ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

CITIES AS SPATIAL CLUSTERS

Ferdinand Rauch

Number 656
May 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

Cities as Spatial Clusters

Ferdinand Rauch*

JEL classification: R12

Keywords: Zipf's Law for cities, Distribution of city sizes

May 27, 2013

Abstract

This paper shows that Zipf's Law for cities can emerge as a property of a clustering process. If initially uniformly distributed people chose their location based on a specific gravity equation as found in trade studies, they will form cities that follow Zipf's Law in expected value. This view of cities as spatial agglomerations is supported empirically by the observation that larger cities are surrounded by larger hinterland areas and larger countryside populations.

*Oxford University, Department of Economics, and CEP, ferdinand.rauch@economics.ox.ac.uk.
Special thanks to Xavier Gabaix, Kristian Giesen, Guy Michaels, Henry Overman, Diego Puga, Stephen Redding, Tony Venables, participants at the Trade Workshop at the LSE and the II. Workshop in Urban Economics in Barcelona as well as two anonymous referees for helpful comments.

1 Introduction

Zipf's Law for cities, which is sometimes referred to as rank-size rule, is a stylized distribution of the size of cities, which suggests that the size of a city is inversely proportional to its rank. The exact distribution specifies that the largest city has twice the population of the second largest, and three times the population of the third largest and so forth.[1] Expressed in more technical terms, Zipf's Law is a Pareto distribution with a shape parameter of minus one. If S_r denotes the size of the r-th largest city, this distribution implies the relationship $rS_r = S_1$. In logarithmic form, this relationship becomes

$$\ln r = \ln S_1 + \beta \ln S_r. \tag{1}$$

Empirical tests of Zipf's Law for cities typically test for a coefficient of $\beta = -1$ in an estimation of equation 1, a coefficient that many studies using different time periods and countries could not reject. Recent evidence that follows a bottom-up economic definition of city boundaries was provided by Makse et al (2011), but there are many more examples.[2] Some studies present evidence against a coefficient of $\beta = -1$. In a meta-study, Nitsch (2005) found a mean coefficient of -1.1 for a selection of recent estimates and could reject a coefficient of -1 in the aggregate of these studies. The alternative of -1.1 is however close to the stylized Zipf coefficient.

The wide range of countries and years for which this empirical regularity has been

[1] The reference is Zipf (1949), who observed the same pattern for the distribution of the frequency of words in languages. A similar pattern was confirmed for firm size, see Axtell (2001). The earliest formulation of this regularity for cities quoted in the literature is the one by Auerbach (1913).

[2] Eaton and Eckstein (1997) provide evidence for Zipf's Law in Japan and France in panels that cover several decades. Giesen and Suedekum (2010) show that Zipf's Law holds for Germany as a whole, and also for German regions in a panel from 1975 to 1997. Black and Henderson (2003) and Giesen and Suedekum (2012) provide evidence for Zipf's law in the US. Holmes and Lee find Zipf's Law using a grid of six × six miles in the US. De Vries (1984) provides evidence for historic European data. Further surveys were provided by Gabaix and Ioannides (2004), and Ioannides and Skouras (2009).

observed suggests that a simple and universal force that is not dependent on economic, cultural or geographic specifics might explain a large part of the pattern of urbanization. Some authors have thus thought that a credible explanation should rely only on few parameters and assumptions that still lead to a pattern similar to Zipf. An early work along these lines is Simon (1955), the most prominent recent attempts into this direction are Eeckhout (2004) and Gabaix (1999).[3]

Eeckhout (2004) could not reject for US 2000 data that the distribution of population density represents random draws from a log-normal distribution, when not only the top but the entire distribution is considered. This is consistent with random growth rates of cities, as shown by Gibrat (1931). It does not however translate easily into an explanation for the empirical regularity. To show that a Pareto distribution with a shape paremeter of -1 can not be statistically distinguished from the right tail of a log normal distribution by some standard tests does not imply that random draws from a log normal distribution will follow Zipf's Law, which would depend on parameters of that log normal distribution such as its standard deviation, and also the number of draws.

Gabaix (1999) suggested that random growth leads to Zipf distributions on the right tail if city sizes in this process are not allowed to shrink below a certain lower bound. This 'grain of sand' that is added to random growth is crucial for this argument, since random growth alone will lead to log-normal distributions, as shown by Gibrat (1931). Behrens, Duranton and Robert-Nicoud (2012) provide a static model that generates approximately Zipf's Law. In this model, city population is a power function of the talent of its residents such that small productivity differences between cities can generate large differences in population.

Several models on size or growth rates of cities implicitly see cities as units that have no spatial relationships to one another or their surrounding areas, examples are the

[3] Gabaix (1999) also discusses some earlier similar attempts.

aforementioned papers by Behrens et al. (2012), Eeckhout (2004) and Gabaix (1999). This paper takes the viewpoint of cities as spatial aggregations of countryside populations and shows that a fairly general aggregation process on a two dimensional surface, that uses gravity equations known from trade studies, leads to a Zipf distribution of city size. This result requires only very few structural assumptions, mainly a specific form of relocation preferences.[4]

In the model discussed below, people are initially uniformly distributed on a surface. In addition, there are uniformly distributed city locations with different values that indicate how suitable they are for settlement. The distribution of these values is irrelevant, only their ranking matters as long as the distribution is not degenerate. Each person has an individual maximum willingness to move, and choses the best location within this search radius, consistent with existing models of search. The strictest assumption that this model requires is a specific functional form for the distribution from which the search radii are taken. This assumption however can be obtained from the data, and follows the gravity equation frequently found in the literature on international trade and migration. In addition to Zipf's Law for cities, the model predicts (a) that a larger city is surrounded by larger area and (b) larger countryside population, and (c) that cities consist mainly of people who were born in or near it. These predictions, for which I provide empirical evidence, are typically not part of models of urbanization.

