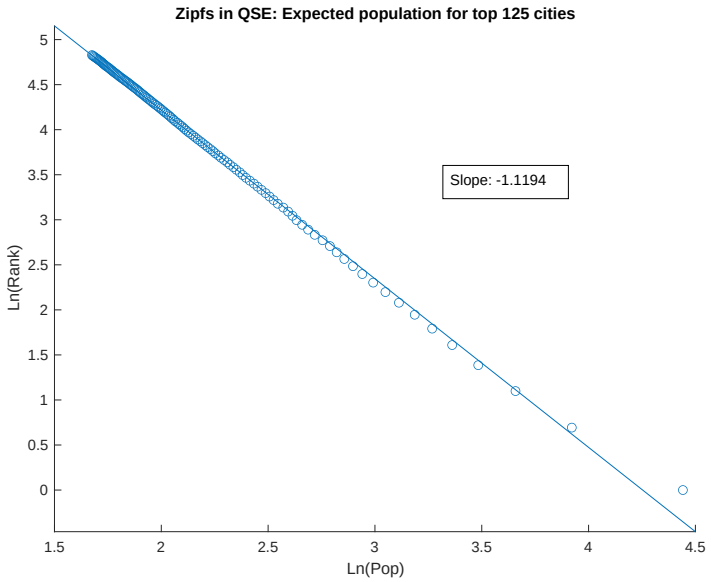
4.2 Gibrat's law

We now demonstrate that the population distribution demonstrates proportional growth in response to increases in the aggregate population $\bar{L}$. We write equation 34 as

$$\tilde{h}_i = \sum K_{i,n} \tilde{h}_n^{\frac{\gamma_2}{\gamma_1}}$$

Where $\tilde{h}_i = h_i \theta^{\frac{1}{1-\frac{\gamma_2}{\gamma_1}}} = h_i \bar{W}^{\frac{\sigma-1}{1-\frac{\gamma_2}{\gamma_1}}}$. Using the labor market clearing condition, as $h_i = L_i^{\tilde{\sigma}\gamma_1}$,

[20] 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.



100-draw MC	Value
Mean	-1.074
Std. Dev	0.166
Median	-1.089
25pct	-1.173
75pct	-0.969
Max	-0.703
Min	-1.515

Figure 4: (Left) Monte Carlo Model Output, (Right) Statistics

$$\bar{L} = \sum_{n \in N} K_{i,n} \tilde{h}_n^{\frac{1}{\bar{\sigma}\gamma_1}} \tag{26}$$

While increasing labor will impact overall welfare given by $\lambda = \bar{W}^{\sigma-1}$, increases will not shift the relative population distribution across space, as the solution for $\tilde{\mathbf{h}}$ is not impacted by changes in $\bar{L}$. This means that a percentage increase in overall population will result in each location experiencing population growth of the same amount, or orthogonality of population growth rates and population size. This establishes that Gibrat's law holds within this model of spatial equilibrium.

This model is static and not suited for the analysis of transition dynamics, but we should note that this result regarding Gibrat's law only exists within equilibrium. This could explain why instances where spatial equilibrium is disrupted or an economy is clearly in transition, as in post-war Japan after U.S. bombing (Davis and Weinstein, 2002) or the Manifest Destiny-era United States with settlement spreading west (Desmet and Rappaport, 2017), reject Gibrat's law for cities even while it has been observed holding in settings that are likely in,

or very near, a spatial equilibrium.

The United States by 1900 was arguably, however, very near a spatial equilibrium. 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 5. 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.

