Uneven Landscapes and City Size Distributions

Sanghoon Lee   Qiang Li*

April 24, 2013

Abstract

This paper proposes a new model generating city size distributions that asymptotically follow the log-normal distribution. The log-normal distribution is consistent with Zipf's law *in the top tail*, which is known to hold for many countries in different periods. The key feature of our model is that it can express city size as a product of multiple random factors (e.g., climate, geographic features, and industry composition). Each factor alone need not generate Zipf's law. Our model provides a justification for classical urban economics models that have been criticized for not delivering Zipf's law, since a single model typically represents only one factor among many present in reality.

JEL Classification Codes: D39, R12

Keywords: City Size Distribution, Zipf's Law, Rank-Size Rule, Log-normal Distribution

*Sanghoon Lee, Sauder School of Business, The University of British Columbia, 2053 Main Mall, Vancouver, BC V6T 1Z2, Canada, Email: sanghoon.lee@sauder.ubc.ca; Qiang Li, Department of Real Estate, National University of Singapore, Singapore 117566. Email: rstlq@nus.edu.sg. We are very grateful to the editor Gilles Duranton, two anonymous referees, Tom Holmes, Keith Head, Vernon Henderson, Xavier Gabaix, Ralph Winter, Nate Schiff, Kristian Behrens, Robert Masson and David Cuberes for valuable comments. We also thank seminar participants at NBER summer institute 2012, University of British Columbia, Yonsei University, the Barcelona Institute of Economics (IEB) Workshop on Urban Economics 2010, Shanghai University of Finance and Economics, the Econometric Society World Congress 2010, Urban Economics Association Annual Meetings 2008 and 2010, and the Federal Reserve Bank of Philadelphia. We are grateful for financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC).

1 Introduction

Many empirical papers have documented that Zipf's law holds *in the top tails* of city size distributions for different countries in different periods (e.g., Rosen and Resnick, 1980; Dobkins and Ioannides, 2001; Ioannides and Overman, 2003; Gabaix and Ioannides, 2004).[1] Zipf's law (or the rank-size rule) indicates that the population size of a city tends to be inversely proportional to its rank: the second largest city in a country is about half the size of the largest city, the third largest city is about one third the size of the largest city, and so forth.

Zipf's law is known to arise when city size follows the Pareto distribution with a shape parameter equal to 1. However, Eeckhout (2004) argues that the log-normal distribution is virtually indistinguishable from the Pareto distribution in the top tail, and can thus be consistent with Zipf's law. In this paper, we propose a new model generating a city size distribution that converges to the log-normal distribution. With certain parameter values, our model can match Zipf's law in the top tail.

The key idea of our model is that many random factors jointly determine city size (e.g., climate, geographic features, and industry composition), and that equilibrium city size can be expressed as a product of these random factors. By applying the central limit theorem (after a log-transformation), we show that city size asymptotically follows a log-normal distribution when we have a sufficient number of factors.

Since modern central limit theorems require only weak conditions, our result applies quite generally; the random factors need not follow any specific distribution, can come from different distributions, and may be correlated with each other to some degree. However, the central limit theorems require that city size be determined by a sufficient number of small factors; we examine this issue using Monte-Carlo simulations. The simulation results show that our model can achieve a good approximation of the Pareto distribution in the top tail,

[1] Soo (2005) finds that Zipf's law does not hold for many countries. Gabaix and Ioannides (2004) argue that these deviations can occur due to idiosyncratic differences (e.g., political economy variables) even if the underlying distribution follows Zipf's law.

even with a small number of factors that are substantially correlated.

Zipf's law requires not only that city size distribution follow a Pareto distribution, but also that its shape parameter be 1. We prove analytically that, in our model, the estimated Pareto shape parameter decreases with an agglomeration economy parameter. Thus, we can make the shape parameter be 1 by adjusting the parameter.

Although our model can match Zipf's law in the top tail, it still remains an open question how well our model will fit the real city size distribution, which may not follow Zipf's law perfectly. We fit our model to U.S. city size distributions, using three different data sets employing different definitions of cities: the Core Based Statistical Areas (CBSAs), Census places, and the area clusters defined in Rozenfeld, Rybski, Gabaix, and Makse (2011). We find that our model does very well fitting the CBSAs and does reasonably well fitting the top tails of Census places or the area clusters.

There are other theoretical papers that explain the empirical city size distribution. The main workhorse in this literature is the random growth of cities (e.g., Simon, 1955; Gabaix, 1999; Eeckhout, 2004; Duranton, 2006, 2007; Rossi-Hansberg and Wright, 2007; Córdoba, 2008; Berliant and Watanabe, 2009). When the growth rate does not depend on city size (i.e., when Gibrat's law holds), city size distribution converges to the log-normal distribution, or to the Zipf distribution when there is a lower reflexive bound on city size. There are two recent static models as well. Hsu (2012) uses the central place theory and Behrens, Duranton, and Robert-Nicoud (2010) use human capital distribution.

Mechanically, our model is a static version of random growth models. The random shocks are stacked in the cross section instead of time. However, being a static model yields unique economic interpretations and implications. A new implication is that we provide a justification for classical urban economics models such as Henderson (1974), which are sometimes criticized for not being able to deliver Zipf's law (e.g., Krugman, 1996; Gabaix, 1999). A typical economic model highlights only one economic factor; our model shows that it is possible to match the empirical pattern by combining many factors, even if each factor does not generate the pattern on its own.[2]

All the models above assume that locations are ex-ante identical. Alternatively, Krugman (1996) points out that locations are ex-ante heterogeneous, and suggests that the cross-city heterogeneity in locational fundamentals may generate Zipf's law.[3] Our paper formalizes Krugman (1996)'s insight, but is original in two ways. First, the mechanism is novel in that the multiplicity of random factors generates the log-normal distribution. Second, random factors are not limited to natural features, but can also be man-made factors (e.g., industry composition and tax policies). The randomness in these factors can be due to exogenous shocks in innovation, voting, and policy-makers' decisions.[4]

Another contribution of our paper is that we introduce a modern version of the central limit theorem to the literature. Eeckhout (2004) first used a central limit theorem to study city size distributions. However, he used the *classical* central limit theorem, which requires growth rate shocks in his model to be independent and identically distributed across periods and cities. This requirement can be at odds with recent findings by Glaeser, Ponzetto, and Tobio (2011), Black and Henderson (2003), and Desmet and Rappaport (2013), that Gibrat's law does not hold in a *short* time span. If Eeckhout (2004) had used the modern version of the central limit theorem, his result would have become compatible with these findings; it would require Gibrat's law to hold only in a *long* time span, consistent with findings by Glaeser et al. (2011).

The rest of the paper is structured as follows. Section 2 provides the model. Section 3 shows that our model generates a city size distribution that converges to the log-normal distribution as the number of factors increases. Section 4 shows that our model can generate Zipf's law in the top tail with certain parameter values. It also shows how many factors we

[2] This mechanism does not work if a theory claims to represent a dominant factor determining city size.

