Understanding the Fractal Pattern of City Sizes: A Location Demand Theory*

Greg Howard†

October 13, 2023

Abstract

The distribution of the population of cities in the United States has a fat tail. I show that this pattern extends to cities in smaller geographic areas such as Census regions, divisions, and states, indicating a fractal nature. To explain this phenomenon, I propose a location-demand-based theory in which people have heterogeneous geographic attachment. I illustrate the mechanism with a toy model and then estimate an empirical location demand model with this feature. The quantitative model matches key features of migration data and has the ability to generate the observed tail distribution. An implication of the theory is that the elasticity of population to shocks or policies is larger in bigger cities.

Keywords: Pareto tails, migration, location choice

JEL Codes: R12, R23

*I would like to thank Vivek Bhattacharya for helpful comments.
†University of Illinois Urbana-Champaign, glhoward@illinois.edu.

fractal: a curve or geometric figure, each part of which has the same statistical character as the whole.

Oxford Languages (2023)

1 Introduction

The size of a city is the aggregation of many individual's decisions over where to live, a complex choice that depends on many factors, the factors of which, in turn, depend on the individuals. Despite this complexity, these aggregations yield striking patterns. The most famous pattern is the adherence of the largest cities to Zipf's Law (Auerbach, 1913; Zipf, 1949), where the population of a city is roughly inversely proportional to the rank of its size.

In this paper, I highlight and then explain an extension of Zipf's Law: that the city-size distribution has a fractal property. Within sub-regions of the U.S., such as halves, Census regions, Census divisions, or even states, cities continue to adhere to Zipf's Law.[1] I offer a location choice theory that can explain this fractal pattern. The theory's main ingredient is that there is heterogeneity in geographic attachment: some people value living in specific locations or regions a lot, and others value it less. Including this property in an otherwise standard location choice model is able to generate the fractal pattern.

While I am not the first to look at sub-regions to examine Zipf's Laws,[2] I believe I am the first to examine multiple levels and to describe it as a fractal property. To do this, I use the populations of core-based statistical areas in the year 2000, which the Census defines as a group of counties with strong commuting ties and an urban area of at least 10,000 people. Using pre-defined sub-regions, I look for whether Zipf's Law holds in these regions, and I

[1]The exact extent to which Zipf's Law holds is controversial in the literature, at least since Eeckhout (2004), which argued that a log-normal distribution fit the entire city-size distribution better. However, it is relatively uncontroversial that the right-tail of the city-size distribution has a fatter tail than log-normal, which is the main fact I wish to focus on. In that framing, I am establishing that there is a fat tail at many levels of sub-regions as well. See Nitsch (2005) and Arshad, Hu and Ashraf (2018) for surveys of available evidence on to what extent Zipf's Law is true.

[2]See Giesen and Südekum (2011), Subbarayan (2009), Kumar and Subbarayan (2014), Ziqin (2016), and Ye and Xie (2012) who look for Zipf's Law in sub-regions of Germany, India, and China.

find that it does across four levels.

Next, I examine whether existing models can easily account for this phenomenon. While it is straightforward to generate the fractal pattern in a conventional model by setting appropriate utilities for cities, it begs the question of what underlies a similar pattern that arises for the utilities themselves.

I take a different approach and propose a toy model that can generate Zipf's Law for cities through the inclusion of heterogeneity in location attachment. In this model, people differ in the maximum distance they are willing to move from their base location, and they choose the best location within their permissible range. In the model, the fractal pattern emerges as larger cities attract people from further away. The resulting Pareto distribution, which is equivalent to Zipf's Law, arises because as higher-utility cities not only draw nearby people with strong location attachment but also far-away people with weak location attachment.

I show suggestive evidence of this mechanism using data from the 2000 Census. Cities with larger populations attract a larger share of people from outside their state, Census division, Census region, and half of the country. Additionally, when focusing on the individuals, I find that people who move further away are more likely to choose larger cities. People who live in their birth state are 10 percent more likely to live in a city that is 10 percent larger. But people that move out of state are 11.5 percent more likely to live in a city that is 10 percent larger.

Finally, I integrate attachment heterogeneity into a conventional location choice model. Specifically, I introduce the assumption that people could have heterogeneous utility weights on the distance from their birthplace, as well as different weights on unobserved idiosyncratic factors. Using the the 2000 Census data, I estimate the model and find that substantial heterogeneity in these weights better matches the migration data. According to my model, people at the 25th percentile of attachment are three times more elastic to distance from their birthplace compared to those at the 75th percentile.

Because it features attachment heterogeneity, the model is able to match the fact that people who move further away are more likely to live in larger cities. More excitingly, it is also able to generate the fat tail of the city-size distribution without assuming that the baseline utility of cities has a fat tail. In fact, the estimated utilities in the model have almost no excess kurtosis compared to a normal distribution.

In addition to being able to explain an interesting phenomenon, the proposed theory has significant implications for urban economics. In particular, the model estimates different population elasticities to utility for each city. This elasticity increases with city size, with the largest cities about 50 percent more elastic than the smallest cities. This result is an important consideration when considering differential impacts of policies across space.

The central feature of my theory is heterogeneity in location attachment, which to my knowledge is a novel feature for trying to explain city sizes. However, in other contexts, there exists empirical evidence of such heterogeneity. In particular, researchers have looked for such heterogeneity that differs by observable features of people. For instance, Diamond (2016) finds that skilled workers exhibit less attachment than unskilled workers, and Cadena and Kovak (2016) finds that immigrants have less attachment than natives.[3] Neither paper considers heterogeneity beyond observable characteristics, but given that it differs with observables, it likely also differs with unobserved characteristics of people.[4]

Given that Zipf's Law is a well-known fact about cities, there exist several explanations in the literature. Most of these are dynamic, which is a different approach to the explanation I provide. Most well known is Gabaix (1999), which shows that if all cities experience random-walk proportional growth, it will converge to a Pareto distribution.[5] This proportional growth is known as Gibrat's Law (Gibrat, 1931) and has been the focus of a large empirical literature since then (e.g. Modica, Reggiani and Nijkamp, 2017; Berry and Okulicz-Kozaryn,

[3]In a sense, Albert and Monras (2022) microfounds a reason that immigrants are particularly unattached to specific locations because much of their consumption occurs via remittances.

[4]Molloy, Smith and Wozniak (2011) documents that migration is also quite heterogeneous by various observable characteristics. To the extent that observed migration is related to attachment, this is additional evidence that there is significant heterogeneity.

[5]Others include Córdoba (2008a,b); Duranton (2006, 2007); and Rossi-Hansberg and Wright (2007).

2012; Malevergne, Pisarenko and Sornette, 2011).[6]

The dynamic view of Gabaix (1999) and others is complementary to the view that I espouse here. There is no inherent contradiction between the two setups: the location choice model I propose could easily be extended to a dynamic framework that features Gibrat's Law (see Section 5.7). However, I hope that the framework I propose is more helpful for microfounding Zipf's Law in a more typical urban framework—which is often static.

More related to this paper are only a handful of papers that consider the city-size distribution as a consequence of location choice models (Behrens, Duranton and Robert-Nicoud, 2014; Lee and Li, 2013; Hsu, 2012). Most related is Behrens et al. (2014) which postulates that Zipf's Law is the consequence of sorting across talent to different sized cities. To the extent that the heterogeneity in location attachment in my model is correlated to talent in their model, the results are related. However, there is no sense of geography in Behrens et al. (2014) so it is not able to consider a fractal pattern. Further, they propose a full model of production in cities and focus most of the empirical tests on that. My paper emphasizes the location choices of workers and focuses empirical tests on that.

Hsu (2012) is also quite related, proposing that cities specialize in goods that are heterogeneous in returns to scale, rather than my model which focuses on how far people are willing to move. While the mathematical underpinnings have some parallels, the economics of the two theories are quite different, with Hsu (2012) focused on heterogeneity in goods rather than people.

Pareto tails are also studied in a variety of other contexts, e.g. Luttmer et al. (2015), Jones and Kim (2018), and Gabaix, Lasry, Lions and Moll (2016) for income and wealth.[7] To explain Pareto tails, economists almost always use a version of random exponential growth or

[6]Of course, Gibrat's Law generates the fractal pattern for free, since any subset of cities will also follow a Pareto distribution. The theory in this paper is orthogonal to Gibrat's Law, in the sense that it is not inconsistent with it, nor does it imply it. I show this in Section 5.7.

[7]Ironically, the term "superstars" is often applied to models that can generate fat-tailed labor income distributions. In urban economics, while the term "superstars" applies to the best cities, it is not usually focused on the distribution of the size of the cities, but rather just refers to a set of cities with high income, education, and prices (to be fair, these are typically large cities).

they assume heterogeneity in the incentives to accumulate income or wealth. My approach here is the latter, which—to my knowledge—has not been applied to cities.

2 The Fractal Pattern of City Sizes

Auerbach (1913) and Zipf (1949) documented what is known as Zipf's Law for Cities,[8] that the rth largest city is about $1/r$ the size of the largest city. This section argues that Zipf's Law holds at smaller geographies of the country, meaning that the distribution of cities is a fractal.