A related work is Hsu (2012), who shows that endogenous firm locations on a circle may lead to cities that are Zipf distributed, if these firms have different production functions. This model differs in at least two important ways from the one discussed here. First, Hsu's model considers location on a one dimensional circle, while the model below considers a two-dimensional surface. Second, the model below does not

[4]This model does not include dynamics, and abstracts from related stylized facts such as Gibrat's Law. If it were extended to a model where people behave in the same way in several time periods, drawing search radii independent of initial location in each period, populations would increasingly move to the best location on the sphere, and vacate all other places. However, this way to incorporate dynamics would violate Gibrat's Law and after some periods also Zipf's Law.

include assumptions on firms and production processes.

The model demonstrates primarily a mathematical property and does not try to build a sophisticated economic argument. It is thus quite simple, which I view as a strength, rather than a weakness. While it would be straight forward to nest this model in an economic model that provides an economic logic for the assumed behavior,[5] I prefer to keep the model in its pure mathematical form, which leaves the mechanics of the argument clear and easy to follow.

2 Model

Consider a borderless surface – for example a sphere – with an area normalized to one. Initially a mass of people is uniformly distributed on the area. There are N potential locations for cities that are spread randomly and independently of initial rural population in that area. This implies that in expected value both initial populations and potential city locations are spread evenly. Some of these locations are better suited for the construction of a city than others, and thus each potential city location j has a score v_j, which indicates how suitable it is as city location.

To describe the clustering behavior of people I assume a simple decision strategy for each individual. Each person has an individual search radius x, and searches for the city with the highest value v_j within distance x of her initial location. This is a fixed search rule, which can be an equilibrium outcome of a search model, see Morgan and Manning (1985). As shown in the following paragraphs this search rule also leads to aggregate behavior that is consistent with the data. The differences in the maximum distance people are willing to move may be accounted for by differences of occupations. Someone who looks for employment as a waiter or taxi driver may

[5]Models with more complex decision processes of consumers could deliver the same result, as long as the aggregate relocation behavior is as described, ie they deliver a fixed radius search rule with one of the distributions analyzed here. Many search models deliver fixed search rule outcomes.

find plenty of offers within her immediate surrounding, while someone who wants to work as a lawyer may have to search using a larger radius within the area to which the law that she studied applies, while a specialized academic researcher may search worldwide for appropriate employment. The utility from moving to a city with value v_j at a distance d_j of the initial position of a person is then

$$U(v_j) = \begin{cases} v_j & \text{if } d_j \leq x \\ 0 & \text{if } d_j > x. \end{cases} \tag{2}$$

Let x^* denote the expected distance that a person moves, and $\bar{x}$ the maximum willingness to travel for this person. People that find no city location within their search radius will not move and stay in their initial rural location. Under the assumptions above every point within distance $\bar{x}$ of this individual's initial location is equally likely to be the preferred location for her. The probability that the highest value v is exactly at a distance of x is $2\pi x/(\bar{x}^2\pi)$. Hence the expected distance in equilibrium is

$$E(x^*|\bar{x}) = \int_0^{\bar{x}} \frac{2\pi x^2}{\bar{x}^2\pi} dx = \frac{2\bar{x}}{3}. \tag{3}$$

This implies that the distribution of the individual maximum willingness to move is proportional to observed moved distances. In this paper I make the central assumption that the search radius of every person is a random draw from the distribution $f(x) = 1/(xk)$, where k is a scaling parameter. This is the distance relationship as frequently found in gravity equations from the trade literature. Gravity type estimation approaches have also frequently been used in the migration literature, see Anderson (2010) for a recent survey. The feature of gravity equations that I use here is the surprisingly robust finding that in regressions of log trade volumes on log distance the coefficient on distance has a coefficient of -1.[6] There is no consen-

[6] See Anderson and van Wincoop (2003) for a prominent use of gravity equations, and Disdier

sus as to why this structural parameter is that constant, however Chaney (2011) provides a theory based on the spread information networks in a one dimensional space.[7] In the next section I will add to that evidence, by showing for 2000 US data that this assumption is consistent with observed migration patterns in the US and internationally, as it should be from equation 3.

On the surface of a sphere the maximum distance one can move is bounded from above. To address this concern I introduce the assumption of a maximum search radius within the population $\bar{x}$ to the model, chosen such that search circles don't overlap with themselves on the surface. On the surface of a sphere of area one, the corresponding maximum value for $\bar{x}$ that fulfills this feature is $\sqrt{\pi}/2$. Note that this is not an assumption that restricts the gravity relationship in the model, it is rather a restriction that ensures that the gravity equation is applied correctly, and that distances that are larger than half the equator do not enter the consideration of people. In addition I show below this restriction is not necessary for the derived results of the model.

The expected value of the number of cities within radius x is equal to $x^2\pi N$, where $x < \bar{x}$. Let $G(v)$ be the cumulative distribution of $g(v)$, the distribution of values of potential city locations v. Note that given the preferences above the shape of the distribution of v is irrelevant as long as it is not degenerate; only the ranking of city values matters. Then if a possible city location draws value v, the probability that it has the highest value within a circle of x is equal to the probability that all $x^2\pi N$ cities in that circle have a smaller value, $G(v)^{x^2\pi N}$. By this problem setup, the expected value of a city size can be written as a function of its value v alone.

and Head (2008) for a survey of coefficients from estimates of the gravity equation and Chaney (2008) for disaggregated estimates.

[7]The mechanisms of his model of contacts and information are fairly different from the one proposed here. One relationship between this paper and his is that he proves gravity for exporters taking Zipf's Law of export sizes as given, while I prove Zipf's Law taking gravity as given. Thus in an interesting way both papers link these fundamental empirical relationships, even if by very different processes.