[3] Davis and Weinstein (2002) and Rappaport and Sachs (2003) empirically found that locational fundamentals play an important role in determining city sizes. Davis and Weinstein (2002), in particular, test locational fundamental theories against dynamic random growth theories, using the extensive bombings over Japanese cities during the Second World War. They favor locational fundamental theories, based on their finding that, after the war, most Japanese cities returned to their original positions in the size hierarchy.

[4] Duranton (2007) shows how randomness in innovation can lead to different industry compositions across cities.

need and how much correlation is allowed among the factors. Section 5 shows how well our model can fit the U.S. city size distribution. Section 6 concludes.

2 Model

Our model builds on Roback (1982). The Roback model predicts the wage and rent levels of a city as functions of its local production and consumption amenities, but does not predict city size. We make two changes to transform the Roback model into a model of city size distribution. First, we add a housing market, which works as the congestion force pinning down the population size of a city. Second, we allow local production, consumption amenities, and land supply to depend on population size, in order to capture agglomeration economies. The resulting model predicts city size as an increasing function of all of the above features.

Other papers have used a similar modeling strategy of adding friction, like the housing market in our paper, to the Roback framework to pin down city size (e.g., Glaeser, Gyourko, and Saks (2006), Rappaport (2008), Glaeser and Gottlieb (2009), and Desmet and Rossi-Hansberg (2013)). Our modeling contribution is that we provide a simple and analytically tractable model that endogenizes agglomeration economies in consumption amenities, production amenities, and land supply.

2.1 Description

There is a continuum of potential city sites, indexed by $s \in [0,1]$. The locations differ *exogenously* in three groups of characteristics: natural consumption amenities $\mathbf{a} \in \mathbb{R}^J$, natural production amenities $\mathbf{o} \in \mathbb{R}^K$, and land supply factors $\mathbf{l} \in \mathbb{R}^M$. The natural features include rivers, mountains, climate, and coastal locations. They also include exogenous random components in man-made features, such as industry composition, road network, zoning, etc. The randomness in these factors is due to randomness in innovation, policy-making, big firms' location choices, etc.

These vectors of exogenous factors $\mathbf{a}, \mathbf{o}, \mathbf{l}$ are aggregated into three scalars of aggregate amenities: consumption amenities $A \in \mathbb{R}$, production amenities $O \in \mathbb{R}$, and land supply $L \in \mathbb{R}$.

$$\begin{aligned} A &= A(N, \mathbf{a}), \\ O &= O(N, \mathbf{o}), \\ L &= L(N, \mathbf{l}). \end{aligned}$$

where N is population size. We allow aggregate amenities A, O, L to depend on population size N to capture agglomeration economies in each channel. For example, firms in a city may become more productive as the city size increases, or land-zoning regulation may depend on the population size of a city.

There are two commodities: housing and a homogeneous good outside housing. The homogeneous good is freely tradable with zero transportation cost, while housing is locally provided. The markets for both goods are perfectly competitive.

$\bar{N}$ workers live in the economy. All workers are homogeneous and freely mobile with zero moving costs. A worker first chooses a city to live in, and then chooses her consumption bundle consisting of the homogeneous good q and housing h. Their utility function $U(q, h; A)$ is strictly increasing in consumption amenities A. Each worker supplies one unit of labour inelastically. The decision of a worker can be summarized by the following optimization problem:

$$\max_{s} V\left(r_s, w_s; A_s\right)$$

where

$$V\left(r_s, w_s; A_s\right) \equiv \max_{q,h} U(q, h; A_s) \text{ subject to } q + r_s h = w_s.$$

where r_s, w_s, and A_s are housing rent, wage, and consumption amenities in city s. Note in the budget constraint that we use the homogeneous good as the numeraire.

Each city has numerous firms producing the homogeneous good. All firms use the same constant-returns-to-scale technology; thus, we can consider one aggregate firm for each city, which behaves like a perfectly competitive firm. The aggregate firm uses labour n and buildings that we assume come from the same stock of housing as workers' housing h. The production function F is increasing in the production amenities O. The decision of a firm in city s can be summarized by the following maximization problem:

$$\max_{n,h} F(n, h; O_s) - w_s n - r_s h$$

where n and h are labour and housing input.

All housing is owned by absentee landlords. Instead of explicitly modeling housing developers, we assume for simplicity that housing supply is a function increasing in rent r and land supply L:

$$H^S(r; L).$$

2.2 Equilibrium

An equilibrium of the model $\{S, \bar{u}, w_s, r_s, N_s | s \in S\}$ consists of the set of populated sites S, equilibrium utility level $\bar{u}$, and wage w_s, rent r_s, and population size N_s for each city $s \in S$, satisfying the following five conditions. First, workers get the same utility across all populated locations:

$$V(r_s, w_s; A(N_s; \mathbf{a}_s)) = \bar{u} \text{ for } s \in S \tag{1}$$

where $\bar{u}$ is the common utility level and S is the set of populated locations (i.e., $S \equiv \{s | N_s > 0\}$).

Second, firms that produce the homogeneous good earn zero profits. Since the firms use constant-returns-to-scale technology, the zero profit condition is equivalent to the unit cost being equal to the unit output price:

$$C(r_s, w_s; O(N_s; \mathbf{o}_s)) = 1 \text{ for } s \in S \tag{2}$$

where C is the unit cost function.

Third, the housing market in each city clears:

$$H^D(N_s, r_s, w_s; A(N_s; \mathbf{a}_s), O(N_s; \mathbf{o}_s)) = H^S(r_s; L(N_s; \mathbf{l}_s)) \text{ for } s \in S \tag{3}$$

where H^D is the aggregate housing demand function of workers and firms.

Fourth, economy-wide labour market clears:

$$\int_S N_s ds = \bar{N}. \tag{4}$$

Fifth, unpopulated sites offer lower utility than the common utility level $\bar{u}$:

$$V(r_s, w_s; A(0, \mathbf{a}_s)) \leq \bar{u} \text{ for } s \notin S. \tag{5}$$

Equations (1) to (3) determine wage w_s, housing rent r_s, and population size N_s for each city $s \in S$, when the common utility level $\bar{u}$ and the set of populated sites S are given. Equation (4) determines the common utility level $\bar{u}$ given the set of populated sites S, since N_s is a function of $\bar{u}$. Equation (5) characterizes the set of populated sites S.

3 The Log-normal Distribution

This section imposes specific functional forms on the model, and shows that our model generates a city size distribution converging to the log-normal distribution. We use the following functional forms for workers' preferences, firms' production technologies, and housing supplies.

$$
\begin{aligned}
(6) \qquad \text{Preference:} \quad & U(q,h;A) = A \cdot q^{\alpha} h^{1-\alpha} \\
\text{Production technology:} \quad & F(n,h;O) = O \cdot n^{\beta} h^{1-\beta} \\
\text{Housing supply:} \quad & H^{s}(r;L) = L \cdot r^{\gamma}
\end{aligned}
$$

where $\alpha, \beta \in (0,1)$ and $\gamma > 0$.