The analysis is straightforward: I rank the cities by population size and plot the log of the rank versus the log of the population. Zipf's Law predicts that the relationship is linear with slope -1.[9]

Figure 1 shows the pattern for entire country in Panel (a). Each dot represents a core-based statistical area (CBSA), and populations are measured in the 2000 Census. The log rank is on the x-axis, and the log-population is on the y-axis. Consistent with the well-known Zipf's Law, there is a fairly linear relationship between the log rank and the log population, and the slope is close to -1.

In Panel (b), I split the country roughly in half, east and west of the Mississippi River. Within each half, the relationship holds. The two lines of best fit are constrained to have slope -1.

In Panel (c), I split the country into the four census regions. Within each region, the relationship holds.

---

[8]For example, Gabaix (1999) is titled "Zipf's Law for Cities: An Explanation," while acknowledging it was previously discovered by Auerbach (1913).

[9]Throughout this section, I do not do any statistical tests to formally ask whether the true distribution is Pareto. See papers such as Black and Henderson (2003), Bee, Riccaboni and Schiavo (2013), González-Val (2010), Levy (2009), Malevergne et al. (2011), Ioannides and Skouras (2013), and Fazio and Modica (2015) that each try to use statistical tests to figure out whether the tail of the city-size distribution is Pareto, with mixed answers. However, anything close to a linear relationship will indicate a fat-tailed distribution. And what I would like to establish is that these fat tails are a property of the city-size distribution at a variety of geographic levels.

In Panel (d), I split the country into the 9 Census Divisions. Again, a line with slope -1 fits the data fairly well.

Finally, in Panel (e), I split the country into states. I show the 35 states which have at least ten CBSAs. Again, the relationship between city size and rank is fairly linear with slope about -1.

Summing up, it appears that Zipf's Law holds across many varieties of subregions of the United States, at least down to the state-level. This is the definition of a fractal: a pattern that is repeated at smaller scales of the whole.

2.1 Challenges for existing location choice models

The standard location choice model has no problem generating this pattern. However, I argue in this section that the parameters that the standard model uses to match these moments lead to some puzzles.

By the "standard model," I consider the following discrete choice problem:

$$\max u_i = \max_k v_k + \epsilon_{ik} \tag{1}$$

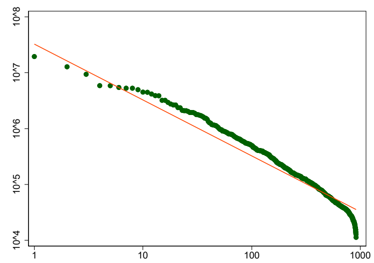
where v_k is the baseline utility of living in city k, that is common to everyone and ϵ_{ik} is the match-specific idiosynchratic utility that depends on the person and the place. The standard assumption is that ϵ_{ik} is idiosynchratic and identically distributed, with an extreme value distribution.

This gives rise to

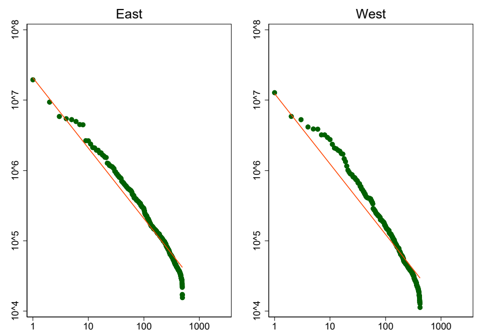
$$\log p_k = v_k + c$$

where p_k is the population of k, and c is a constant across cities that guarantees that the total population across cities is equal to the total population of the country.

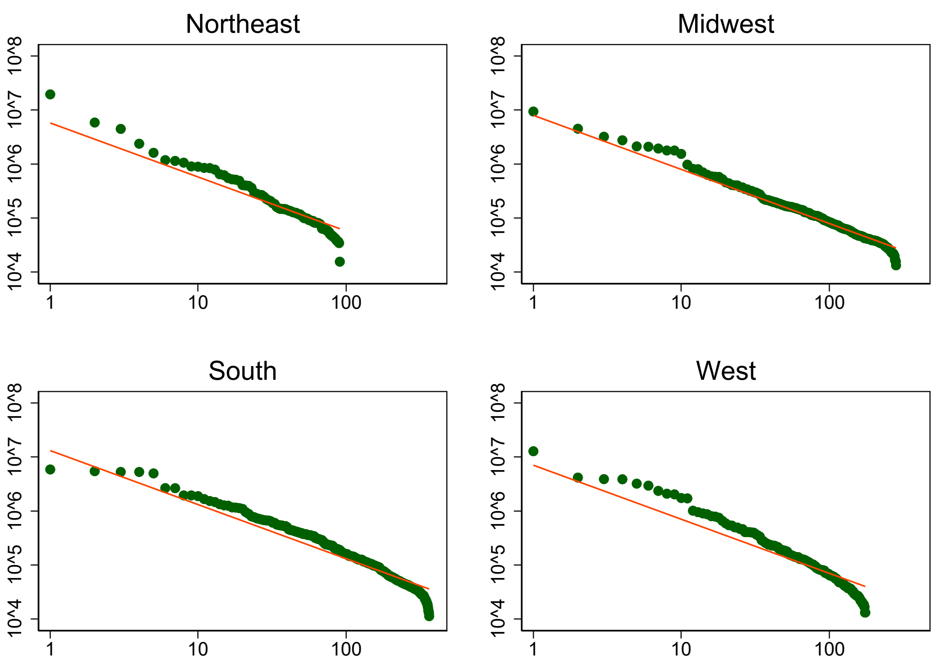
So to match the fact that city populations follow the patterns in the previous section, the standard model only requires that $\exp(v_k + c)$ also follows those patterns. Hence, to explain



(a) Entire Country

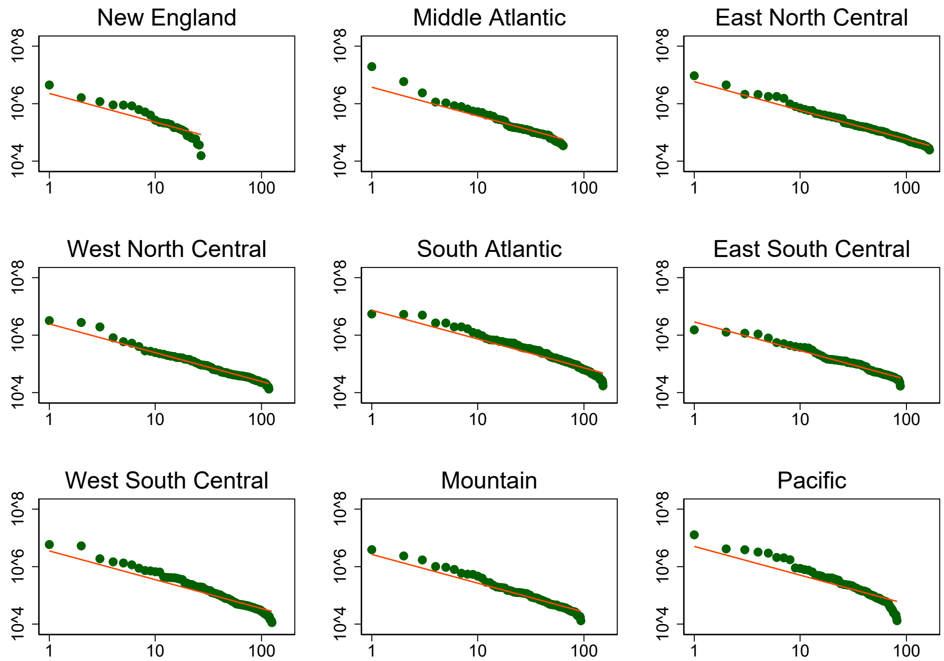


(b) Halves

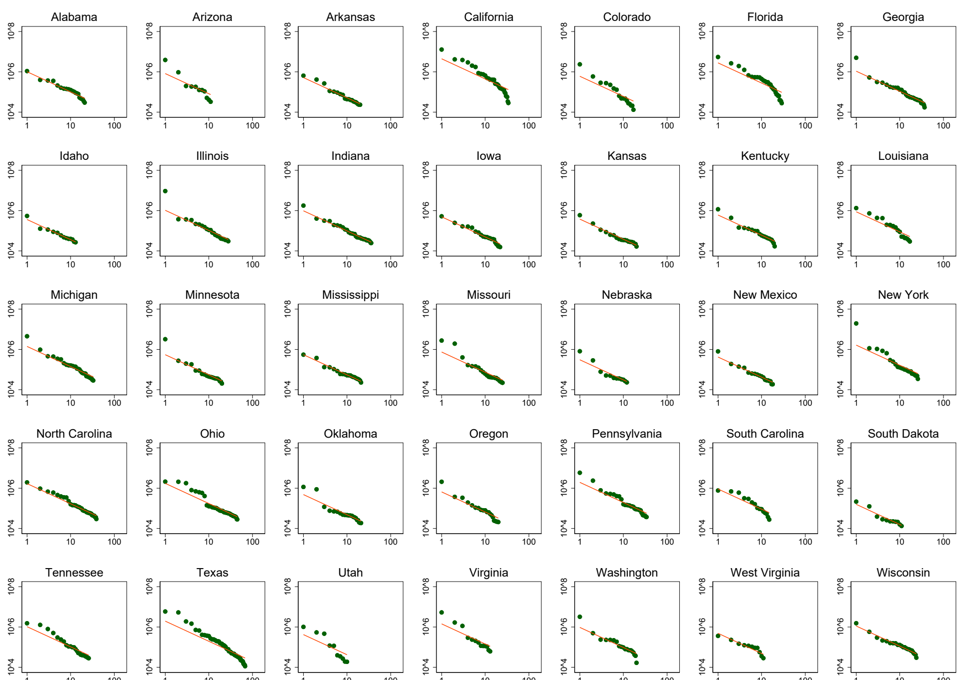


(c) Census Regions

Figure 1: See notes on next page



(d) Census Divisions



(e) 35 States with at least 10 CBSAs

Figure 1: Zipf's Law for city sizes, for different geographies. The x-axis is the rank of the city size within each region, and the y-axis is the population of the city. A city is defined as a CBSA in 2005. Panel (a) shows the whole country. Panel (b) splits the cities into the four Census regions, and Panel (c) shows the nine Census divisions. Panel (d) shows the 35 states with 10 or more CBSAs. The lines of best fit in each figure are constrained to have slope -1.