The solution for the expected value of city sizes in this model requires two elements. First, the expected number of people with a given search radius that move to a location with value v. A person with search radius $\bar{x}_i$ will move to location j with value v_j with probability $G(v)^{x^2\pi N}$ if j lies within distance $\bar{x}$ of i. For city j, there are $\pi x^2 f(x)$ persons with search radius x that are close enough to consider location j. Second, since the outcome of interest is a relationship concerning ranks and not value draws, I need a mapping of the cumulative probability density function of location value draws to ranks. Note that as long as $g(v)$ is continuous, for a draw from $g(v)$ each value between 0 and 1 of $G(v)$ has the same probability. Thus for a mapping from draws from $G(v)$ to ranks the same rules apply as for uniform distributions bounded by 0 and 1. Let R denote the rank of a city, such that the highest valued city has a rank of $R = 1$, and the second most valuable city a value of $R = 2$ and so on. It holds that the expected value of the cdf value of the largest value draw equals $1 - 1/(N + 1)$, and the expected value of the second highest draw equals $1 - 2/(N + 1)$, and generally the expected value of the draw with rank R is $1 - R/(N + 1)$. I use this mapping from expected values of ranks to ranks to rewrite the size of a city as a function of its rank R:

$$S(R) = \int_0^{\bar{x}} \left(1 - \frac{R}{N+1}\right)^{x^2\pi N} \pi x^2 f(x)dx = \frac{(1 - \frac{R}{N+1})^{\bar{x}^2\pi N} - 1}{2kN \ln\left(1 - \frac{R}{N+1}\right)}. \tag{4}$$

Note that the scaling parameter k is irrelevant for the relative size of cities, and thus the shape of this city density is solely dependent on the parameters N and $\bar{x}$. If the number of initial possible locations for cities N is large, we can use the following three approximations for cities with small rank R: $\lim_{N\to\infty} \ln(1 - R/(N + 1)) = -R/(N + 1)$, $\lim_{N\to\infty}(1 - R/(N + 1))^N = e^{-R(N+1)/N}$ and $\lim_{N\to\infty}(N + 1)/N = 1$. These approximations simplify the problem to:

$$\lim_{N\to\infty} S(R) = \frac{1 - \frac{1}{e^{\pi R\bar{x}^2}}}{2Rk} \approx \frac{1}{2Rk} \tag{5}$$

for large N at the right tail of the distribution. The assumption of a large N does not lead to meaningful testable implications, because it is not required in the setup of this model that a potential city location develops into a city.

Note that the assumption of a finite number N would lead to results that look fairly similar to the rank-size rule, albeit expressed in a more complex mathematical form. For example, when I consider the number 135 for N, which is the number that Gabaix (1999) uses for his estimation of the Zipf equation, and compute the 135 stylized city sizes, and regress the exact Zipf city sizes on the approximate biased city size I find a slope coefficient of 0.992 with a t-statistic of over 2670.

Equation 5 corresponds almost to Zipf's Law (equation 1) with the exception that the problem leaves a bias of $e^{\pi^2 \bar{x} R}$, which decreases the size of large cities below their predicted Zipf values. This bias disappears for large values of $\bar{x}$, the maximum observed search distance. Further, regardless of $\bar{x}$ this bias disappears quickly with increasing ranks. The maximum value of $\bar{x}$ with the property that search circles do not overlap themselves on a sphere is $\sqrt{\pi}/2$. For this parameter the term $e^{\pi R \bar{x}^2}$ is approximately equal to 1/12 for the largest unit, 1/140 for the second largest, 1/1640 for the third, and $1/19,300$ for the fourth. Note that this bias comes from the restriction of maximum search distances. The model could we written as one in which the surface stretches infinitely in every direction (as in Chaney 2011). Then the bias converges to zero. At least in the US the equation with the bias may be closer to the data than the one without the bias, as the largest cities are typically below their stylized population levels.

The mapping from the size of a city to its rank $E[S(R)] = \frac{1}{2Rk}$ equals a Pareto distribution with shape parameter minus one, and hence is consistent with Zipf's Law.

3 Supportive empirical evidence

The functional form of $f(x)$

Equation 3 suggests that the search radiuses are directly proportional to observed migration distances. In this section I demonstrate that the particular functional form $f(x) = 1/(xk)$ is indeed consistent with observed migration patterns in the US and internationally.

I obtain evidence for moved distances from US census data using state level data from the 5 percent sample data from the 2000 US census. See the appendix for details on the sources and construction of the data. To remove particular geographic features such as oceans from the data as much as possible, I exclude the states Alaska and Hawaii and people born outside of the US for this exercise. I create bins of size 100 kilometers, and assign each person into the bin in which the air-line distance between the state of her birth and the state of her current residence falls. The label of each bin is the mean of its bound, for example all people who live between 100 and 200 kilometers from their birthplace are assigned to the bin indexed as 150. The large majority of people, over 68 percent of observations, fall into the bin indexed 50 with a distance between 0 and 100 kilometers, which suggests that the majority of US citizens live near their place of birth. The largest remaining distance in this dataset is the distance between California and Maine, with roughly 4200 kilometers. Into this bin falls a share of 0.00017 of all people, those who either were born in California and live in Maine, or those who were born in Maine and live in California.

An extremely rapid decrease of population with distance suggests that a log-log regression may be an appropriate approach to describe a tendency in this relationship, as in estimations of the well known gravity equation, the standard approach to address similar estimations in the trade literature. Using the data described in the last paragraph, I estimate a standard gravity equation, represented by the regression of the log number of people that moved from state j to state i $\ln(f_{ij})$ on the log of the distance between states i and j $\ln(d_{ij})$ and the log populations of both states, $\ln(P_i)$ and $\ln(P_j)$. The estimated gravity equation is of the form

$$\ln f_{ij} = \beta_0 + \beta_1 \ln d_{ij} + \beta_2 \ln P_i + \beta_3 \ln P_j + \epsilon_{ij}, \tag{6}$$

and gives estimates as follows (robust standard error in parentheses): $\beta_0 = 8.03(0.04)$, $\beta_1 = -1.045(0.0345)$. In this estimation with 1824 observations a test that the slope coefficient $\beta_1 = -1$ can not be rejected at the 5 percent level of significance (the p-value of this t-test is 0.187). This results also holds when the control variables P_i and P_j are excluded from the estimation. As a further robustness check I do not consider migration within US states, but international migration to the US. In this specification the estimate of β_1 is $-0.89(0.07)$, and again I can not reject at the 5 percent level of significance that this coefficient equals -1 (the p-value of this t-test is 0.102). Figure 1 shows the fit of the prediction for a bin size of 100 kilometers. The equation above seems to provide a reasonable fit to the data. In Figure 1 I show the number of log number of people that move a certain distance against that distance. I do not use the log of distance to leave the displayed distance values easier to interpret.