We use the following functional forms for aggregate consumption amenities A_s, production amenities O_s, and land supply L_s:

$$
(7) \qquad A_s = \prod_{j=1}^{J} A_{j,s}, \; O_s = \prod_{k=1}^{K} O_{k,s}, \; L_s = \prod_{m=1}^{M} L_{m,s},
$$

where

$$
\begin{aligned}
& A_{j,s} = a_{j,s} N_s^{\lambda_j}, \\
(8) \qquad & O_{k,s} = o_{k,s} N_s^{\mu_k}, \\
& L_{m,s} = l_{m,s} N_s^{\nu_m}.
\end{aligned}
$$

Each factor $A_{j,s}$, $O_{k,s}$, $L_{m,s}$ consists of exogenous features $a_{j,s}$, $o_{k,s}$, $l_{m,s}$ and agglomeration economy terms $N_s^{\lambda_j}$, $N_s^{\mu_k}$, $N_s^{\nu_m}$. We allow agglomeration economy parameters λ_j, μ_k, ν_m to differ across different factors.

These functional forms are quite standard and have some empirical and theoretical support. For example, Davis and Ortalo-Magne (2011) show that expenditure shares on housing are constant over time and across MSAs, which justifies the Cobb-Douglas preference. In production technology, the Cobb-Douglas function is arguably the most commonly used production function in economics, although this ubiquity alone may not justify its plausibility. For housing supply, we use the constant elasticity supply function, which is commonly used in empirical literature. Aggregate land supply L can be interpreted as the angle in which a monocentric city can expand. For example, imagine a monocentric city located on a coast, so that it can expand only in 180 degrees. Its land supply L would be half as large as that of a city that could expand in full 360 degrees.

The key requirement for the functional forms is that they all be multiplicatively separable. The multiplicative terms become linear when we take logs; this feature allows us to express the log city size as the sum of the log exogenous features $a_{j,s}$, $o_{k,s}$, $l_{m,s}$. We apply the central limit theorem over the sum and obtain the log-normal distribution for city size. Therefore, our mechanism does not work under utility or production functions that are not multiplicatively separable (e.g., Constant Elasticity of Substitution (CES) functions).

With these functional forms, we derive our theoretical results. We first show that the equilibrium population size of each city is unique and stable. Suppose that S and $\bar{u}$ are fixed. Using equations (2) and (3), we can solve for wage w_s and rent r_s as the functions of N_s. By substituting these into w_s and r_s in the indirect utility function V, we can express the indirect utility function in terms of only population size N.

$$\tilde{V}_s(N) = \Phi \left(\frac{1}{N}\right)^{\frac{1-\Omega}{\Phi_A}} \left\{ \prod_{j=1}^{J} (a_{j,s})^{\Phi_A} \prod_{k=1}^{K} (o_{k,s})^{\Phi_O} \prod_{m=1}^{M} l_{m,s} \right\}^{\frac{1}{\Phi_A}} \tag{9}$$

where Φ_A, Φ_O, and Φ are *positive* constants and $\Omega \equiv \Phi_A \sum_{j=1}^{J} \lambda_j + \Phi_O \sum_{k=1}^{K} \mu_k + \sum_{m=1}^{M} \nu_m$ is the aggregate agglomeration economy parameter, a weighted sum of all the agglomeration economy parameters in all factors determining consumption and production amenities and land supply.[5] This indirect utility function $\tilde{V}_s(N)$ is the utility city s offers when its population size is N. Equilibrium population size N_s is determined by the remaining equation (1):

$$\tilde{V}_s(N_s) = \bar{u}. \tag{10}$$

Suppose $\Omega < 1$. Equation (9) implies that $\tilde{V}_s(N)$ is continuous and strictly decreasing in

[5] $\Phi_A = \frac{1+\beta\gamma}{1-\alpha\beta}, \Phi_O = \frac{\alpha+\gamma}{1-\alpha\beta}, \Phi = (1-\alpha)^{1-\alpha}\alpha^{\alpha}(1-\beta)^{-\frac{(-1+\beta)(\alpha+\gamma)}{1+\beta\gamma}}\beta(1-\alpha\beta)^{\frac{-1+\alpha\beta}{1+\beta\gamma}}$.

N with $\lim_{N \to 0} \tilde{V}_s(N) = \infty$ and $\lim_{N \to \infty} \tilde{V}_s(N) = 0$. Thus, the intermediate value theorem implies that there exists a unique positive N_s satisfying equation (10) for any positive $\bar{u}$. In addition, this population size N_s is stable in the Krugman (1991) sense. For example, suppose that a negative temporary shock hits city s and its population size decreases below equilibrium city size N_s. With a smaller population size, the utility the city offers is greater than the common utility level $\bar{u}$, and thus its population size returns to equilibrium size N_s. Thus, if $\Omega < 1$, each site is populated with a unique and stable population size. On the other hand, if $\Omega > 1$, equilibrium city size becomes unstable (i.e., $\tilde{V}_s(N)$ is increasing in N), and we obtain a degenerate equilibrium with all people going to one city. In this case, a population distribution function N_s over $s \in [0, 1]$ would generate a *spike* at one particular s where all the mass $\bar{N}$ is located.

The intuition behind this stability result is the following: as city size decreases, housing prices decrease, allowing the city to offer higher utility. However, the downside of losing population is the loss in the agglomeration economies in consumption and production amenities, and in land supply. If the agglomeration economy parameter Ω is less than 1, the housing effect dominates, so the city offers better utility with a smaller population size, which makes the equilibrium city size stable. On the other hand, if Ω is greater than 1, the agglomeration effect dominates, so a city offers better utility with a larger population size. This effect makes the equilibrium city size unstable, and leads to the degenerate equilibrium where all people live in only one city.

The strictly decreasing $\tilde{V}_s(N)$ implies that cities are generally operating under aggregate decreasing returns. This feature is at odds with the usual bell-shaped curve that many urban economists believe is driving city size.[6] Conceptually, we may be able to generate the bell-shaped curve by tweaking the agglomeration economies for small cities, or by introducing fixed costs in developing a city. We do not attempt either, in order to keep our model

[6] Many urban economists believe that the indirect utility function is increasing for small population size but is decreasing for large population size, creating the bell-shaped curve. This happens because agglomeration economies dominate agglomeration costs when city size is small, while agglomeration costs dominate agglomeration economies when city size is big.

tractable. Moreover, another equilibrium city size that may exist on the increasing portion of the indirect utility curve would not be stable.