the fractal patterns of Section 2, we can assume that v_k has a fractal pattern, but instead of a Pareto distribution, it has an exponential distribution. Of course, this begs the question of why v_k would have an exponential tail and a fractal pattern.

In addition, it gives rise to two facts that seem at odds with the data or at least would require further explanation. First, as I just stated, the standard model requires v_k must have an exponential tail. This is somewhat at odds with the fact that there do not seem to be huge differences in consumption or real income in big cities versus small (Diamond and Moretti, 2021; Diamond and Gaubert, 2022), much less a fat tail in consumption across space. Of course, it could be amenities that have an exponential tail or the curvature of the way consumption maps onto utility that generates exponential tails, but these beg the question of why those phenomenon exist.

Second, to generate large cities in every region, it requires a limited amount of spatial correlation in the v_i. Since there is substantial overlap in the distribution of city sizes across all the regions and divisions, that also requires a substantial overlap in the distributions of v_k's, implying zero or negative correlation across space. However, many of the things that we think determined the utility of living in the city are correlated across space. In particular, the recent spatial and trade literature has emphasized the importance of market access, which is inherently correlated across space. In addition, wages, rents, and measures of natural amenities are also highly correlated across space.

While neither of these puzzles is entirely unexplainable, it could also be that different changes to the standard model, that focus on how people choose their locations, rather than the v_k themselves, could help to resolve these puzzles. I show one such change in the following sections.

3 Toy Model of the Fractal Pattern of City Sizes

In this section, we propose a toy model of location choice that can generate the pattern in Section 2.

Consider a continuum of locations, indexed by $k \in [0, 1]$. Each location has baseline utility v_k, which is bounded. Assume that v_k has a well-defined and unique maximum on any interval $[x2^{-n}, (x+1)2^{-n})$ for any whole number n and whole number $x < 2^n$.[10]

Consider a continuum of people endowed with two characteristics, indexed by n and j. Each person is one of N types, indexed by $n = 0, 1, 2, ..., N-1$. Their type determines the range they can move before facing high moving costs. Assume each type has mass 1, so the total population is N. In addition, each person is endowed with a base location j, which is distributed uniformly across locations. n and j are independent of one another.

Agents maximize their utility, which is composed of the baseline utility of the city and a moving cost:

$$u_{nj} = \max_k v_k - \delta^n_{jk}$$

where

$$\delta^n_{jk} = \begin{cases} 0 & \text{if } \lfloor 2^n k \rfloor = \lfloor 2^j j \rfloor \\ \infty & \text{otherwise} \end{cases}$$

where $\lfloor \cdot \rfloor$ is the floor function. So an agent faces infinite moving costs from their base location if the location they consider is not in the same region as their base location, where a region is defined by splitting the number line into 2^n evenly-spaced units.

In other words, people in the model divide the country into regions depending on their type, and will pick the highest-utility city in their region.

[10] A Brownian motion would have this property, so it does not seem that restrictive.

Proposition 1. *Define the size of the r^{th} largest city to be p_r.*[11] *Then:*

$$\lim_{N\to\infty} p_r = 2/2^{\lfloor \log_2 r \rfloor} \approx 2/r$$

Proof: Define an n-region to be an interval $[x2^{-n}, (x+1)2^{-n}]$ for some x. For each n, every location is in one n-interval.

Consider the location with highest v_k. Call it k^0. Everyone of type n that lives in the same n-region moves to k^0. The number of people of type n that move there is the size of the n-interval that k^0 is in. This size is 2^{-n}. Summing over n, the population of k^0 is

$$\sum_{n=0}^{\infty} 2^{-n} = 2$$

Next consider the 1-interval that does not contain k^0 (i.e. if the best location is between .5 and 1, consider $[0, .5)$ and if the best location is between 0 and .5, consider $[.5, 1))$. Consider the best location in this 1-interval, and call it k^1. For all $n > 1$, k^1 will attract everyone that lives in the same n-region. Summing over n, the population of k^1 is

$$\sum_{n=1}^{\infty} 2^{-n} = 1$$

Now let's consider the general case. For any n, there are a total of 2^n n-intervals. 2^{n-1} of those n-intervals have a location that was the best location in an $n-1$-interval. We will ignore those. Only consider the population of the best location in the other 2^{n-1} n-intervals. Call these locations $\{k^n\}$. Since those will attract all the people of type m in the same m-interval, for all $m > n$, the population at those locations will be

$$\sum_{m=n}^{\infty} 2^{-m} = 2^{1-n}$$

[11] Assume ties are broken randomly, or that all cities tied for rank r are given the rank of r.

Hence, there will be 2^{n-1} locations that have population 2^{1-n}.

It is easy to rank the cities in terms of population since the population at k^n is bigger than k^m if and only if $n < m$. Since there are 2^{n-1} locations for each n (except for $n = 0$, which has one city), the rank of a city of type n is strictly greater than 2^{n-1} and weakly less than 2^n for all $n > 1$. Therefore a city of rank r has type $n = \lfloor \log_2 r \rfloor$. And therefore the population of the city of rank r is

$$\lim_{N \to \infty} p_r = 2/2^{\lfloor \log_2 r \rfloor}$$

□

Proposition 2. *Similarly, for any region $[x2^{-n}, (x+1)2^{-n})$, where x and n are whole numbers and $x < 2^n$, the population of the largest city is bounded below by*

$$\lim_{N \to \infty} p_1 \geq 2^{-n}$$

and the population of the r^{th} largest city is

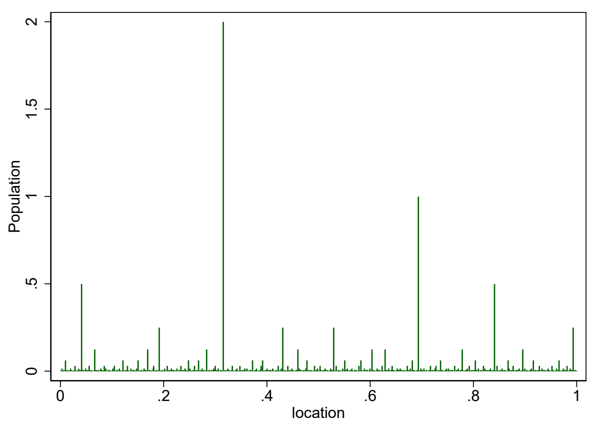
$$\lim_{N \to \infty} p_r = 2^{-n}/2^{\lfloor \log_2 r \rfloor} \approx 2^{1-n}/r$$

for $r \geq 2$.

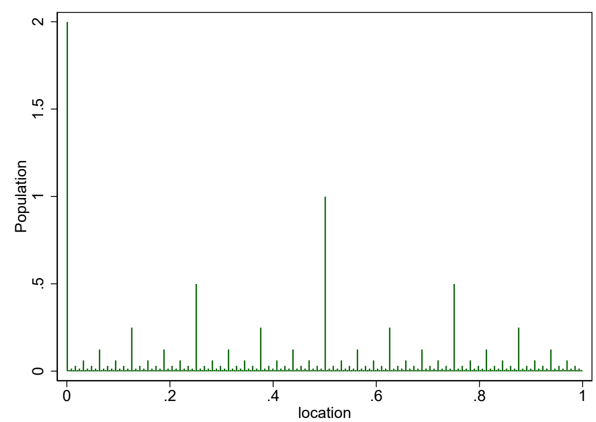
Proof: Consider the best location in the n-interval. If it is the best location in the $n-1$ interval, it will have population that is weakly greater than 2^{n-1}. Otherwise, it will have population 2^{-n}. Hence, the weak inequality for the largest city. For subsequent cities, the proof is identical to Proposition 1. □

The propositions imply that the distribution of population is approximately Pareto, i.e. there is a negative linear relationship between $\log r$ and $\log p_r$. This is true not only at the aggregate level, but also for any given region. Hence, there is a fractal pattern in the city sizes, as documented in the previous section.