This exercise is not a direct estimation of the discussed model, as population in the data is not uniformly distributed in the initial period. Thus equation 6 may capture equilibrium dynamics that are different from the forming dynamics described by my model, mainly because the size of the non-movers will be determined by the degree to which the US already was equilibrium in the initial period. To address this concern I reestimate equation 6 excluding people that live in their state of birth, and thus only consider people who decide to move to another state. In an estimation for the remaining population I find a coefficient of log distance of -0.97 with a robust standard error of 0.06, which also does not reject a coefficient of -1 at conventional levels of statistical significance. Note also that Figure 1 omits people living in their

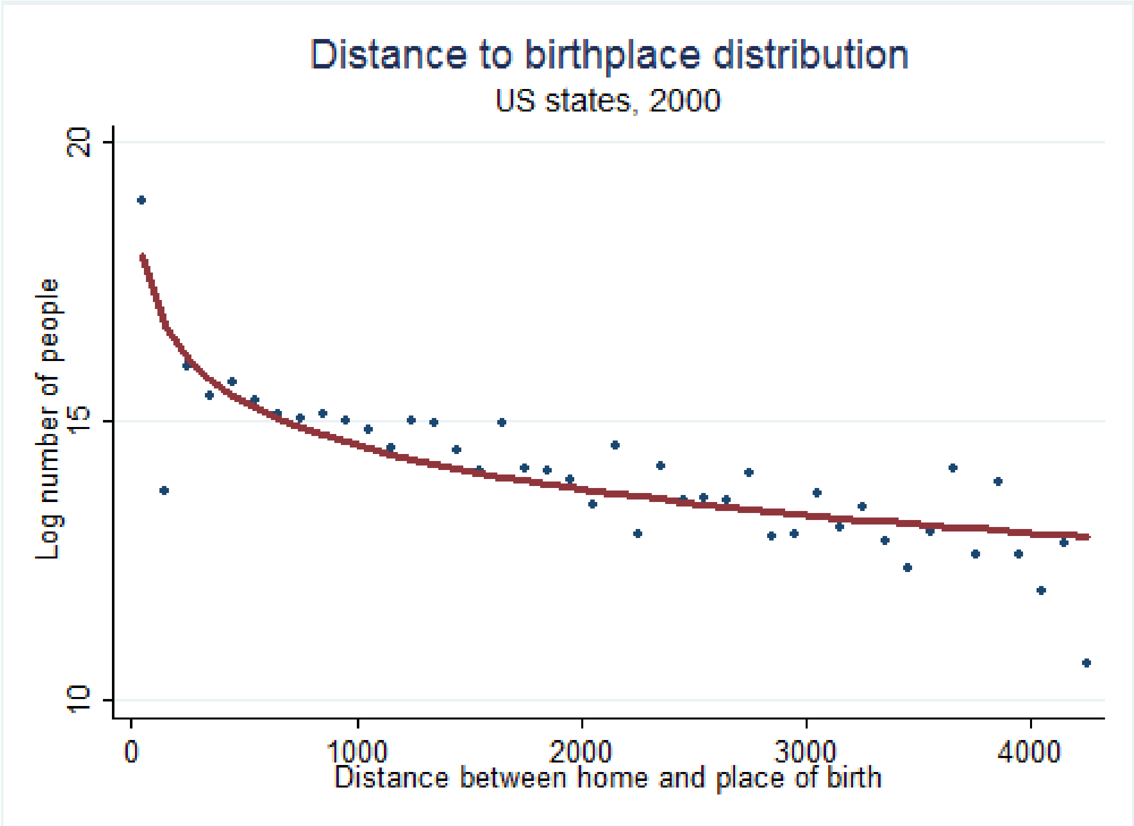

Figure 1: Migration gravity. The figure shows the number of people by the distance between their state of residence and their state of birth. Distances are grouped by bins of certain ranges. The bin size is 100 kilometers. For example, all observations with distances larger than 0 and smaller or equal than 100 kilometers are pooled together in the first bin. People that remained in their state of birth are excluded.

state of birth.

Another concern is that this observation may just be a mechanical implication from a different process given the geographic features of the US. For example, if the number of people moving from state s_i to state s_j is a constant for all i and j, I may still find a declining relationship in the figure, since the number of distances is not uniformly distributed, and people may move to destinations outside of the US. To address this concern I repeat the analysis for people who moved out of Nebraska (a state in the center of the United States), and only include destination states that are within a circle with its center in Nebraska that fully overlaps territory in the United States. These are all the states within a distance closer than 1,000 kilometers to

Nebraska.[8] In this restricted sample I cannot reject a slope coefficient of value -1 at a 95 percent level of significance either.

Another concern is that the assumptions of the model suggests that initially populations are uniformly distributed. The birthplaces in the census data are not uniformly distributed. To correct for this I use a weighting method, dividing the count of people moving from state S_1 to state S_2 by the population of state S_1. In this way each potential birthplace location has the same weight for distance S_2. When I estimate the gravity equation using weighted data the estimated equation is $ln(nr_i) = 9.5(0.03) \; - 0.95(0.04) \; ln(d_i)$. Robust standard errors are displayed in brackets, the number of observations is 2,550. Hence also in this estimation I can not reject a coefficient of -1 on the log distance at 5% level of statistical significance.

I repeat the exercise for migration within the US on the county level in Figure 2 using data on migration between counties from the US Census (2013). Given that there are many more pairwise distances than between states I now use smaller bins of 10 kilometers. Again the gravity for migration holds with great accuracy. The estimation equation corresponding to the trend line in Figure 2 gives now a slope coefficient of -0.9938 with a robust standard error of 0.0573. The p-value of a test that this coefficient is -1 is 0.91. Thus also data for migration between US counties suggest that migration follows gravity patterns.