From equation (10) we obtain equilibrium population size N_s:

$$N_s = \left\{ \left(\frac{\Phi}{\bar{u}} \right)^{\Phi_A} \prod_{j=1}^{J} (a_{j,s})^{\Phi_A} \prod_{k=1}^{K} (o_{k,s})^{\Phi_O} \prod_{m=1}^{M} l_{m,s} \right\}^{\frac{1}{1-\Omega}} . \tag{11}$$

Equilibrium population size N_s is strictly increasing in production amenities $a_{j,s}$, consumption amenities $o_{k,s}$, and supply factors $l_{m,s}$, and strictly decreasing in the common utility level $\bar{u}$.

So far we have taken the equilibrium utility level $\bar{u}$ and the set of populated sites S as given. The equilibrium utility level $\bar{u}$ is unique given the set of populated sites S. This property follows from equation (4) because the population size of each city is continuous and strictly decreasing in $\bar{u}$ with $\lim_{\bar{u}\to 0} N_s(\bar{u}) = \infty$ and $\lim_{\bar{u}\to\infty} N_s(\bar{u}) = 0$, as can be seen in equation (11). The intermediate value theorem implies that equation (4) is satisfied for only one value of $\bar{u}$. The set of populated sites S is equal to the set of all locations $[0, 1]$, because the indirect utility $\lim_{N\to 0} \tilde{V}_s(N) = \infty$ and zero population size is not stable.

Proposition 1 *Suppose* $\Omega < 1$.

(a) Every site is populated and its population size is unique and stable.

(b) Population size N_s *of city* s *is increasing in consumption amenity factor* $a_{j,s}$*, production amenity factor* $o_{k,s}$*, and land supply factor* $l_{m,s}$ *(*$j \in J, k \in K$*,* $m \in M$*).*

Now we derive our key result that if the exogenous factors a_j, o_k, and l_m are randomly distributed, population size N converges in distribution to the log-normal distribution as the number of these factors increases. We interpret $a_{j,s}$, $o_{k,s}$, and $l_{m,s}$ as the realizations of random variables a_j, o_k, and l_m and thus no longer show city index s. Taking the log transformation of equation (11) we obtain

$$\log N = \frac{1}{1-\Omega}\left\{\sum_{j=1}^{J}\log a_j^{\Phi_A} + \sum_{k=1}^{K}\log o_k^{\Phi_O} + \sum_{m=1}^{M}\log l_m + \Phi_A \log\left(\frac{\Phi}{\bar{u}}\right)\right\}. \tag{12}$$

Mathematically, $a_j^{\Phi_A}$, $o_k^{\Phi_O}$, and l_m play the same roles. In order to simplify notations, we introduce a new symbol x_i, which we use for all three types of exogenous random factors. We can reorder the attribute terms

$$(a_1^{\Phi_A}, ..., a_J^{\Phi_A}, o_1^{\Phi_O}, ..., o_K^{\Phi_O}, l_1, ..., l_M)$$

as we like and assign $x_1, ..., x_I$ where $I \equiv J + K + M$. We rewrite equation (12) using the new notations as

$$\log N^I = \frac{1}{1-\Omega}\left\{\sum_{i=1}^{I}\log x_i + \Phi_A \log\left(\frac{\Phi}{\bar{u}}\right)\right\}. \tag{13}$$

We will apply the central limit theorem to $\sum_{i=1}^{I}\log x_i$. To further simplify notations, we define $X_i \equiv \log x_i$.

$$\begin{aligned}\log N^I &= \frac{1}{1-\Omega}\left\{\sum_{i=1}^{I} X_i + \Phi_A \log\left(\frac{\Phi}{\bar{u}}\right)\right\} \\ &= \frac{1}{1-\Omega}\left\{\sum_{i=1}^{I} \hat{X}_i + \sum_{i=1}^{I} \bar{X}_i + \Phi_A \log\left(\frac{\Phi}{\bar{u}}\right)\right\}\end{aligned} \tag{14}$$

where $\hat{X}_i$ is demeaned X_i (i.e., $\hat{X}_i \equiv X_i - \bar{X}_i$ where $\bar{X}_i \equiv E(X_i)$), so that $E\left(\hat{X}_i\right) = 0$. To show that city size distribution converges to the log-normal distribution, it suffices to show that $\sum_{i=1}^{I} \hat{X}_i$ in equation (14) converges to the normal distribution as I increases.

The classical central limit theorem states that $\sum_{i=1}^{I} \hat{X}_i$ converges in distribution to normal distribution if $\hat{X}_i$s are independent and identically distributed. Since this requirement is too restrictive in our context, we use a modern central limit theorem that relaxes these requirements: different random variables $\hat{X}_i$ can come from different distributions and can, to some degree, be correlated. Kourogenis and Pittis (2008) provide an excellent survey of modern central limit theorems. The version we use corresponds to Theorem 4 in Kourogenis and Pittis (2008), which in turn is based on Corollary 1 in Herrndorf (1984). We begin by describing the correlation structure among the random variables using $\alpha - mixing$.

Definition 2 *For a sequence $\hat{X}_1, \hat{X}_2, ...$ of random variables, let $\{\alpha_i | i \in \mathbb{N}\}$ be a sequence such that*

$$(15) \qquad |P(A \cap B) - P(A)P(B)| \leq \alpha_i \text{ for } A \in \sigma(\hat{X}_1, ..., \hat{X}_n) \text{ and } B \in \sigma(\hat{X}_{n+i}, \hat{X}_{n+i+1}, ...)$$

where $\sigma(\mathbf{X})$ is defined as the σ-algebra generated by $\mathbf{X}$ and $P(.)$ is a probability measure defined on A and B. If $\alpha_i \to 0$ as $i \to \infty$, the sequence $\hat{X}_1, \hat{X}_2, ...$ is said to be $\alpha - mixing$.[7]

The $\alpha - mixing$ means that random variables are approximately independent when they are sufficiently far apart from each other. As an extreme example, suppose that $\alpha_i = 0$. Two events A and B generated by $\left\{\hat{X}_1, ..., \hat{X}_n\right\}$ and $\left\{\hat{X}_{n+s}, \hat{X}_{n+s+1}, ...\right\}$ are completely independent from each other because equation (15) implies $P(A \cap B) = P(A)P(B)$. In our context the $\alpha - mixing$ means that if we arrange random features affecting city size such that more correlated ones are positioned closer together, two random features that are sufficiently far apart from each other would be virtually independent. For example, we can think of current climate and agriculture productivity as being set close together, while current climate and natural resources are relatively distant from each other. Now we state the central limit theorem.