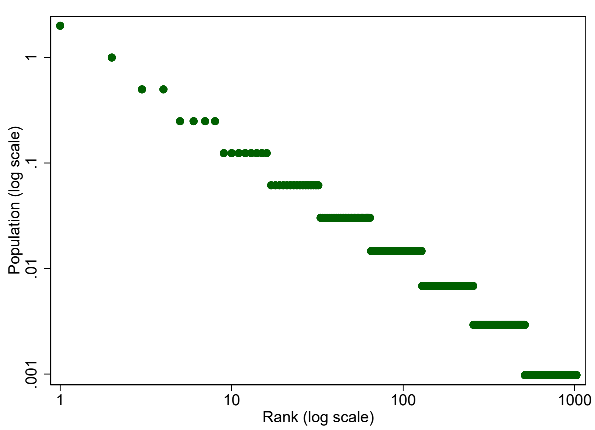
In Figure 2, I show the distribution of population in the toy model for two simulations. For both simulations, I assume $N = 10$, so the most-attached agents will not move outside their $\frac{1}{1024}$-sized region. The least-attached agents will consider the whole line. Panel (a) shows the population in each location, for one random draw of all the v_k, in which I assume they are independently normally distributed. However, the fractal pattern does not depend at all on how I pick the v_k. In panel (b), I show the population distribution if I assume that $v_k = -k$, so that places to the left are always higher utility. In this graph, the fractal pattern is perhaps even more evident to the naked eye.



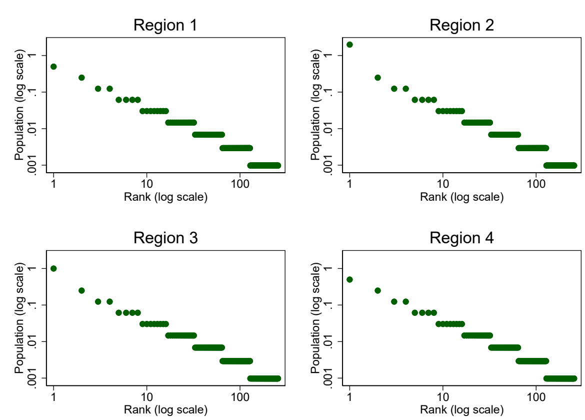
(a) Population distribution over the unit interval, v_k are independent and random



(b) Population distribution over the unit interval, $v_k = -k$



(c) Zipf's Law in the toy model, any v_k



(d) Zipf's Law in the toy model's four 2-regions, v_k correspond to Panel (a)

Figure 2: Simulations of the toy model for $N = 10$

Panels (c) and (d) show that the toy model matches the fractal pattern of the data. In Panel (c), I plot the log population versus the log rank (where I break ties randomly). This graph is identical for either way of generating the v_k. The linear relationship is evident. In Panel (d), I show the same graph, but breaking the interval into four equally sized regions. Except for the rank 1 point, the graphs would be identical for any v_k. For these particular graphs, I use the same v_k that generated Panel (a). This means that not only does Zipf's Law hold on the entire until interval, it also holds for subregions of the interval, as it did in the data.

Of course, the model presented here is extremely stylized, and is not meant to match the way real people choose where to live. Nonetheless, because of the starkness of the assumptions and the conclusions, the toy model can tell us what assumptions are or are not critical to generate the fractal patterns in the real world.

In particular, the model helps resolve some of the challenges that I raised in the previous section. First, the model allows arbitrary spatial correlation in the baseline utilities, while still generating big cities in all regions. Second, the model does not require that the utilities themselves have any particular distribution. In fact, the differences across utility in this model can be arbitrarily small, and utility can have a fat tail, a thin tail, or no tail at all.

What is critical is heterogeneity in attachment. In the model, it is essential that some people need to live very close to their base location, while others are able to live further away.

4 Empirical Evidence of Model's Mechanisms

If the key assumption of the previous model is that people have heterogeneous attachment to regions, we may be able to look for evidence of this in the data. In this section, I show two types of evidence. The first focuses on the cities: bigger cities should have a higher proportion of their population from further away places. The second is complementary but focuses on people: people that move further away are more likely to end up in big cities.

For the first piece of evidence, I measure the share of population in each city that was born in the same state, division, region, or half of the country as the city itself. I then compared across city sizes using a binned scatter plot.

For the second piece of evidence, I run the following regression:

$$\log m_{j\to k} = \beta(\delta_{jk}) \log p_k + \alpha_j(\delta_{jk}) + \epsilon_{jk}$$

where $m_{j\to k}$ is the number of people who were born in location j and live in location k, p_k is the population in k, δ_{jk} is the distance between j and k. If some people have less geographic attachment they are more likely to move further away, and will end up in high-utility cities with lots of population. Hence, we would expect the $\beta(\delta_{ji})$ to be increasing in δ_{ji}.

4.1 Data

For this exercise, I use data from the 2000 Census, available via Ruggles, Flood, Sobek, Brockman, Cooper, Richards and Schouweiler (2023). For each person, the data records the birthplace, by state, and the current location, by public use microdata area (PUMA). I use the geographic correspondence engine from Center (2016) to assign PUMAs to CBSAs. If a PUMA is not entirely within a CBSA, I apportion the person based on the population-weighted fraction of the PUMA in the CBSA. I then aggregate the data to state-CBSA pair, recording the total number of people that were born in state j and moved to CBSA k.

4.2 Results

The first piece of evidence is presented in Figure 3. I split cities into 50 bins of about 18 cities each, and plot the mean share of population that is from out-of-area versus the log population. For all four definitions of area—state, division, region, and half—the share coming from out of that area is increasing in the city size. This is qualitatively consistent with the toy model, where bigger cities could attract people from further away.[12]

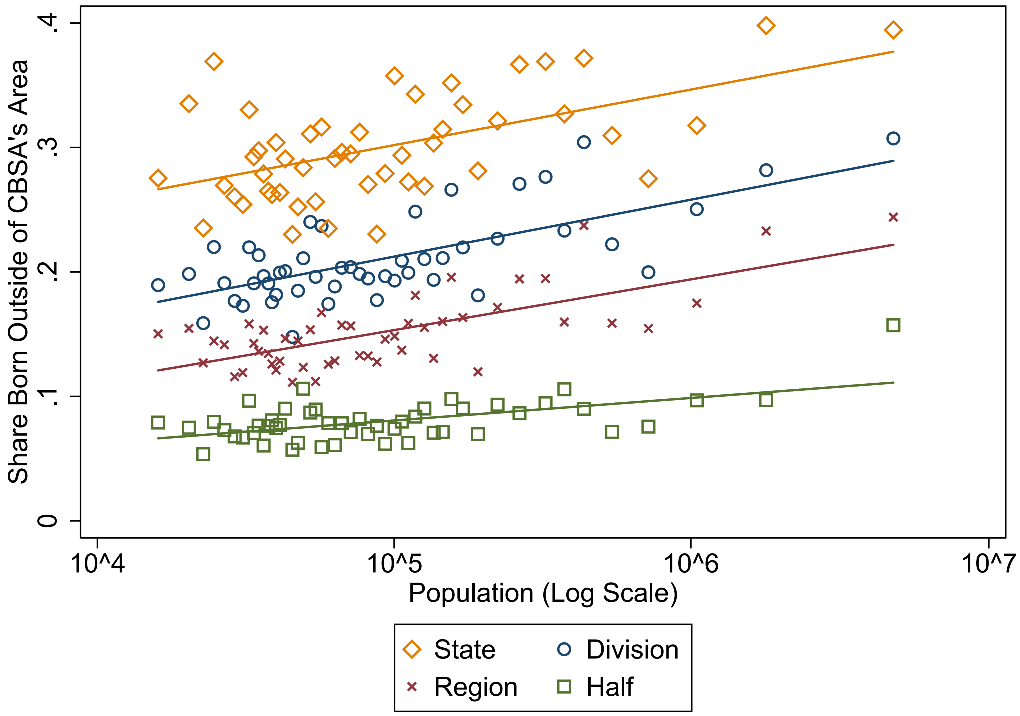


Figure 3: Share of the city that comes from out of the area, where area is defined as the state of the city, the Census division of the city, the Census region of the city, or the half of the country (split by the Mississippi). Cities are split into 50 bins, and the mean of each bin is plotted in the graph, along with a linear fit line. Cities are assigned to an area based on which state the plurality of their residents live in.

This fact is a challenge for a typical model without attachment heterogeneity. If we took the standard model and added homogenous moving costs with respect to distance, it would not be able to generate this pattern. We will show this later in Appendix Figure A3.

We start with a simple split, where we consider two bins of δ_{jk}, based on whether k

[12]Of course, in the toy model, the shares from outside the area in the smallest cities would be zero, but there are good reasons to not take those predictions literally. Here are a few: birthplace may not be the right proxy for base location; The areas as measured in the data may not correspond precisely to the n-regions in the previous section; and the model is a toy model where we might want to take some qualitative predictions seriously but not the quantitative ones. In the next section, we look at a quantitative model of the U.S., and I revisit this figure.

Note also that there is a lot of noise in the figure, even after the binning. This seems to be driven by the fact that some cities are central in their areas while others are on the edges and are able to attract people from nearby states more easily. As long as this centrality is not systematically correlated to city size, it should not bias these estimates.