On the basis of these estimation results, I denote the distribution of maximum distances x in the population by density $f(x)$. Combining the estimated slope coefficient of -1 with equation 3 I chose the functional form $f(x) = 1/(xk)$ for the distribution of search radii, where k is a scaling parameter. Thus the way in which I model individual clustering behavior is in aggregate consistent in shape and magni-

[8] These states are in ascending order of distance to Nebraska: Kansas, South Dakota, Iowa, Minnesota, Colorado, Missouri, Oklahoma, North Dakota, Wyoming, Illinois, Arkansas and New Mexico. Distance is measured as the distance between centroids of states.

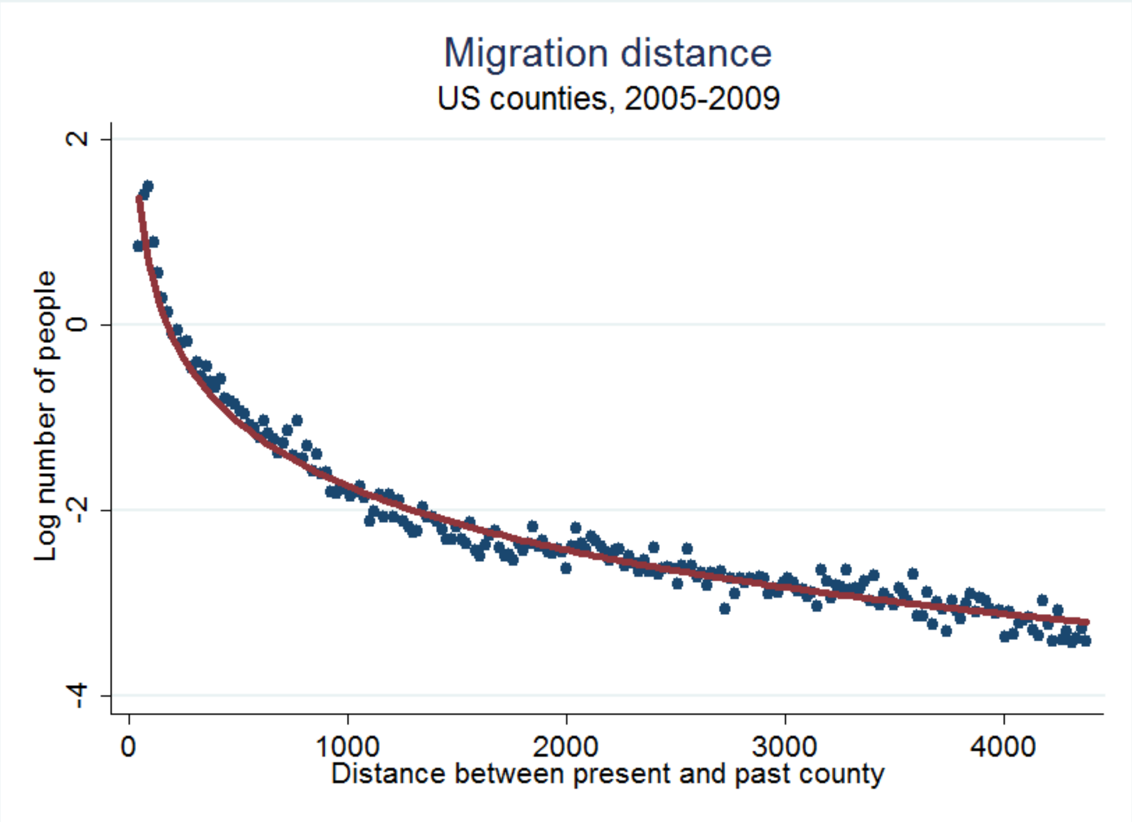

Figure 2: Migration gravity. The figure shows the log number of people by the distance between their county of residence in the years 2009 and 2005. Distances are grouped by bins of certain ranges. The bin size is 10 kilometers. For example, all observations with distances larger than 0 and smaller or equal than 10 kilometers are pooled together in the first bin. People that did not move between counties are excluded.

tude with observations on gravity. Many trade studies estimate similar relationships for the functional form that relates distance and the value of shipments for many types of traded goods or services.

Note further from Figure 1 that people are more likely to live near their place of birth than in places far away. This feature is an equilibrium prediction of the model presented here, but typically not part of other models of urbanization.

Other functional forms for $f(x)$

The functional form $f(x) = 1/(xk)$ is not only consistent with the other, it is also, by coincidence, the only one in the set of more general Pareto functions $1/(kx^p)$ for which the integral solved in equation 4 does not involve solving the incomplete

Gamma function. The general solution for the integral specified in equation 4 under the assumption that N is large and $f(x) = 1/(kx^p)$ is

$$S(R) = -\frac{\Pi^{\frac{p-1}{2}} R^{\frac{p-3}{2}}}{2k} \Gamma\left(\frac{3-p}{2}, \Pi\bar{x}^2 R\right),$$

where Γ is the incomplete Gamma function. For finite values of $\bar{x}$ this is hard to simplify further. This can however be solved for values of $0 < p < 3$ under the assumption that $\bar{x}$ becomes infinite. Note that I can not say how much this simplification biases the stylized relationship in general.[9] Allowing this approximation, the solution further simplifies to involve the complete Gamma function

$$S(R) = \frac{\Pi^{\frac{p-1}{2}} R^{\frac{p-3}{2}}}{2k} \Gamma\left(\frac{3-p}{2}\right).$$

For any value of $0 < p < 3$ this Gamma function gives a positive constant that does not matter to the coefficient of interest. Then the stylized relationship between log rank and log size of cities is

$$\log S(R) = c - \frac{3-p}{2} \log R,$$

where c is a constant that does not depend on R. This is equivalent to a Pareto distribution with a shape parameter of $(3-p)/2$. In the case of $p = 1$, which is the focus of this paper, this indeed gives Zipf's Law with a slope coefficient of -1. In the neighborhood of -1 gravity exponents the relationship remains stable. Similar to the results obtained by Chaney (2011), large deviations from the -1 coefficient in migration patterns would break down Zipf's Law in this model and lead to a different power law. The relationship would remain of Pareto shape.

[9] In the special case of $p = 1$ the bias is the one derived and discussed before in this paper, as both models correspond to one another.