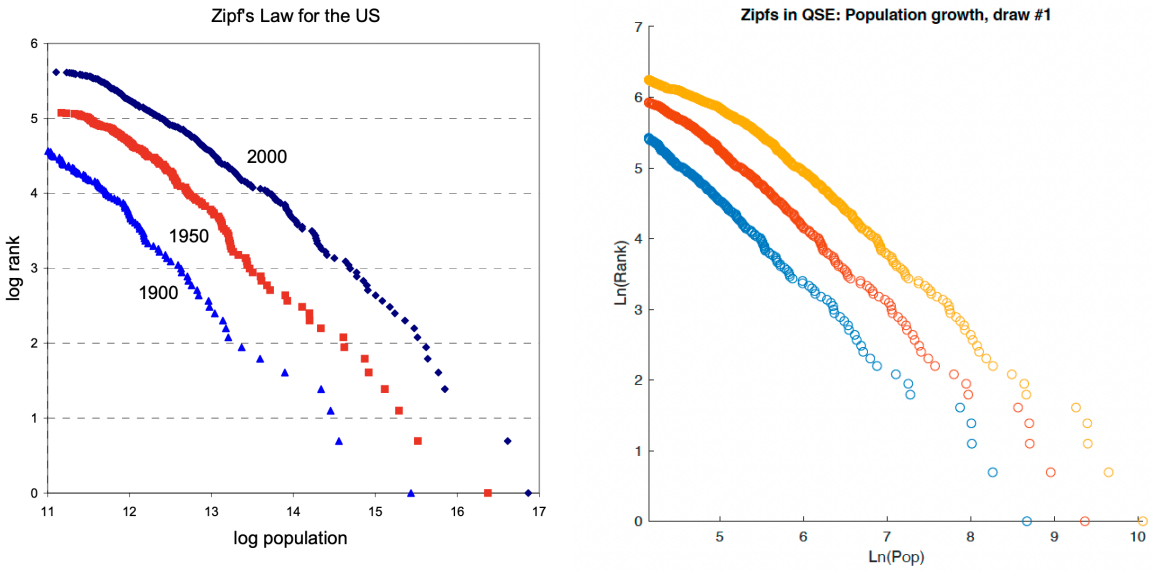


Figure 5: (Left) Rossi-Hansberg and Wright (2007), (Right) Model output

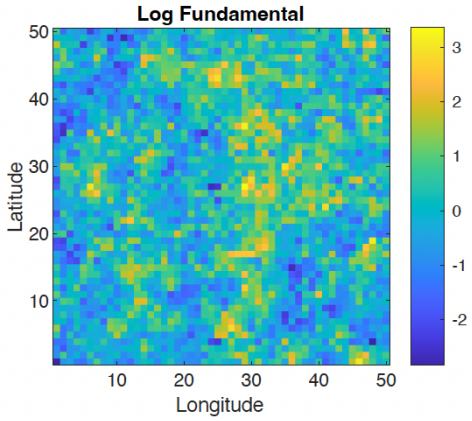
4.3 Central and border cities

One key difference between this project and the prior literature on Zipf's Law is the existence of a role for space, allowing us to consider the impact of spatial distortions on population distributions. A commonly identified tendency in the data is that developed countries have slopes closer to -1, while many developing nations have flatter slopes (Jiang et al., 2014;

Duben and Krause, 2021) with larger-than-expected major cities.

This is not true ubiquitously for developing nations. To be precise, it appears to be a particularly pronounced among African countries. A potential explanation is the distortion caused by "artificial" borders in Africa (using language as in Alesina et al. (2011)). Countries with "natural" borders tend to have central capitals; countries with "unnatural" borders have unusual capital locations. This is the result of the political economy that formed these entities—a naturally forming state would develop around a defensible, highly productive location, while a state formed by a colonial power based on an arbitrary division of land could have its most productive location along one of these divisions and far from the center. Examples of several countries with central or non-central largest cities, and their corresponding Zipf's plots, are in the appendix.

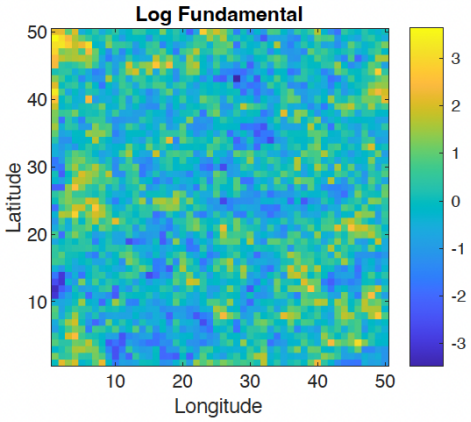
Central capital example



Average over simulations

100x MC	Value
Mean	-1.113
Std. Dev	0.148
Median	-1.116

Border capital example



Average over simulations

100x MC	Value
Mean	-0.942
Std. Dev	0.190
Median	-0.937

Figure 6: Central vs. Border "Capital" example

Within the QSE model with spatially-correlated fundamentals, forcing the most productive location to either be in the center or near a border generates expected deviations in the population distribution. Central capital simulations have steeper slopes, while border

capital simulations have flatter slopes. The latter implies larger than expected large cities; see Figure 6.