Table 1: Regression Results

	(1) Log Migration	(2) Log Migration
Log Population	0.993***	1.151***
	(0.0169)	(0.0196)
Observations	907	44427
Sample	Within State	Outside State
Birth State FE	Yes	Yes

Standard Errors Clustered by Receiving Destination

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

is in state j or k is outside of state j. For cities that span multiple states, we assign the state that has the greatest share of the city's population. The results of the regression are in Table 1. If someone lives in the same state they were born, they are 9.9 percent more likely to live in a city that is 10 percent bigger. But if they live outside the state they were born, then they are 11.5 percent more likely to live in a city that is 10 percent bigger. This provides some suggestive evidence in favor of the toy model's main mechanism, that there exists heterogeneity in location attachment.

We can be much more flexible in terms of distance than just inside/outside the birth state. The results of the regression in which we discretize δ_{jk} into 21 bins are presented in Figure 4. In the closest bin, where people are not living that far from their birthplace, people do move to larger cities, with an elasticity non significantly different than 1. But for people that are moving further away, the elasticity of migration to population is higher. This increases between 300 km and about 1500 km, but does decrease a bit for very far distances, which largely involve moving from coast to coast or to or from Alaska or Hawaii. A natural interpretation is that the people that move further are more likely to have less attachment to a specific location, and so are able to pick high-utility areas more than the people that are close to their birthplace, who are more likely to be attached to that area, and care less about utility.

Neither of the two facts in this section are consistent with a standard gravity model in

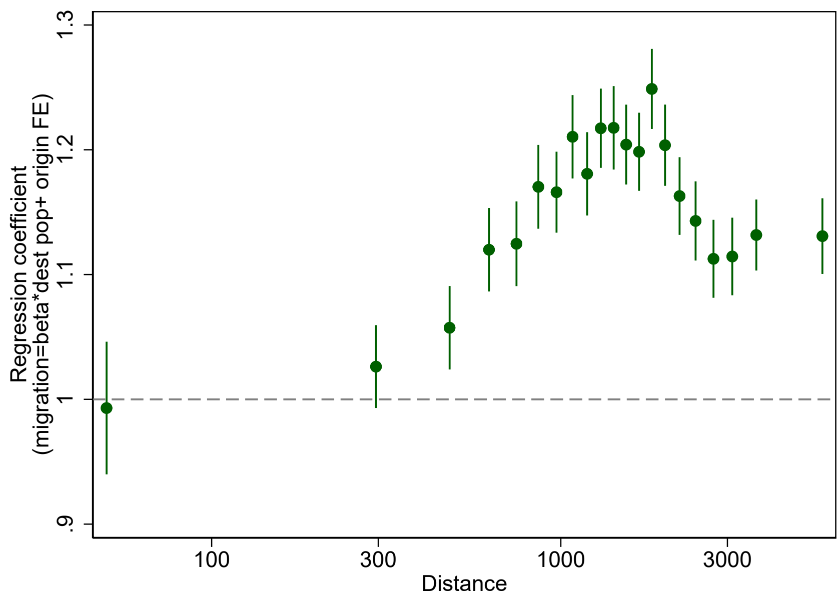


Figure 4: The importance of city size, by distance. The graph shows the coefficients δ_{ij} and the 95 percent confidence interval of the regression $\log m_{j\to i} = \beta_{\delta_{ji}} \log p_j + \delta_{ji} + \epsilon_{ji}$. The furthest left point consists of cities that are in the same state as the state of birth. Outside the state of birth, the distance is measured from centroid to centroid, and binned into 20 groups.

which migration can be expressed as:

$$m_{jk} = X_j Y_k / f(\text{distance}_{jk})$$

where X_j and Y_k are city-specific terms and f is an increasing function. If this equation determines migration, then the first fact must be false: if two cities are in the same location, then the share of the people from any given location j is the same, regardless of what Y_k is. Similarly, if this equation is true, the second fact must also be false. Because for two cities that are the same distance away, the amount of migration will be proportional to Y_k. This proportion does not depend on how far people move. Hence, to match these facts, we will need to consider a model that does not generate a standard gravity equation, such as having heterogeneity in geographic attachment.

5 Incorporating the Mechanism into a Standard Model

In this section, I modify the standard model to have heterogenous attachment, and estimate that model from data. I then look at whether the model can generate the pattern from the previous section, and the fractal pattern that motivates the paper.

5.1 A discrete choice model with heterogeneous attachment

The model is a discrete choice model over possible places to live. As in the toy model, each person has a base location and heterogeneous attachment. The heterogeneous attachment is modeled as different weights that people place on the disutility of living far from home, as well as different weights on the match-specific utility draws.

In math, a person i, born in location j, maximizes their utility by picking a location k to live in according to the following formulation:

$$u_{ij} = \max_k v_k - \beta_i \delta_{jk} + \gamma_i \epsilon_{ik}$$

where v_k is a city-specific indirect utility, reflecting the amenities, wages, and rents in that city, δ_{jk} is the distance from j to k,[13] and ϵ_{ik} is an i.i.d. extreme value draw for each person-city pair. β_i and γ_i are weights that vary by person dictating how important distance and the idiosyncratic term are in their decision.

This model is observationally-equivalent to a model in which we multiply the entire right-hand side by $\frac{1}{\gamma_i}$, which ends up being more tractable to estimate. I assume that β_i and γ_i are drawn from a jointly-log-normal distribution, which I parameterize in the following manner:

$$\begin{bmatrix} \log \beta_i/\gamma_i \\ \log 1/\gamma_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\beta^2 & \rho\sigma_\beta\sigma_\gamma \\ \rho\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix} \right)$$

[13] When we take this to the data, we will use the log distance from the state's centroid to the city's centroid, except for city's that are in the state, for which we will estimate this term.

The 0 is a normalization because of the city-fixed effects. The entire distribution is governed by four parameters: μ which governs the median importance of distance; σ_β, which represents how much variation there is in relative attachment to birth-location compared to the idosynchratic term; σ_γ, which measures how much variation there is in the relative importance of the city-fixed-effect compared to the idiosynchratic term; and ρ which measures how correlated the two terms are.

Intuitively, people with small γ_i will end up moving to high-utility big cities. Similarly, people with small β_i will end up moving further away. People with less attachment in total (low β_i and γ_i) will end up moving further away and to more populous cities, as we saw in the data.

Given the structure of the data, which is the same as in Section 4, we add a parameter to the model, which is that we measure δ_{jk} in the following manner:

$$\delta_{jk} = \log \text{distance}_{jk} \cdot 1_{[k \text{ not in } j]} + \lambda 1_{[k \text{ in } j]}$$

So if the city k is outside of state j, we assume δ_{jk} is the log of the number of kilometers between them, and if city k is in state j, we assume δ_{jk} is equal to λ, which we will estimate.

5.2 Estimation Strategy

There are $5+K-1$ parameters of the model to estimate, where K is the number of cities: μ, σ_β σ_γ, ρ, λ, and each of the v_k, one of which I can normalize. For any given set of parameters, I can solve for the total number of migrants from j to k:

$$m_{jk}^{\text{model}} = p_j \int_{\beta_i} \int_{\gamma_i} \frac{\exp(\frac{1}{\gamma_i} v_k - \frac{\beta_i}{\gamma_i} \delta_{jk})}{\sum_\ell \exp(\frac{1}{\gamma_i} v_\ell - \frac{\beta_i}{\gamma_i} \delta_{j\ell})} f(\beta_i, \gamma_i) d\beta_i d\gamma_i$$

where p_j is the number of people born in j and $f(\beta_i, \gamma_i)$ is the probability density function of the multivariate log-normal distribution. In practice, I discretize the log-normal distribution over a 31×31 grid.[14]

To do the estimation, my objective is to minimize the sum of the squared difference between the log migration from j to k in the model and the data, conditional on exactly matching the number of people born in j and the total number of people choosing k.

$$\min \sum_{jk} (\log m_{jk}^{\text{model}} - \log m_{jk})^2$$

subject to

$$\sum_j m_{jk}^{\text{model}} = \sum_j m_{jk} \quad \text{and} \quad \sum_k m_{jk}^{\text{model}} = \sum_k m_{jk}$$

The estimation itself has two loops to it. In the inner loop, I take as given the μ, σ_β, σ_γ, ρ, and λ, and I estimate all the v_k's through an iterative procedure.

$$v_k^{n+1} = v_k^n + \theta_k \log \frac{\sum_j m_{jk}}{\sum_j m_{jk}^{\text{model}}}$$

where θ_k is the model-implied elasticity of the city's population to v_k.[15] For an initial guess v^0, I iterate until the vector v^n converges. This guarantees that the total number of people moving to k matches in the model and the data.

In the outer loop, I use the MATLAB function fminunc to minimize the sum of the squared differences in migration in the model and the data, over μ, σ_β, σ_γ, ρ, and λ.

[14] I use an evenly spaced grid from -$3\sigma_\beta$ to $3\sigma_\beta$ for $\log \beta_i/\gamma_i$ and an evenly spaced grid from $-3\sigma_\gamma$ to $3\sigma_g amma$ for $\log 1/\gamma_i$.