For other functions such as $f(x) = 1/\log(x)$ or $f(x) = 1/(x+k)$ the integral is likely nonelementary.

Correlation between a city and its hinterland

Gabaix (1999) and other previous models that try to explain the Pareto distribution at the right tail view cities as independent observations with no mutual or any other spatial relationships. This model on the other hand has clear testable spatial predictions. Given that a city consists of the people from the area around it, and that in the shadow of a city with a large value draw smaller cities find it difficult to develop, a larger city should be surrounded by larger area and also larger hinterland populations. I show that both correlations are observed in US Minor Civil Divisions data. For this exercise I define the largest N units as cities, and the other units as countryside. I then assign each countryside unit to its nearest city, and compute the conditional correlations.

To test these hypotheses I use Minor Civil Division (MCD) data from the US census. MCDs are county subdivisions that fulfill two essential requirements that make them suitable for this exercise: They are small units (one county incorporates on average 11 MCDs) and they cover the entire area and population of the US. The small size of MCDs allows to distinguish urban and rural units more clearly than at the level of counties or MSAs. After dropping Alaska and Hawaii there are about 35,000 units left, which cover the entire area and population of the remaining states.

Among these MCDs I chose N cities, based on the units with the highest population density. I chose density and not population since the area of MCDs is considerably larger in the West than in the North East, and thus urban areas are more straight forwardly identified with the use of density than population. I consider each of the remaining MCDs a countryside unit, and link it to its nearest city. Then for each city I sum the population and area of these non-city MCDs. I regress the log area and the log population of the surrounding areas on the population of the city.

Table 1 reports these correlations, using robust standard errors in parentheses. In the first two columns I define the densest 1,000 MCDs as cities, and the remaining ones as countryside. I link each countryside unit with its nearest city. In column (1) I sum the area of the hinterland for each city and include the own area of a city. The model assumes that a city is essentially a point on the surface, and thus the MCDs own area is part of the surrounding area. I find evidence of strongly positive correlation between the population of a city and its surrounding areas. In column (2) I regress a cities population on the population of its hinterland. Here I exclude the population of the city itself. Not all cities have surrounding MCDs to which they are closest, thus the number of observations drops. Also this estimation suggests a strong positive correlation.

Columns (3) and (4) repeat the exercise, but include state fixed effects to account for regional variation of circumstances across the US. The columns suggest that the described positive correlation is observed within states as well as across states. In columns (5) and (6) I repeat the exercise, but chose only the 150 densest units as cities, a number that is close to the number of cities included in famous tests of Zipf's Law (for example Gabaix 1999 or Krugman 1996). Again I observe a positive correlation for both variables. In columns (7) and (8) I exclude all MCDs that are within 25 kilometers from a city, which also leads to the exclusion of the own area of a city. This robustness addresses the concern that the observed positive correlations may be driven by suburban areas or populations, which we expect to be larger for larger cities by the monocentric city model. Again the positive correlations are observed.[10] I verify that I get similar results when I aggregate MCDs to the level of Metropolitan Statistical Areas.

[10]The association of a city with its hinterland is further supported by Nitsch (2003), who showed that Vienna shrank when it lost part of its hinterland.

							Drop $\leq$ 25 km	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log area	Log rural population	Log area	Log rural population	Log area	Log rural population	Log area	Log rural population
Log city population	0.443***	0.375***	0.118**	0.125***	0.380**	0.155*	0.193***	0.261***
	(0.0545)	(0.0358)	(0.0554)	(0.0383)	(0.181)	(0.0911)	(0.0642)	(0.0596)
Constant	7.834***	7.525***			9.949***	11.33***	12.58***	8.424***
	(0.561)	(0.350)			(2.093)	(1.027)	(0.680)	(0.616)
Observations	1,000	740	1,000	740	150	132	386	386
Number of cities N	1,000	1,000	1,000	1,000	150	150	1,000	1,000
R^2	0.071	0.158	0.420	0.400	0.048	0.025	0.03	0.061
State fixed effects	No	No	Yes	Yes	No	No	No	No

Table 1: City size explaining its surrounding area (including its own), and surrounding countryside population (excluding own population). Cities are defined as the largest N units according to population density in the MCD data. Robust standard errors displayed in parentheses. Stars denote statistical significance at the 1 percent (***), 5 percent (**) and 10 percent (*) level.

Smaller coefficients for β

As mentioned, Nitsch (2005) expressed some doubt in an empirical meta study of 29 papers on the fit of Zipf's Law, and suggested a slightly smaller pooled estimate of $\hat{\beta} = -1.1$,[11] which would imply that city sizes are closer to each other than suggested by Zipf. Even some studies that confirm Zipf's Law nevertheless notice the absence of very large cities relative to Zipf's Law (for example Rossi-Hansberg and Wright 2007). The bias in equation 5 has the same feature by subtracting more from larger cities. While the magnitude of this bias as derived in equation 5 is small for the largest unit, and disappears very fast for the following ones, it would become more visible for smaller values of $\bar{x}$ and could eventually lead to slope coefficients β of higher absolute value at the right tail of the distribution. Note also from Figure 1 in Gabaix (1999) that the largest 5 cities are below their stylized values, and among them are the most deviating outliers.

Simulation

As additional evidence I simluate the model numerically. I create a sphere with a surface of area 1, and chose random points for the location of people and cities on its surface such that both are uniformly distributed, independently from one another. I then assume a distribution of search radiuses, allocate search radiuses to people randomly, and move every person to the best location within their search radius.

Two features of these simulations are worth pointing out. First, if search radiuses are taken from the distribution $f(x) = 1/(kx)$, the choice of k matters in finite data, contrary to the expected value case solved with an infinite number of people and potential city locations in the theory part. In final datasets a choice of a small k might mean that no person finds a city within the search radius they are given, and a large value of k might mean that every person moves to the same city. For a choice

[11]This smaller coefficient is also observed in some cross-country studies, see Rosen and Ressnick (1980) or Soo (2005).

of k near to these extreme cases of large and small simulation might lead to some populated cities, but the distribution is then often not Pareto. Second, contrary to the expected value case, simulations allow only integer outcomes for the size of a city. Thus for a large choice of parameters there is a long left tail of expected city sizes smaller one that will be transformed into cities of either size 0 or 1. Thus I would expect to find a Pareto distribution of the generated cities only in the right tail of the distribution.