5 Conclusion

Zipf's law is a striking empirical regularity in the distribution of populations within countries, and holds across continents and millennia. In this paper, we have shown how economic geography models can generate population distributions that satisfy Zipf's law under general conditions by modeling a more realistic, spatially-correlated distribution of fundamentals. By integrating Zipf's law into spatial equilibrium models, we demonstrate how insights from economic geography theory about the importance of both place and space can bring about this characteristic population distribution. This conception of Zipf's law's emergence better and more reliably captures features of the real world than random growth-based theories that are prevalent in the literature; random growth theories successfully generate similar-looking population distribution, but fail to capture many other elements of the distribution of city populations. Notably, random growth models cannot capture the apparent lack of randomness in the spatial distribution of cities and the tendency of cities to grow rapidly following negative shocks, the latter being a violation of the random growth basis of these models. Additionally, our proposed explanation for Zipf's law and its place in the burgeoning QSE literature offers the ability to consider the role of spatial distortions, locational fundamentals, and networks of and interactions between cities in influencing population distributions.

References

Alesina, A., Easterly, W., and Matuszeski, J. (2011). Artificial states. *Journal of the European Economic Association*, 9(2):246–277.

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.

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.

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

Cristelli, M., Batty, M., and Pietronero, L. (2012). There is more than a power law in zipf. *Scientific Reports*, 2.

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.

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

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.

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*.

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

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*.

Kedem, B. and Chiu, L. S. (1987). On the lognormality of rain rate. *Proceedings of the National Academy of Sciences*, 84(4):901–905.

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*.

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*.

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*.

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.

A Proof of main-text propositions

A.1 Proof of Proposition 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:

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

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

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

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

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

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.2 Proof of Proposition 2, (i) and (ii)

First, we demonstrate that if there is a regular spatial equilibrium then equation 19 is the unique relationship between w_i and L_i. Re-write equation 27 as

$$\phi_i = \frac{L_i^{1-\alpha(\sigma-1)} A_i^{1-\sigma} w_i^{\sigma}}{w_i^{1-\sigma} U_i^{1-\sigma} L_i^{\beta(1-\sigma)}} \tag{27}$$

where we do not yet assume that ϕ_i is the same across $i \in N$. Under the assumption of identical trade costs, we can substitute in equations 17 and 18 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 27 to re-write the above as

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

$$\implies \quad \phi_i = \frac{\sum_{n \in N} F_{i,n} \phi_n^{\beta}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}}$$

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

$$\frac{\phi_i^{\beta}}{\phi_i^{\beta-1}} = \frac{\sum_{n \in N} F_{i,n} \phi_n^{\beta}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}} \tag{28}$$

$$\implies \quad \frac{\phi_i^{\beta}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta}} = \frac{\phi_i^{\beta-1}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta-1}}$$

and define $\gamma_i = \frac{\phi_i^{\beta}}{\sum_{n \in N} F_{i,n} \phi_n^{\beta}}$. Substituting this term in to equation 28 and defining $j_{1,i} = \phi_i^{\beta}$ and $j_{2,i} = \phi_i^{\beta-1}$, we get

$$j_{1,i} = \sum_{n \in N} \gamma_i F_{i,n} j_{1,n} \tag{29}$$

$$j_{2,i} = \sum_{n \in N} \gamma_i F_{i,n} j_{2,n} \tag{30}$$

Both of these can be written in matrix form as

$$\underbrace{\begin{bmatrix} j_{1,1} \\ j_{1,2} \\ \dots \\ j_{1,N} \end{bmatrix}}_{\mathbf{j}_1} = \underbrace{\begin{bmatrix} \gamma_1 F_{1,1} & \gamma_1 F_{1,2} & \dots & \gamma_1 F_{1,N} \\ \gamma_2 F_{2,1} & \gamma_2 F_{2,2} & \dots & \gamma_2 F_{2,N} \\ \dots & \dots & \dots & \dots \\ \gamma_N F_{N,1} & \gamma_2 F_{N,2} & \dots & \gamma_N F_{N,N} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} j_{1,1} \\ j_{1,2} \\ \dots \\ j_{1,N} \end{bmatrix}}_{\mathbf{j}_1} \tag{31}$$