Theorem 3 *(Herrndorf (1984)) Let $\left\{\hat{X}_i, i \in \mathbb{N}\right\}$ be an α-mixing sequence of random vari-*

[7] Roughly speaking, σ-algebra (or σ-field) generated by $\mathbf{X}$ is a set of all events that are distinguishable by $\mathbf{X}$ values. Refer to Billingsley (1995) for details.

ables satisfying the following conditions.

$$\begin{aligned}
1)\ E\left(\hat{X}_i\right) &= 0 \\
2)\ \lim_{i\to\infty} E\left(\frac{\left(\sum_{j=1}^{i}\hat{X}_j\right)^2}{i}\right) &= \sigma^2, 0<\sigma^2<\infty \\
3)\ \sup_{i\in\mathbf{N}} E\left|\hat{X}_i\right|^b &< \infty \text{ for some } b>2 \\
4)\ \sum_{s=1}^{\infty}(\alpha_s)^{1-\frac{2}{b}} &< \infty.
\end{aligned}$$

Then $\frac{1}{\sigma\sqrt{i}}S_i$ *converges in distribution to the standard normal distribution* $N(0,1)$.

The first condition is satisfied since we construct $\hat{X}_i$ so that $E\left(\hat{X}_i\right)=0$. The second condition means that the variance of the partial sum behaves nicely. The third condition requires that the moments of order $b>2$ be uniformly bounded. The fourth condition puts a restriction on the $\alpha-$mixing rate. The third and fourth conditions are linked by b. As b increases, the third condition becomes harder to satisfy, and the fourth condition becomes easier to satisfy. By applying Theorem 3, we obtain our main result.

Proposition 4 *If the sequence* $\hat{X}_1, \hat{X}_2, \ldots$ *satisfies the conditions listed in Theorem 3, city size converges in distribution to the log normal. Asymptotically, city size* N^I *with* I *random factors follows* $\log N\left(\frac{1}{1-\Omega}\left\{\sum_{i=1}^{I}\bar{X}_i+\Phi_A\log\left(\frac{\Phi}{\bar{u}}\right)\right\}, \frac{\sigma^2 I}{(1-\Omega)^2}\right)$.

4 Zipf's Law

The previous section shows that our model can generate the log-normal distribution when we have enough random factors. However, Zipf's law, which is known to hold in the top tail of city size distribution, emerges when city size follows the Pareto distribution with its shape parameter equal to 1. This section shows that our model is consistent with Zipf's law in the top tail, in the following ways. First, we analytically prove that the model, with certain parameter values, can generate a city size distribution whose estimated Pareto shape parameter is 1. Second, we use Monte-Carlo simulations to show that we get a good approximation of the Pareto distribution in the top tail even with a small number of factors that are substantially correlated.

We begin by showing that Zipf's law emerges when city size follows the Pareto distribution with its shape parameter equal to 1. Suppose that city size N follows the Pareto distribution with the scale parameter $\widetilde{N}$ and the shape parameter η:

$$CDF(N) = 1 - \left(\frac{\widetilde{N}}{N}\right)^{\eta} \text{ for } N > \tilde{N}.$$

When there are M cities in total, the rank R_s of city s with population size N_s can be approximated as

$$R_s \approx M(1 - CDF(N_s)) = M\left(\frac{\widetilde{N}}{N_s}\right)^{\eta}. \tag{16}$$

When $\eta = 1$, we obtain Zipf's law:

$$N_s \approx \frac{1}{R_s} \cdot M\widetilde{N}.$$

In other words, the population size N_s of a city is inversely proportional to its rank R_s.

A typical test of Zipf's law examines whether the Pareto coefficient η is equal to 1, and how well the empirical city size distribution fits the Pareto distribution. By taking logs on equation (16), we obtain

$$\begin{aligned} \log R_s &\approx \log M + \eta \log\left(\widetilde{N}\right) - \eta \log N_s \\ &= C - \eta \log N_s \end{aligned} \tag{17}$$

where $C \equiv \log M + \eta \log\left(\tilde{N}\right)$. One can estimate η by regressing $\log R_s$ on $\log N_s$ and check the goodness of fit by calculating R^2.

We now prove that our model can generate the Pareto coefficient equal to 1 by adjusting model parameters. Using equation (13) we can express the Zipf coefficient η as

$$\eta = -\frac{Cov\left(\log N, \log R\right)}{Var\left(\log N\right)} = \left(1-\Omega\right)\frac{-Cov\left(\sum_{i=1}^{I}\log x_i, \log R\right)}{Var\left(\sum_{i=1}^{I}\log x_i\right)}. \tag{18}$$

Thus, the Zipf coefficient η depends on the agglomeration economy parameter Ω, the distribution of random factors $\{x_i\}$, and the number of random shocks I. $Cov\left(\sum_{i=1}^{I}\log x_i, \log R\right)$ is negative because $\sum_{i=1}^{I}\log x_i$ raises population size as shown in equation (14) and thus improves the ranking of a city (i.e., lowers R). Therefore, equation (18) implies that the Zipf coefficient η falls with Ω.

Proposition 5 *The Zipf coefficient η decreases with Ω.*

Proposition 5 links the Zipf coefficient to the agglomeration economies. This result is intuitive. A smaller Zipf coefficient results when city size distribution is more uneven. Greater agglomeration economies make big cities even bigger, thereby creating a more uneven city size distribution. Dobkins and Ioannides (2001) show that the Zipf coefficient for the U.S. has declined in the twentieth century. One explanation based on our model is that agglomeration economies became more important over this period.

Proposition 5 suggests that we can match the Zipf coefficient equal to 1 by adjusting Ω. Note that this result does not depend on the number of factors or the degree of correlation. Gabaix and Ioannides (2004) show that the OLS estimate of η is downward biased in a small sample and provide the magnitude of the bias for various sample sizes. This bias is not a problem for us, because we can match the Zipf coefficient with the bias taken into account.

Now we run Monte-Carlo simulations to examine how many factors would be required or how much correlation would be allowed. Since Proposition 5 shows that we can match the Zipf coefficient to 1 by adjusting Ω regardless of the number of random factors and the correlation structure, we focus on R^2 from the Zipf regression (17).

We proceed as follows: we fix the number of factors and the degree of correlation among the factors, generate 25,000 city sizes using our model, truncate the distribution to include only the top cities using three different cutoffs: 135, 366 and 940, and run the Zipf regression for each group of cities with different cut-offs. We use 25,000 clusters because this number is close to the number of places (25,359) in Eeckhout (2004) and clusters (23,499) in Rozenfeld et al. (2011). We truncate them at the top 135, 366, and 940 cities because these numbers are the thresholds commonly used in the literature (135 for major cities, 366 for MSAs, and 940 for CBSAs that include both MSAs and μSAs). We repeat this 10,000 times and calculate average R^2 and its standard deviation. We vary the number of factors and the degree of correlation and repeat the whole process.