Figure 3 shows the log rank - log size relationship for such a simulation with the following parameters: $100,000$ people, $1,000$ potential city locations, $k = 1/10$, display the 100 largest units. I expect to find cities distributed according to a Pareto distribution with a shape parameter of -1, and add the stylized distribution in that figure. While there is some noise visible, the data follow the stylized distribution quite closely. A test of the relationship in this sample gives a coefficient for the Pareto shape parameter of -1.015 with a standard error of 0.019. Thus I cannot reject that the shape parameter is -1 on the five percent level of significance in this simulation. I repeat this simulation 100 times, and reject at five percent level only 6 times.

4 Conclusion

The contribution of this paper is to propose a theory of city creation that can explain Zipf's Law in a novel, non-dynamic way. In addition, the model accounts for the fact that larger cities are surrounded by larger areas, and also by larger countryside populations, and that cities consist by majority of people that were born near or in it. Existing explanations of Zipf's Law typically do not predict these latter features. To arrive at these properties, the paper studies clustering behavior of uniformly distributed populations on a surface that move to locations of different values. In

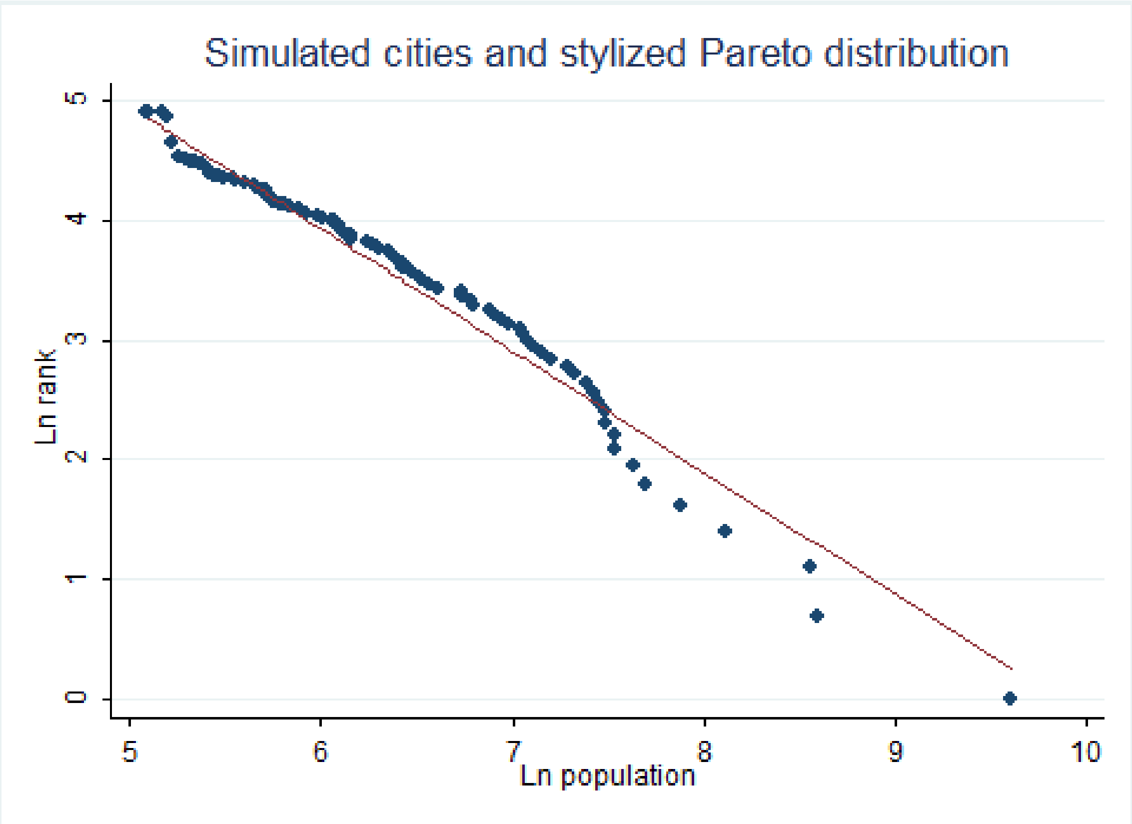

Figure 3: The scatter shows the relationship between log size and log rank for simulated data for the 100 largest cities. The line shows a stylized Pareto distribution with shape parameter -1. Cities were generated following the algorithm of the model, using a sphere of area 1, 100,000 people and 1,000 potential city locations. The value of k in this simulation is 1/10.

this model, each person searches possible locations within a heterogeneous radius, and then moves to the best location she or he can find within this circle. This behavior is consistent with data on aggregate population movement in the United States.

The model tries to highlight a mathematical property that may explain the distribution of cities in the simplest possible form. Therefore it is very simple, for example in the way in which it models preferences and thus behavior. It also abstracts from geographic, cultural or economic differences, and it considers only one period. If these or other elements were added into a more complex model, the general results of this paper may persist, as long as the aggregate clustering behavior remains in place. However, I see the simplicity of the argument as a strength rather than a weakness, as only a simple explanation with few parameters can account for the seeming universal observation of Zipf's Law.

References

[1] Anderson, James E., 'The Gravity Model', NBER Working Paper, 2010.

[2] Anderson, James E. and Eric van Wincoop, 'Gravity with Gravitas: A Solution to the Border Puzzle', *American Economic Review*, Vol 93 (March 1), 2003.

[3] Auerbach, Felix, 'Das Gesetz der Bevölkerungskonzentration', *Petermann's Geographische Mitteilungen 59*, 1913.

[4] Axtell, Robert L., 'Zipf Distribution of U.S. Firm Sizes', *Science* 7, Vol. 293, no. 5536, 2001.