[15] This elasticity can be solve in closed-form. It is given by the following weighted-average of i-specific elasticities:

$$\theta_k = \frac{\sum_j \int_{\beta_i} \int_{\gamma_i} \frac{1}{\gamma_i} m_{ijk}(1 - m_{ijk}) f(\beta_i, \gamma_i) d\beta_i d\gamma_i}{\sum_j m_{jk}}$$

which m_{ijk} is the migration probability of person i in j moving to k:

$$m_{ijk} = \frac{\exp(\frac{1}{\gamma_i} v_k - \frac{\beta_i}{\gamma_i} \delta_{jk})}{\sum_\ell \exp(\frac{1}{\gamma_i} v_\ell - \frac{\beta_i}{\gamma_i} \delta_{j\ell})}$$

Parameter	Baseline Model	Constrained Model
$\log(\mu)$	2.065	1.02
λ	4.0549	2.13
σ_β	0.8382	.001
σ_γ	0.3263	.001
ρ	-0.4154	0

Table 2: Estimated Parameter Values. In the Baseline Model, all five parameters are unconstrained. In the Constrained model, only the first two parameters are estimated.

5.3 Estimates

The model estimates significant dispersion over both β_i/γ_i and γ_i, as well as a negative correlation between the two. See Table 2 for a listing of the specific values of the five parameters (the v_k will be discussed below).

We can use the estimates of the standard deviations and correlations to solve for the standard deviations and correlation of $\log \beta_i$ and $\log \gamma_i$ directly.

$$\begin{bmatrix} \log \beta_i \\ \log \gamma_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\beta^2 + \sigma_\gamma^2 - 2\rho\sigma_\beta\sigma_\gamma & -\rho\sigma_\beta\sigma_\gamma + \sigma_\gamma^2 \\ -\rho\sigma_\beta\sigma_\gamma + \sigma_\gamma^2 & \sigma_\gamma^2 \end{bmatrix} \right)$$

The standard deviation of $\log \beta_i$ is 1.01. The standard deviation of $\log \gamma_i$ is still 0.33. The correlation between the two is 0.66. So people that have less strong preferences over distance also tend to have less strong preferences over idiosynchratic factors. So the model does a better job matching the data when there are some people that are less "attached" to any specific location.

In particular, there is a lot of heterogeneity in how elastic people are to distance, with the 75th percentile of elasticity being more than 3 times as elastic as a person in the 25th percentile.[16] This is consistent with the toy model's, whose primary feature was that some people were willing to move further than others.

[16]To calculate this, I use the z-scores from a normal distribution and calculate: $\exp((z_{.75} - z_{.25})\sigma_\beta)$.

5.4 Comparing the model to the empirical evidence

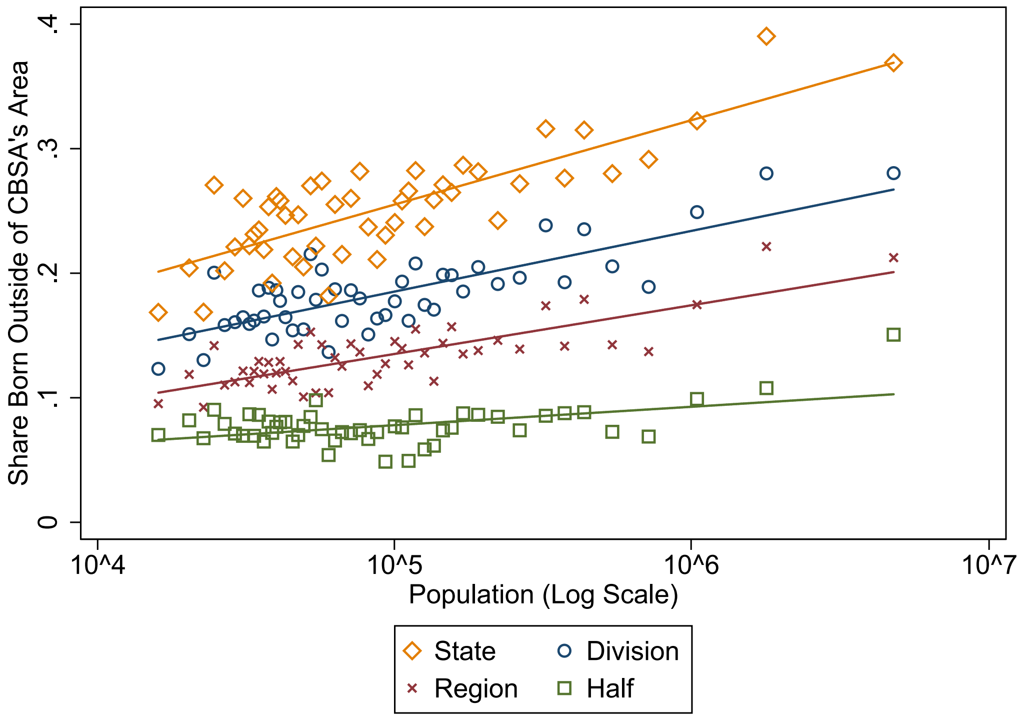


Figure 5: Share of the city that comes from out of the area, in the model

In fact, the model is able to replicate the main features of the data that was highlighted in Section 4. Figure 5 replicates Figure 3, but uses the migration flows implied by the model. As in the data, the share of population from outside the area is increasing in city size and the magnitudes are comparable.

In contrast, in Appendix Figure A3, we replicate the same figure, but using a constrained version of the model in which we do not allow for heterogeneity in moving costs. In that version of the model, the shares are mostly independent of city size.

Similarly, if we run the same regression regarding the proclivity to move to large cities based on distance using the model's output, we get almost the same coefficients as in Table 2.[17] In Figure 6, I replicate Figure 4, using the model's predicted migration, and it also predicts that people that move far from their birthplace are more likely to live in big cities.

[17]The exact coefficients are 0.98 instead of 0.99, and 1.13 instead of 1.15.

In Appendix Figure A1, I show that in a model where I constrain σ_β and σ_γ to be close to zero, the model cannot reproduce this feature.

Intuitively, this is achieved because of the negative correlation between $-\log \beta_i/\gamma_i$ and $\log 1/\gamma_i$. People that care a lot about whether to live in a high-utility city care less about how far away it is.

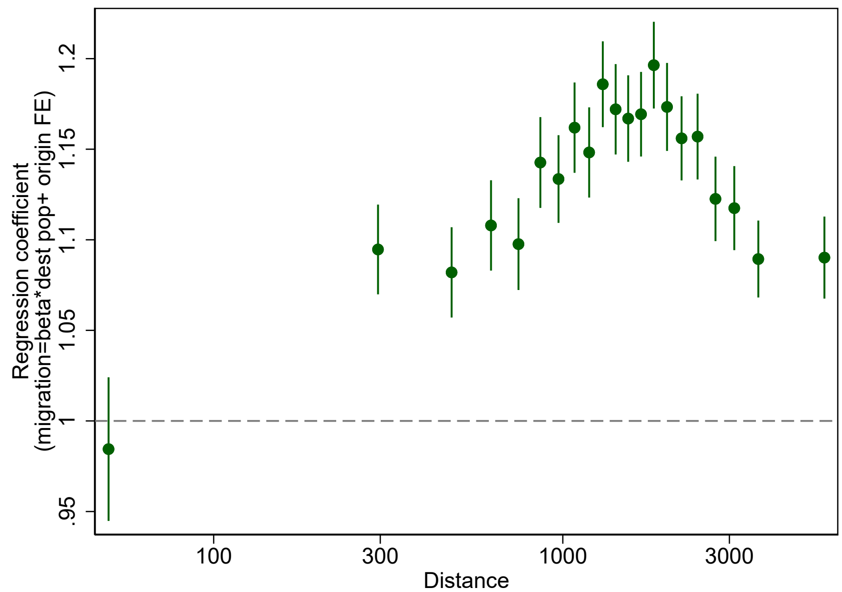


Figure 6: The importance of city size, by distance, in the model

5.5 An explanation of the Pareto tail

The next step is to evaluate whether the mechanisms in the model can generate the fractal pattern in the data. In particular, a major question is whether the model can produce a fat tail. One reason to think that it can is that there is a non-linear relationship between log population and v_k. In Figure 7, I show a scatter plot of the log population versus the v_k. I also include a locally weighted scatterplot smoothing curve, which shows a fair bit of concavity in the relationship between the two variables. This means that at high levels of v_k, a small increase in v_k increases log population more than it does at low levels. This means that even if v_k does not have fat tails, it can create a fat tail in log population.