We generate random factors x_is such that each factor follows the uniform distribution $[0, 1]$ and that the correlation between x_i and x_j is $\rho^{|i-j|}$. (See Appendix A for a detailed procedure.) We use the uniform distribution because it is very different from the Zipf distribution, and consider this simulation as the worst scenario. We do not report Ω, Φ_A, Φ and $\bar{u}$ because these parameters do not affect R^2 once the random factors x_is are determined.[8]

Figure 1 and Table 1 show the simulation results. We compare them with our benchmark R^2 that we obtain by running the Zipf regressions with 2009 U.S. population data.[9] The benchmark R^2s are .978 for the top 135 cities, .976 for the top 366 cities, and .974 for the top 940 cities. When the factors are uncorrelated (i.e., $\rho = 0$), the simulation takes 9, 8, and 9 factors to reach the benchmark R^2s for the top 135, 366, and 940 cities respectively. With $\rho = 0.2$, it takes 13, 12, and 11 factors. With $\rho = 0.4$, it takes 19, 18, and 16 factors.

The simulation results show that our simulated distribution can approximate the Zipf

[8]In order to see this, we obtain the following equation by inserting $\log N^I$ in equation (14) into equation (17):

$$\log R^j = C - \eta \cdot \frac{1}{1-\Omega} \left\{ \sum_{i=1}^{I} \log x_i^j + \Phi_A \log \left(\frac{\Phi}{\bar{u}} \right) \right\}$$

It is clear from this equation that changes in these parameters are fully absorbed by changes in OLS estimates $\hat{\eta}$ and $\hat{C}$ and thus do not affect predicted log rank and thus R^2.

[9]Data are available at http://www.census.gov/popest/data/metro/totals/2009/

distribution with reasonable numbers of random factors. Moreover, had we used other distributions than the uniform distribution, the simulation could have converged much more quickly. Our next Monte-Carlo simulation makes this point.

In our second simulation, we use the following process to generate random factors.

$$x_i = \sum_j^J \varepsilon_j. \tag{19}$$

where each ε_j follows i.i.d. uniform distribution $[0, k]$ where $k \in \mathbb{R}^+$. The idea is that each random factor x_i is made up of sub-factors ε_js. Figure 2 shows how the distribution for x_i changes as J increases. The solid line shows the p.d.f. for x_i with J sub-factors.[10] The central limit theorem states that x_i converge in distribution to the normal distribution as J increases. The dashed line shows the p.d.f. of the normal distribution.

We run a Monte-Carlo simulation similar to the first one, except that we vary J instead of ρ. The results in Figure 3 and Table 2 show that our model requires a much smaller number of random factors when each random factor follows a distribution closer to the bell-shaped normal distribution. $J = 1$ case is identical to $\rho = 0$ case in the first simulation and takes 8 factors to reach the benchmark R^2. When $J = 2$, the simulations reach the benchmark R^2 with 4 factors for all the groups of cities. With $J = 3$, the simulation reaches the benchmark R^2 with 3, 2, and 3 factors for the top 135, 366, and 940 cities respectively.

5 The Goodness of Fit to U.S. City Size Distribution

In this section, we fit our model to U.S. city size distribution data. The literature has used three different data sets employing different definitions of cities: CBSAs, Census places, and the area clusters defined by Rozenfeld et al. (2011). We fit our model to each data set and see how well it performs.

[10]The p.d.f. of x_i: $\frac{d}{dx}\int_{\varepsilon_1,,,\varepsilon_{J-1}\in[0,1]} Min(1, Max(0, x - \sum_{j=1}^{J-1}\varepsilon_j))$.

5.1 CBSAs

We begin with 2009 CBSA data. The data set has 940 CBSAs, which consist of 366 MSAs and 574 μSAs. The total number of people living in CBSAs is 288 million, which accounts for 94 percent of all population in the U.S. in 2009. MSAs have been a dominant definition of cities for urban economists. However, its population cutoff is 50,000, and MSAs thus miss small cities below this threshold. To address this issue, the Census Bureau first defined μSAs in 2003, with the same definition used to define MSAs, except that μSAs would require a lower population cutoff of 10,000. Duranton (2007) popularized the use of all CBSAs including μSAs in studying city size distribution.

Equation (13) summarizes our model's city size prediction. We rewrite the equation here for convenience.

$$\log N^I = \frac{1}{1-\Omega}\sum_{i=1}^{I}\log x_i + \Psi \tag{20}$$

where $\Psi \equiv \frac{\Phi_A}{1-\Omega}\log\left(\frac{\Phi}{\bar{u}}\right)$. In order to fit our model to the data, we have to pin down the distribution for x_i, the number of factors I, Ω, and Ψ. The key difficulty is that we do not know the distribution for x_i and thus Ω cannot be separately identified from x_i distribution. For example, for any values of Ω and $\{x_i\}_{i\in I}$, if we define $\Omega' \equiv 1-\varphi(1-\Omega)$ and $x_i' \equiv x_i^{\varphi}$ for any $\varphi \in \mathbb{R}$, we obtain the same city size because $\frac{1}{1-\Omega'} = \frac{1}{\varphi}\frac{1}{1-\Omega}$ and $\log x_i' = \varphi \log x_i$. For this reason we focus only on how well our model fits data for different x_i distributions, instead of trying to separately estimate each parameter; it is meaningless to report the point estimates of these parameters, since a simple transformation of x_i and Ω can generate the same fit.

As in our second Monte-Carlo simulation, we assume that x_i can be expressed as a sum of underlying sub-factors ε_j following the Uniform distribution (see equation (19)). We fit our model to data using three different numbers of sub-factors $J = 1, 2, 3$ and show that they fit similarly well. See Figure 2 for the distribution of x_i for each number of sub-factors.

We estimate the three parameters I, Ω, and Ψ by matching the first three theoretical moments of log city size: $E\left(\log N^I\right)$, $Var\left(\log N^I\right)$, and $E\left(\log N^I - E\left(\log N^I\right)\right)^3$ with corresponding empirical moments from the CBSA data. When we calculate the theoretical moments, we have to take into account the truncation issue of the CBSA data set. The CBSA data set has 940 cities. When we assume that the number of all cities is 25,000, as in the previous section, the CBSAs account for the top 3.76 percent of all cities.

We use Monte-Carlo simulations to calculate these theoretical moments for the truncated distribution. We proceed as follows. First, we fix $I \in \{1, 2, 3, ..., 50\}$. Second, we generate 1 million draws of $\sum_{i=1}^{I} \log x_i$ which is the only stochastic part in equation (20).[11] Third, since the CBSA data is truncated at the top 3.76 percent, we keep only the same percentage of simulated $\sum_{i=1}^{I} \log x_i$ in the top tail. Fourth, we calibrate Ω to match the second moment $Var\left(\log N^I\right)$. Fifth, we calibrate Ψ to match the first moment $E\left(\log N^I\right)$. Sixth, we calculate the distance between the simulated third central moment $E\left(\log N^I - E\left(\log N^I\right)\right)^3$ and its corresponding empirical moment. Now we go back to the first step and choose a different I and repeat the whole process. When we finish the iterations for all Is, we pick the I that minimizes the distance in the third central moment calculated in the sixth step. Note that the first and the second moments are matched exactly in the fourth and the fifth steps.