[5] Behrens, Kristian, Gilles Duranton and Frederic Robert Nicoud, 'Productive Cities: Sorting, Selection and Agglomeration', Draft, 2012.

[6] Black, Duncan and Vernon Henderson, 'Urban evolution in the USA', *Journal of Economic Geography*, Vol 3 (4) 2003.

[7] Chaney, Thomas, 'Distorted Gravity: the Intensive and Extensive Margins of International Trade', *American Economic Review*, Vol. 98, September 2008.

[8] Chaney, Thomas, 'The Gravity Equation in International Trade: An Explanation', Working Paper, September 2011.

[9] Disdier, Anne-Celia and Keith Head, 'The Puzzling Persistence of the Distance Effect on Bilateral Trade', *The Review of Economics and Statistics*, Vol 90(1), 2008.

[10] Eaton, Jonathan and Zvi Eckstein, 'Cities and growth: Theory and evidence from France and Japan', *Regional Science and Urban Economics*, Vol 27 (4-5), pages 443-474, 1997.

[11] Eeckhout, Jan, 'Gibrat's Law for (All) Cities', *American Economic Review*, December, 2004.

[12] Gabaix, Xavier, 'Zipf's Law for Cities: An Explanation', *The Quarterly Journal of Economics*, Vol 94 (5), 1999.

[13] Gabaix, Xavier, and Yannis Ioannides, 'The Evolution of City Size Distributions', Handbook of Regional and Urban Economics, volume 4, Chapter 53, V. Henderson and J-F. Thisse eds, North-Holland, p.2341-2378, Vol 4, pages 2341-2378, 2004.

[14] Gibrat, R., 'Les in*é*galit*é*s *é*conomiques', Paris, France: Librairie du Recueil Sirey, 1931.

[15] Giesen, Kristian and Jens Suedekum, 'Zipf's law for cities in the regions and the country', *Journal of Economic Geography*, Vol 11(4), 2010.

[16] Giesen, Kristian and Jens Suedekum, 'The Size Distribution across all Cities: A Unifying Approach', *CESifo Working Paper No 3730*, February 2012.

[17] Holmes, Tom and Sanghoon Lee, 'Cities as Six-by-Six Mile Squares: Zipf's Law?', in Agglomeration Economics, edited by Ed Glaeser, University of Chicago Press, 2010.

[18] Hsu, W.-T., 'Central Place Theory and City Size Distribution', *The Economic Journal*, 122, September 2012.

[19] Ioannides, Yannis and Spyros Skouras, 'Gibrat's Law for (All) Cities: A Rejoinder', Working Paper, 2009.

[20] Krugman, Paul, 'The Self-Organizing Economy', Cambridge, MA: Blackwell, 1996.

[21] Makse, Hernan, Xavier Gabaix, Hernan Rozenfeld and Diego Rybski (2011), 'The Area and Population of Cities: New Insights from a Different Perspective on Cities', *American Economic Review*, 2011.

[22] Michaels, Guy, Ferdinand Rauch and Stephen Redding, 'Urbanization and Structural Transformation', *Quarterly Journal of Economics*, 127(2), 535-586, May 2012.

[23] Morgan, Peter and Richard Manning, 'Optimal Search', *Econometrica*, Vol. 53, No. 4, 1985.

[24] Nitsch, Volker, 'Does history matter for urban primacy? The case of Vienna', *Regional Science and Urban Economics* 33, 401-418, 2003.

[25] Nitsch, Volker, 'Zipf zipped', *Journal of Urban Economics*, 2005.

[26] Rosen, K. and M. Resnick, 'The size distribution of cities: an examination of the Pareto law and primacy', *Journal of Urban Economics* 8, 165-186, 2005.

[27] Rossi-Hansberg, Esteban and Mark Wright, 'Urban Structure and Growth', *Review of Economic Studies* 74, 2007.

[28] Simon, Herbert, 'On a Class of Skew Distribution Functions', *Biometrika*, 1955.

[29] Soo, Kwok 'Zipf's Law for Cities: A Cross-Country Investigation', *Regional Science and Urban Economics* 35(3): 239-263, 2005.

[30] US Census, 'County-to-County Migration Flows, 2005-2009 5 year ACS', downloaded April 2013 from http://www.census.gov/hhes/migration/data/acs/county-to-county.html.

[31] United Nations, 'World Urbanization Prospects: The 2007 Revision', United Nations Publishing, ISBN 978-92-1-151447-6, 2007.

[32] de Vries, Jan, 'European Urbanization, 1500-1800', Methuen and Co. Ltd. London, 1984.

[33] Zipf, George K., 'Human Behavior and the Principle of Least Effort' Addison-Wesley, Cambridge, MA, 1949.

5 Appendix 1: Data

State level data for 2000

The state dataset is created by merging two datasets: First the US census data as provided by IPUMS (Integrated Public Use Microdata Series, www.ipums.org). I keep information on the place of birth and the state for each person. Birthplace information is not provided on a more disaggregated level than state level by this source. I add information on mutual distances between states from GIS, using digital maps provided by NHGIS (The National Historical Geographic Information System, www.nhgis.org), where distance is measured as the great circle distance between the centroids of the two states. I exclude Alaska and Hawaii from this dataset because of their peculiar geographic relationships to the rest of the US. I keep the person weights created by the census and provided by IPUMS, and compute all provided statistics using these weights to reflect true populations.

MCD 2000 data

This dataset uses Minor Civil Divisions (MCDs) for the year 2000, as provided by the US census factfinder, factfinder.census.gov. MCDs are county subdivisions from the US census, that are small units that still cover the entire surface and population of the US. I merge this data with data from a map from NHGIS (the National Historical Geographic Information System, www.nhgis.org) to add information on latitude, longitude and area. I add MSA (Metropolitan Statistical Area) boundaries from TIGER (Topologically Integrated Geographic Encoding and Referencing) from the US census. I compute these three variables in GIS from the provided map. I exclude Alaska and Hawaii because of their peculiar geographic relationships to the rest of the US. Distance in the dataset is then measured as great circle distance. See the US census documentation or Michaels, Rauch and Redding (2011) for a more detailed description and summary statistics of MCDs.