$$\underbrace{\begin{bmatrix} j_{2,1} \\ j_{2,2} \\ \dots \\ j_{2,N} \end{bmatrix}}_{\mathbf{j}_2} = \underbrace{\begin{bmatrix} \gamma_1 F_{1,1} & \gamma_1 F_{1,2} & \dots & \gamma_1 F_{1,N} \\ \gamma_2 F_{2,1} & \gamma_2 F_{2,2} & \dots & \gamma_2 F_{2,N} \\ \dots & \dots & \dots & \dots \\ \gamma_N F_{N,1} & \gamma_2 F_{N,2} & \dots & \gamma_N F_{N,N} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} j_{2,1} \\ j_{2,2} \\ \dots \\ j_{2,N} \end{bmatrix}}_{\mathbf{j}_2} \tag{32}$$

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 ϕ 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 19 is the unique relationship between w_i and L_i in a regular spatial equilibrium.

Given the validity of equation 19, the equilibrium condition is as in equation 20 and can be expressed as

$$\bar{W}^{\sigma-1} L_i^{\tilde{\sigma}\gamma_1} = \sum_{n \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \tau_{i,n}^{1-\sigma} A_n^{\tilde{\sigma}\sigma} U_n^{\tilde{\sigma}(\sigma-1)} \left(L_n^{\tilde{\sigma}\gamma_1} \right)^{\frac{\gamma_2}{\gamma_1}} \tag{33}$$

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

$$\theta h_i = \sum K_{i,n} h_n^{\frac{\gamma_2}{\gamma_1}} \tag{34}$$

$$\bar{W}^{\sigma-1} \begin{bmatrix} L_1^{\tilde{\sigma}\gamma_1} \\ L_2^{\tilde{\sigma}\gamma_1} \\ \dots \\ L_N^{\tilde{\sigma}\gamma_1} \end{bmatrix} = \begin{bmatrix} K_{1,1} & K_{1,2} & \dots & K_{1,N} \\ K_{2,1} & K_{2,2} & \dots & K_{2,N} \\ \dots & \dots & \dots & \dots \\ K_{N,1} & K_{N,2} & \dots & K_{N,N} \end{bmatrix} \begin{bmatrix} \left(L_1^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \\ \left(L_2^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \\ \dots \\ \left(L_N^{\tilde{\sigma}\gamma_1}\right)^{\frac{\gamma_2}{\gamma_1}} \end{bmatrix}$$

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 $\frac{\gamma_2}{\gamma_1}$.

Fujimoto and Krause (1985) provide a generalization of the Perron-Frobenius theorem which extends to non-linear operators of this kind. 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), 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 θ which solve the equilibrium condition. This establishes existence and uniqueness of the equilibrium.

To establish that each location is inhabited if $\gamma_1 > 0$ (establishing regularity), note that combining equations describing the value of trade (equation 7) and total income (equation 10) into the welfare expression (equation 9):

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

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), Fujimoto and Krause (1985) establish that the iterative procedure takes much the same form as that in Proposition 1.

$$\mathbf{h}_{k+1} = \frac{\mathbf{T}(\mathbf{h}_k)}{||\mathbf{T}(\mathbf{h}_k)||} \tag{35}$$

which will converge to the unique solution regardless of the initial guess of $\mathbf{h}_0$. 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

$$h_{i,k+1} = \frac{A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}}{\sum_{i \in N} A_i^{\tilde{\sigma}(\sigma-1)} U_i^{\tilde{\sigma}\sigma} \sum_{n \in N} \tau_{i,n}^{1-\sigma} U_n^{\tilde{\sigma}(\sigma-1)} A_n^{\tilde{\sigma}\sigma} h_{n,k}^{\frac{\gamma_2}{\gamma_1}}} \tag{36}$$

B 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^2
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).

For all correlation calculations, the variables were natural logged.

Spatial Correlation Structure of Geographic and Climate Attributes in the U.S.