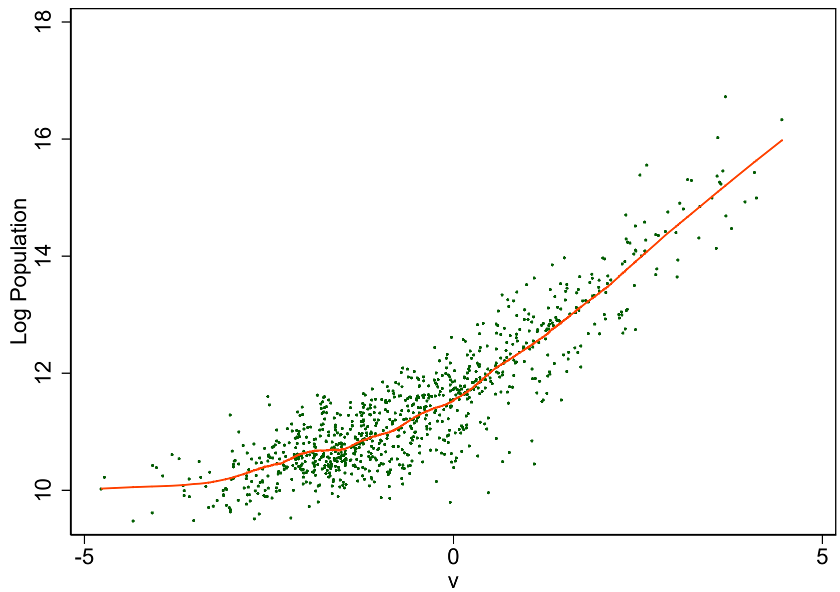


Figure 7: The relationship between log population and v_k.

Indeed, the distribution of v_k that is estimated by the model does not have fat tails. The kurtosis of the distribution is 3.06, only 0.06 more than a normal distribution. To contrast, the kurtosis of the log population distribution is 4.32 (or excess kurtosis 1.32). So what that means is that the model, with its heterogeneity of attachment, is able to generate the heavy tails of the population distribution, without relying on heavy tails in the city utilities.

Without the heterogeneity in attachment (i.e. the constrained model discussed previously), the model is not able to generate the heavy tails on its own. The relationship between log population and v_k is much more linear, which I show in Appendix Figure A2. And the kurtosis of the v_k in the constrained model is 3.62.

5.6 Elasticities in the Model

One of the primary implications of the model is that changes in v_k will have different proportional effects on cities of different sizes. With the model, we can estimate these elasticities directly. In Figure 5, I show the semi-elasticity of population with respect to v_k, against the city size. I also include a locally weighted scatterplot smoothing curve to emphasize the increasing relationship. For the most populous cities, the semi-elasticity is about 50 percent larger than for the least populous city. Note that the average elasticity is not an interesting statistic, since we normalized the mean of the log elasticity to be zero. This means that identical policies that have identical welfare effects in different cities will induce different population responses, by up to 50 percent. It might also help explain the slow decline of cities because as they get smaller, their residents are less likely to move away in response to continued shocks.

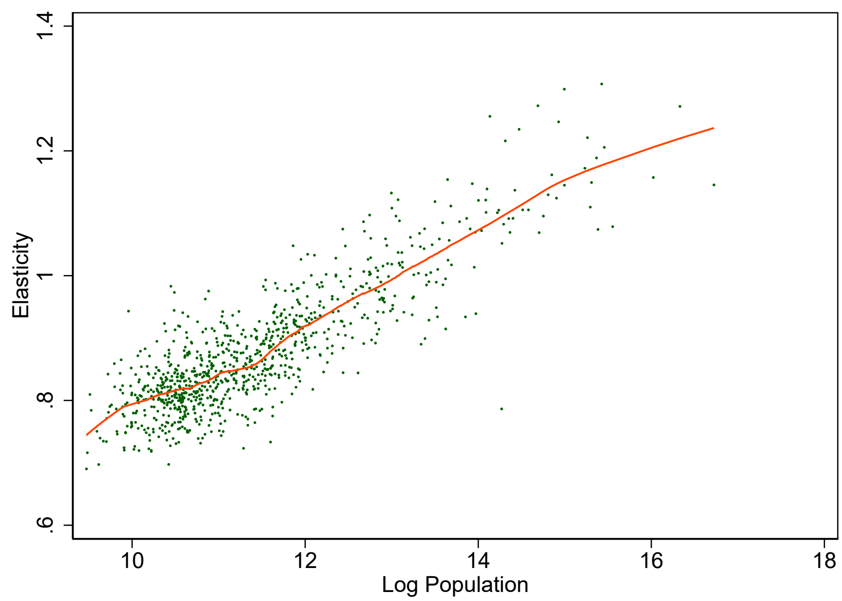


Figure 8: The relationship between population semi-elasticity to v_k and log population.

5.7 Relationship to Gibrat's Law

The model I presented in this paper is not dynamic, but we can consider a simple extension in which the v_t's evolve over time. Importantly because the relationship between v_t and log population is concave, there is no need for v_k to be a random walk to generate Gibrat's Law. Rather, it can have a stationary distribution and still generate Gibrat's Law. To see this in continuous time,[18] assume

$$\dot{v} = -Rv + AZ$$

where $\dot{v}$ is the instantaneous rate of change of V, R is a persistence parameter, A governs the variance of v, and Z is a Brownian motion. If $R > 0$, then v_t has an ergodic distribution.

Because the population elasticity in increasing in v, the evolution of $\log p$ is

$$\dot{\log p} = \left(-R\epsilon + \frac{1}{2}\frac{\partial \epsilon}{\partial v}\right) v + \epsilon AZ$$

where ϵ is the local semi-elasticity of p to v_k.

The extra term, which is due to Ito's Lemma, might offset the persistence parameter for the right combination of values, which could mean that log population growth is a random walk. Hence, there is no contradiction between Gibrat's Law and the thin-tailed v_k that I found in the model.

6 Conclusion

In this paper, I show that the city size distribution is in fact a fractal, and has a fat tail at various subgeographies. I propose a toy model featuring heterogeneity in geographic attachment. The toy model matches the fractal pattern exactly. Next, I show empirical evidence in support of the model's main mechanism: in general people are 10 percent more likely to live in a city that is 10 percent bigger, but this elasticity increases in the distance away from the person's birthplace. Outside their home state, a person is 11.5 percent more likely to live in a city that is 10 percent bigger. Building attachment heterogeneity into an otherwise standard location choice model, I show that the model can match this empirical fact, and also generate the fat tail of the city size pattern.

[18]For this exercise, I am considering movement in one city, and ignoring the spillovers from other cities. I do not expect it to change the results.

References

Albert, Christoph and Joan Monras, "Immigration and spatial equilibrium: the role of expenditures in the country of origin," *American Economic Review*, 2022, *112* (11), 3763–3802.

Arshad, Sidra, Shougeng Hu, and Badar Nadeem Ashraf, "Zipf's law and city size distribution: A survey of the literature and future research agenda," *Physica A: Statistical Mechanics and its Applications*, 2018, *492*, 75–92.

Auerbach, Felix, "Das gesetz der bevölkerungskonzentration.," *Petermanns Geographische Mitteilungen*, 1913, *59*, 74–76.

Bee, Marco, Massimo Riccaboni, and Stefano Schiavo, "The size distribution of US cities: Not Pareto, even in the tail," *Economics Letters*, 2013, *120* (2), 232–237.

Behrens, K., G. Duranton, and F. Robert-Nicoud, "Productive cities: Sorting, selection, and agglomeration," *J. Polit. Econ.*, 2014, *122*, 507–553.

Berry, B.J. and A. Okulicz-Kozaryn, "The city size distribution debate: Resolution for US urban regions and megalopolitan areas," *Cities*, 2012, *29*, S17–S23.

Black, Duncan and J Vernon Henderson, "Urban evolution in the USA," *Journal of Economic Geography*, 2003, *3* (4), 343–372.

Cadena, Brian C and Brian K Kovak, "Immigrants equilibrate local labor markets: Evidence from the Great Recession," *American Economic Journal: Applied Economics*, 2016, *8* (1), 257–290.

Center, Missouri Census Data, "Geocorr 2014: Geographic Correspondence Engine," `https://mcdc.missouri.edu/applications/geocorr2014.html` 2016.

Córdoba, J.-C., "On the distribution of city sizes," *J. Urban Econ.*, 2008, *63*, 177–197.

Córdoba, J.C., "A generalized Gibrat's law," *Int. Econ. Rev.*, 2008, *49*, 1463–1468.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American Economic Review*, 2016, *106* (3), 479–524.

_ **and Cecile Gaubert**, "Spatial sorting and inequality," *Annual Review of Economics*, 2022, *14*, 795–819.

_ **and Enrico Moretti**, "Where is standard of living the highest? Local prices and the geography of consumption," Technical Report, National Bureau of Economic Research 2021.

Duranton, G., "Some foundations for Zipf's law: Product proliferation and local spillovers," *Reg. Sci. Urban Econ.*, 2006, *36*, 542–563.

_ , "Urban evolutions: The fast, the slow, and the still," *Am. Econ. Rev.*, 2007, *97*, 197–221.

Eeckhout, J., "Gibrat's law for (all) cities," *Am. Econ. Rev.*, 2004, *94*, 1429–1451.

Fazio, Giorgio and Marco Modica, "Pareto or log-normal? Best fit and truncation in the distribution of all cities," *Journal of Regional Science*, 2015, *55*, 736–756.

Gabaix, Xavier, "Zipf's law for cities: an explanation," *The Quarterly Journal of Economics*, 1999, *114* (3), 739–767.

_ , Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The dynamics of inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

Gibrat, Robert, *Les Inégalités économiques*, Paris, France, 1931.