Figure 4 shows the result. To see how well our model performs, we employ the tools developed in Duranton (2007). We simulate our estimated model to generate 1,000 sets of city size distribution, each of which consists of 25,000 cities. Thus, for each rank of cities, we have 1,000 city sizes across different sets, from which we calculate the mean of log city sizes and the 95 percent confidence interval. In each diagram in Figure 4, the solid line in the middle is the mean log city size; the other two solid lines around the mean log city size indicate the 95 percent confidence interval. The dots are the CBSA data. Figure 4 also reports R^2s developed in Duranton (2007) and the population error used in Ioannides and Skouras (2013). Duranton (2007) defines R^2 as $1 - msd/var$ where $msd = \frac{1}{m}\sum$(Actual log

[11] Since we need $I \times J$ draws of $\varepsilon_{i,j}$s to get one draw of $\sum_{i=1}^{I} \log x_i$, one million draws of $\sum_{i=1}^{I} \log x_i$ requires that we draw $I \times J$ million draws of $\varepsilon_{i,j}$s.

Size − Mean log Simulated Size$)^2$ and var is the empirical variance for log city size. The population error is roughly the number of people that have to switch from a simulated city size distribution to the real city size distribution. We calculate it as $\frac{1}{2}\sum_i |\hat{s}_i - s_i|$ where $\hat{s}_i$ is the mean of the ith largest simulated city sizes and s_i is the ith largest city size in data.

Figure 4 also includes the fit of the log-normal distribution and the Zipf distribution to CBSA data as benchmarks. We fit the log-normal distribution by matching the first and the second moments, and fit the Zipf distribution by matching the first moment only. The results show that our model fits the CBSA data significantly better than the log-normal distribution or the Zipf distribution. Since our model fits the largest cities extremely well, the population error is much smaller than the log-normal distribution or the Zipf distribution. The R^2s are also higher in our model fit.

It may not be so surprising that our model fits the data better than the log-normal distribution or the Zipf distribution, because we have more parameters to match the data. For example, even though our model converges to the log-normal distribution as we add more factors, we can choose the number of factors to get a better fit. However, it is still surprising that our model does so well in the very top tail. In fact, as we will see in the next section, our model does not do as well when matched to the Census place data or the area cluster data.

5.2 Census Places and the Area Clusters

Now we fit our model to the other two data sets. The first data set is the Census 2000 places. The U.S. Census Bureau defines a place as a concentration of population, consisting of incorporated places such as cities, towns, or villages, and census-designated places (CDP), which the Census Bureau has defined for statistical purposes. Unlike the CBSAs, the place data set does not have a truncation issue; Eeckhout (2004) uses the data set to argue that size distribution for all cities follows a log-normal distribution. However, Levy (2009) and Ioannides and Skouras (2013) show that a log-normal distribution does not fit the entire distribution of places.

The second data set is the area clusters defined by Rozenfeld et al. (2011). They define cities by connecting nearby census tract centroids. If a distance between two centroids is within a threshold distance, they classify them as being in the same city. They define a city as the maximal connection of census tract centroids. They identify the threshold distance as 3km, so that the resulting cluster distribution in the top tail most closely fits the size distribution of MSAs. They define small clusters below the MSA cutoff of 50,000 using the same 3km threshold.

We use the same estimation procedure that we used to fit our model to the CBSA data, except that we do not drop small simulated cities, because these data sets are not truncated. The first row of Figure 5 shows the result; our model does not fit either data. This result is consistent with Ioannides and Skouras (2013)'s finding that the log-normal distribution, to which our model converges, does not fit the whole Census place data or the area cluster data. We show only the $J = 2$ case where we use two sub-factors, but the other cases also perform similarly.

There are two potential reasons why our model does not fit either the Census places or the area clusters. First, as noted in Ioannides and Skouras (2013), small cities may be subject to different economic fundamentals. For example, housing congestion in our model would not be as relevant in very small cities. Second, the data may be sensitive to the details of the definition for very small cities. For example, as discussed in Holmes and Lee (2010), the Census place definition depends heavily on arbitrary political decisions that determine which places become incorporated as Census places. With the area clusters, why the same threshold distance should be used to define cities throughout all city sizes is unclear; the same distance is likely to have different economic implications in large and small cities.

To see how well our model performs in the top tails, we fit our model to the top 3.76 percent of cities in the two data sets. The 3.76 percent is the percentage accounted for by CBSAs when we assume that the total number of cities is 25,000. The second row in Figure 5 shows that our model performs much better, although not as well as it did with the CBSAs.

6 Conclusion

This paper proposes a new model of city size distributions. Our model generates city size distributions that asymptotically follow the log-normal distribution. When we tune parameters, the model can match Zipf's law in the top tail, which is often observed in data. The key insight of our theory is that many factors determine city size; this plurality of determinants may lead to the empirical pattern we observe in the data. The key implication is that our model provides one mechanism by which a model can be consistent with empirical regularities in city size distribution, even when the model alone cannot generate empirical regularities by itself. This implication is important because classical urban economics models are sometimes criticized for not being able to generate Zipf's law.

References

Behrens, K., G. Duranton, and F. Robert-Nicoud (2010). Productive cities: Sorting, selection and agglomeration. Working paper, Department of Economics, University du Quebec a Montreal.

Berliant, M. and H. Watanabe (2009, December). Explaining the size distribution of cities: x-treme economies. MPRA Paper 19080, University Library of Munich, Germany.

Billingsley, P. (1995, April). *Probability and Measure* (3 ed.). Wiley-Interscience.

Black, D. and V. Henderson (2003). Urban evolution in the USA. *Journal of Economic Geography 3*(4), 343–372.

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

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

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

Desmet, K. and J. Rappaport (2013). The settlement of the united states, 1800 to 2000: The long transition towards gibrat's law. *Federal Reserve Bank of Kansas City Working Paper No. RWP 13-02.*.

Desmet, K. and E. Rossi-Hansberg (2013). Urban accounting and welfare. *American Economic Review*.

Dobkins, L. H. and Y. M. Ioannides (2001, November). Spatial interactions among u.s. cities: 1900-1990. *Regional Science and Urban Economics 31*(6), 701 – 731.

Duranton, G. (2006, July). Some foundations for zipf's law: Product proliferation and local spillovers. *Regional Science and Urban Economics 36*(4), 542–563.

Duranton, G. (2007, March). Urban evolutions: The fast, the slow, and the still. *American Economic Review 97*(1), 197 – 221.

Eeckhout, J. (2004, December). Gibrat's law for (all) cities. *American Economic Review 94*(5), 1429–1451.

Gabaix, X. (1999, August). Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114*(3), 739–767.

Gabaix, X. and Y. M. Ioannides (2004, December). The evolution of city size distributions. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4 of *Handbook of Regional and Urban Economics*, Chapter 53, pp. 2341–2378. Elsevier.