Correlations between attributes of centroid grid point and all points in X mile-radius from centroid

	50	100	200	300	400	500	600	1000
Ruggedness	0.77	0.68	0.60	0.55	0.53	0.51	0.50	0.41
Elevation	0.96	0.92	0.86	0.81	0.77	0.74	0.70	0.47
Land Suitability	0.97	0.93	0.89	0.85	0.82	0.80	0.78	0.66
Dist to River	0.97	0.94	0.89	0.83	0.76	0.68	0.60	0.39
Dist to Coast	1.00	1.00	0.99	0.98	0.97	0.95	0.93	0.77
Avg Temp	0.97	0.95	0.92	0.92	0.94	0.92	0.90	0.81
Avg Mon Precip	0.99	0.97	0.93	0.90	0.87	0.85	0.84	0.72
Dist to Harbor	0.99	0.97	0.94	0.91	0.88	0.85	0.83	0.71
Growing Days	0.98	0.96	0.92	0.88	0.86	0.85	0.83	0.78
Malaria	0.93	0.88	0.82	0.73	0.65	0.61	0.57	0.45
Land Area	0.76	0.69	0.64	0.62	0.61	0.59	0.58	0.57

Description: All data at quarter-point latitude and longitude level; variables which are means were averaged over the time period 1960-1990. Values shown are for correlations of grid points in the United States with all grid points in the United States, Canada, and Mexico. All values were logged.

Source: Data as collected by Henderson et al. (2018)

Table 3: Spatial Correlation Structure of U.S. Geographic and Climate Attributes

C County Attribute Analysis

In this section, we complement our exercise from Section 1.1 with U.S. county-level data, in the spirit of Behrens and Robert-Nicoud (2015).

C.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 ordinally structure all county variables from *worst* to *best*.[21] 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.

C.2 Attributes and space

We first calculate cross-correlations between our six attributes for all counties; the results are provided in Table 4. 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

[21] 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 4: Correlation Matrix of U.S. Geographic and Climate Attributes

5. 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 complete this exercise for all counties within a certain radius from centroid counties and for randomly-selected counties in a 20-mile wide ring from the centroid county. In the latter case, the column labeled "400" is randomly-selecting a county 390-410 miles away from the centroid county.

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.

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
Correlation between centroid and all counties in *X* mile disk from centroid								
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
Correlation between centroid and random county in (*X-10, X+10*) mile ring from centroid								
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 5: Spatial Correlation Structure of U.S. Geographic and Climate Attributes