Giesen, Kristian and Jens Südekum, "Zipf's law for cities in the regions and the country," *Journal of Economic Geography*, 2011, *11* (4), 667–686.

González-Val, Rafael, "The evolution of US city size distribution from a long-term perspective (1900–2000)," *Journal of Regional Science*, 2010, *50*, 952–972.

Hsu, W.T., "Central place theory and city size distribution," *Econ. J.*, 2012, *122*, 903–932.

Ioannides, Yannis M and Spyros Skouras, "US city size distribution: Robustly Pareto, but only in the tail," *Journal of Urban Economics*, 2013, *73*, 18–29.

Jones, Charles I and Jihee Kim, "A Schumpeterian model of top income inequality," *Journal of Political Economy*, 2018, *126* (5), 1785–1826.

Kumar, Gopalan and Arjunan Subbarayan, "The temporal dynamics of regional city size distribution: Andhra Pradesh (1951–2001)," *Journal of Mathematics and Statistics*, 2014, *10* (2), 221–230.

Lee, S. and Q. Li, "Uneven landscapes and city size distributions," *J. Urban Econ.*, 2013, *78*, 19–29.

Levy, Moshe, "Gibrat's law for (all) cities: Comment," *American Economic Review*, 2009, *99* (5), 1672–1675.

Luttmer, Erzo Gerrit Jan et al., *An Assignment Model of Knowledge Diffusion and Income Inequality*, Federal Reserve Bank, Research Department, 2015.

Malevergne, Y., V. Pisarenko, and D. Sornette, "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities," *Phys. Rev. E*, 2011, *83*, 036111.

Modica, M., A. Reggiani, and P. Nijkamp, "Methodological advances in Gibrat's and Zipf's laws: A comparative empirical study on the evolution of urban systems," in H. Shibusawa, K. Sakurai, T. Mizunoya, and S. Uchida, eds., *Socioeconomic Environmental Policies and Evaluations in Regional Science: Essays in Honor of Yoshiro Higano*, Singapore: Springer, 2017, pp. 37–59.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak, "Internal migration in the United States," *Journal of Economic perspectives*, 2011, *25* (3), 173–196.

Nitsch, V., "Zipf zipped," *J. Urban Econ.*, 2005, *57*, 86–100.

Oxford Languages, "Define fractal," `https://www.google.com/search?q=define+fractal` 2023. Accessed June 16, 2023.

Rossi-Hansberg, E. and M.L. Wright, "Urban structure and growth," *Rev. Econ. Stud.*, 2007, *74*, 597–624.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler, "IPUMS USA: Version 13.0 [dataset]," `https://doi.org/10.18128/D010.V13.0` 2023.

Subbarayan, Arjunan, "The size distribution of cities in Tamilnadu (1901–2001)," *International Journal of Agricultural & Statistical Sciences*, 2009, *5* (2), 373–382.

Ye, Xinyue and Yichun Xie, "Re-examination of Zipf's law and urban dynamic in China: a regional approach," *The Annals of Regional Science*, 2012, *49*, 135–156.

Zipf, George Kingsley, *Human Behavior and the Principle of Least Effort*, New York, Hafner Publishing Company, 1949.

Ziqin, Wang, "Zipf law analysis of urban scale in China," *Asian Journal of Social Sciences and Studies*, 2016, *1*, 53.

A Appendix Figures

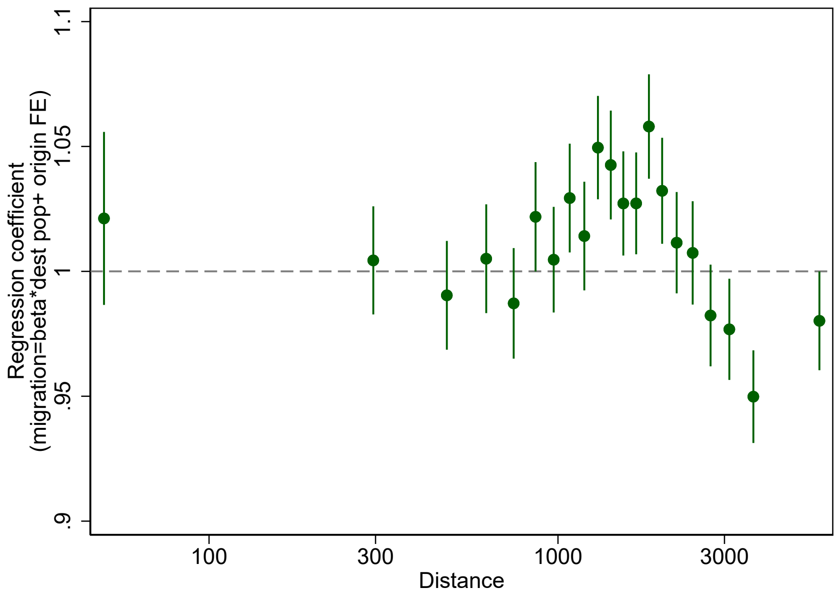


Figure A1: The importance of city size, by distance, in the constrained model

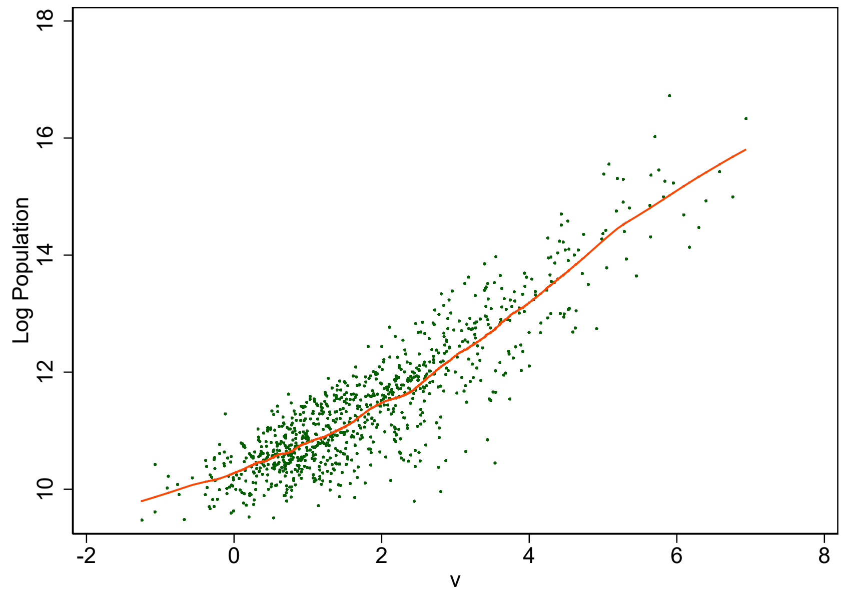


Figure A2: The relationship between log population and v_k, in the constrained model.

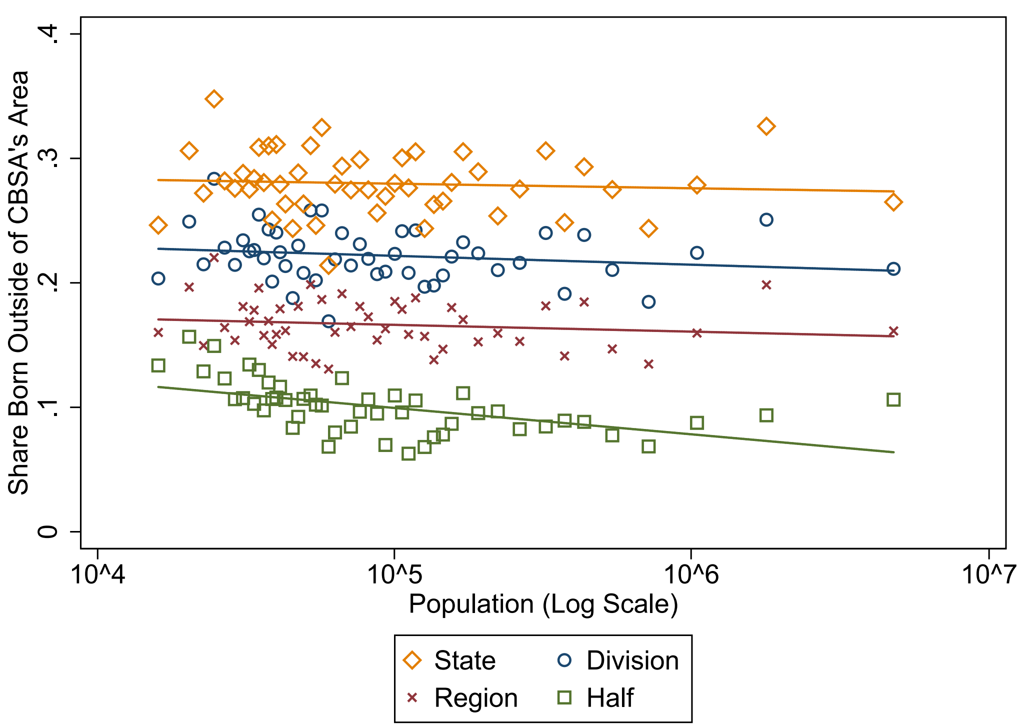


Figure A3: Share of the city that comes from out of the area