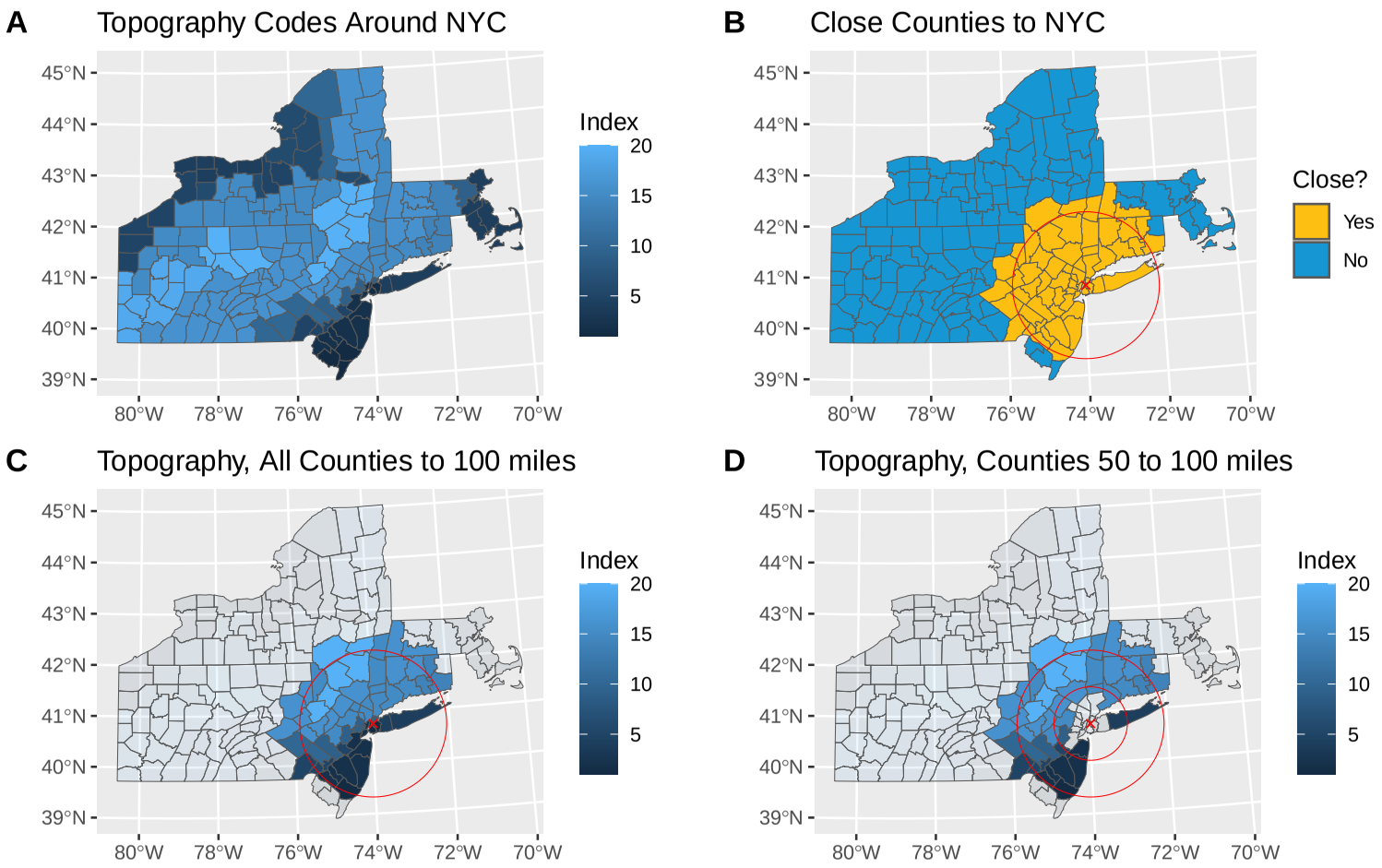


Figure 7: An example of how we calculate spatial correlations. **A:** Topography index for all counties in New York, Pennsylvania, New Jersey, Connecticut, and Massachusetts. **B:** Counties within 100 mile radius of the center (red "X") of New York county (yellow) and all other counties outside that 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

	Dependent variable:			
	Population			
	(1)	(2)	(3)	(4)
January Temp	0.860***	0.583***	0.269***	0.192**
	(0.068)	(0.140)	(0.091)	(0.082)
Jan Hrs Sun	−0.405***	−1.023***	−0.550***	−0.588***
	(0.119)	(0.218)	(0.143)	(0.130)
(Inv) July Temp	2.940***	−4.109***	−0.728	−0.563
	(0.521)	(0.692)	(0.456)	(0.411)
(Inv) July Humidity	−0.460***	−0.754***	−0.297**	−0.233**
	(0.095)	(0.179)	(0.117)	(0.106)
(Inv) Topography	−0.017	−0.011	−0.055***	−0.072***
	(0.030)	(0.032)	(0.021)	(0.019)
Pct Area Water	0.263***	0.175***	0.065***	0.051***
	(0.014)	(0.014)	(0.010)	(0.009)
Land Area	−0.015	−0.052	0.043*	0.070***
	(0.035)	(0.037)	(0.024)	(0.022)
Constant	18.172***	−9.190***	7.754***	9.830***
	(2.056)	(3.216)	(2.120)	(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^2	0.212	0.377	0.736	0.786
Adjusted R^2	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 6: County Population on Attributes

D Additional Extension Figures

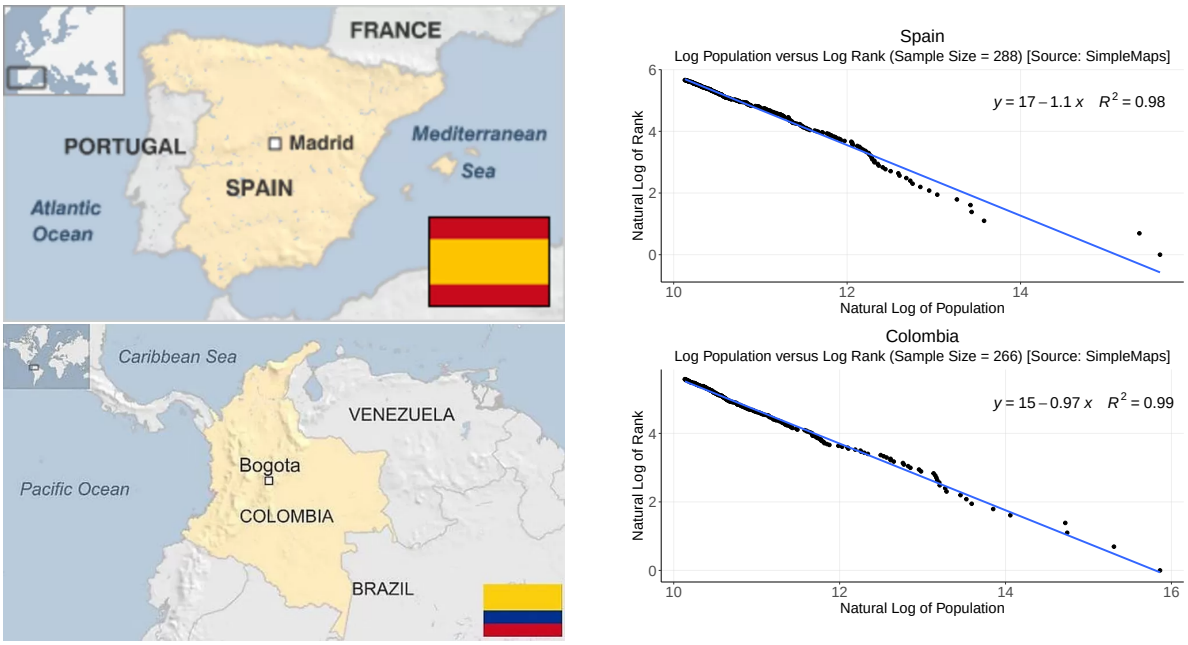


Figure 8: Countries with central capitals (Spain, Colombia)

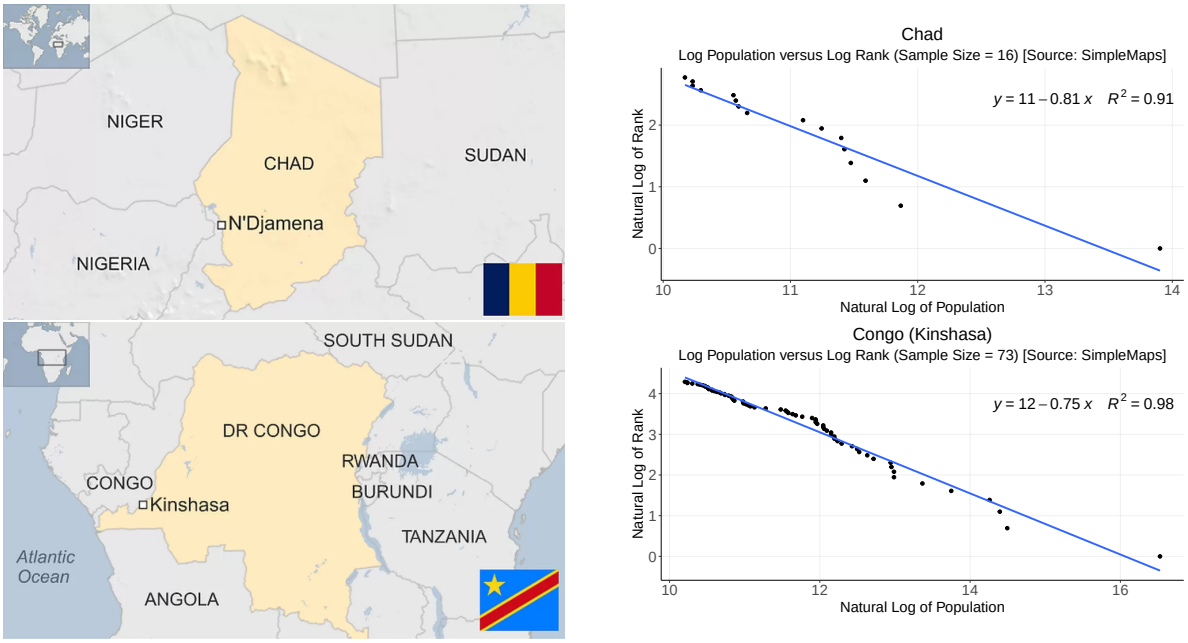


Figure 9: Countries with border capitals (Chad, Democratic Republic of Congo)