Glaeser, E. L. and J. D. Gottlieb (2009, December). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature 47*(4), 983–1028.

Glaeser, E. L., J. Gyourko, and R. E. Saks (2006). Urban growth and housing supply. *Journal of Economic Geography 6*(1), 71 – 89.

Glaeser, E. L., G. A. Ponzetto, and K. Tobio (2011, April). Cities, skills, and regional change. NBER Working Papers 16934, National Bureau of Economic Research, Inc.

Henderson, J. V. (1974, September). The sizes and types of cities. *American Economic Review 64*(4), 640–56.

Herrndorf, N. (1984, February). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability 12*(1), 141–153.

Holmes, T. J. and S. Lee (2010, September). Cities as six-by-six-mile squares: Zipf's law? In *Agglomeration Economics*, NBER Chapters, pp. 105–131. National Bureau of Economic Research, Inc.

Hsu, W.-T. (2012). Central place theory and city size distribution. *The Economic Journal*, no–no.

Ioannides, Y. and S. Skouras (2013). Us city size distribution: Robustly pareto, but only in the tail. *Journal of Urban Economics 73*(1), 18 – 29.

Ioannides, Y. M. and H. G. Overman (2003, March). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics 33*(2), 127–137.

Kourogenis, N. and N. Pittis (2008). Mixing conditions, central limit theorems and invariance principles: A survey of the literature with some new results on heteroscedastic sequences. *SSRN eLibrary*.

Krugman, P. (1991, June). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–99.

Krugman, P. (1996, December). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies 10*(4), 399 – 418.

Levy, M. (2009, September). Gibrat's law for (all) cities: Comment. *American Economic Review 99*(4), 1672–75.

Rappaport, J. (2008, November). Consumption amenities and city population density. *Regional Science and Urban Economics 38*(6), 533–552.

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

Roback, J. (1982, December). Wages, rents, and the quality of life. *Journal of Political Economy 90*(6), 1257–1278.

Rosen, K. T. and M. Resnick (1980, September). The size distribution of cities: An examination of the pareto law and primacy. *Journal of Urban Economics 8*(2), 165 – 186.

Rossi-Hansberg, E. and M. L. J. Wright (2007, April). Urban structure and growth. *Review of Economic Studies 74*(2), 597 – 624.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011, August). The area and population of cities: New insights from a different perspective on cities. *American Economic Review 101*(5), 2205–25.

Schumann, E. (2009). Generating correlated uniform variates. *COMISEF Wiki*, http://comisef.wikidot.com/tutorial:correlateduniformvariates/.

Simon, H. A. (1955, December). On a class of skew distribution functions. *Biometrika 42*(3/4), 425–440.

Soo, K. T. (2005, May). Zipf's law for cities: A cross-country investigation. *Regional Science and Urban Economics 35*(3), 239–263.

A How to Generate Correlated Uniformly Distributed Random Variables

We follow the instruction by Schumann (2009). Suppose that we want to generate a vector of random variables $\mathbf{U}$ that are uniformly distributed and correlated with the Bravis-Pearson correlation matrix $\mathbf{\Sigma}$.

For random variables X_1 and X_2 following any c.d.f. F, it is known that $F(X_1)$ and $F(X_2)$ follow Uniform distribution. It is also known that the Bravis-Pearson correlation between $F(X_1)$ and $F(X_2)$ is equal to the Spearman (rank) correlation between X_1 and X_2:

$$\rho^B_{F(X_1),F(X_2)} = \rho^S_{X_1,X_2}.$$

Thus, if we generate X_1 and X_2 with the Spearman correlation ρ, we obtain two uniformly distributed variables $F(X_1)$ and $F(X_2)$ that have the Bravis-Pearson correlation ρ.

Since we cannot specify the Spearman correlation when we generate random numbers in a computer, we have to find the Bravis-Pearson correlation ρ^B corresponding to the Spearman correlation ρ. For the normal distribution, the exact relationship between the two correlations is known as

$$\rho^B = 2\sin\left(\frac{\pi}{6}\rho\right).$$

In summary, we use the following steps to generate a vector of random variables $\mathbf{U}$ that are uniformly distributed and correlated with the Bravis-Pearson correlation matrix $\mathbf{\Sigma}$. First, we compute $\mathbf{\Sigma}^{adj} = 2\sin(\pi/6 \cdot \mathbf{\Sigma})$. Second, we create $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Sigma}^{adj})$. Third, we compute $\mathbf{U} = F(\mathbf{Z})$.

Table 1: Average R^2's from Monte-Carlo Simulation A

Correlation	$\rho = 0$		$\rho = 0.2$		$\rho = 0.4$	
# of Factors	Mean	S.D.	Mean	S.D.	Mean	S.D.
Panel A: Top 135 Cities						
6	0.973	(0.015)	0.966	(0.016)	0.953	(0.018)
7	0.975	(0.014)	0.970	(0.015)	0.959	(0.017)
8	0.977	(0.013)	0.973	(0.015)	0.964	(0.017)
9	0.978	(0.013)	0.974	(0.014)	0.967	(0.016)
10	0.979	(0.012)	0.976	(0.014)	0.969	(0.015)
11	0.980	(0.012)	0.977	(0.013)	0.971	(0.015)
12	0.980	(0.012)	0.978	(0.013)	0.973	(0.015)
13	0.981	(0.012)	0.979	(0.013)	0.974	(0.014)
14	0.981	(0.012)	0.979	(0.013)	0.975	(0.014)
15	0.981	(0.012)	0.979	(0.012)	0.976	(0.014)
16	0.981	(0.012)	0.980	(0.012)	0.977	(0.013)
17	0.982	(0.011)	0.980	(0.012)	0.978	(0.013)
18	0.982	(0.011)	0.981	(0.012)	0.978	(0.013)
19	0.982	(0.011)	0.981	(0.012)	0.978	(0.013)
20	0.982	(0.011)	0.981	(0.012)	0.979	(0.013)
Panel B: Top 366 Cities						
6	0.969	(0.011)	0.960	(0.011)	0.944	(0.012)
7	0.973	(0.010)	0.965	(0.011)	0.952	(0.012)
8	0.975	(0.010)	0.969	(0.011)	0.957	(0.012)
9	0.977	(0.010)	0.971	(0.010)	0.961	(0.011)
10	0.979	(0.009)	0.974	(0.010)	0.965	(0.011)
11	0.980	(0.009)	0.975	(0.010)	0.967	(0.011)
12	0.980	(0.009)	0.976	(0.010)	0.969	(0.011)
13	0.981	(0.009)	0.978	(0.009)	0.971	(0.010)
14	0.982	(0.009)	0.978	(0.009)	0.972	(0.010)
15	0.982	(0.009)	0.979	(0.009)	0.974	(0.010)
16	0.983	(0.009)	0.980	(0.009)	0.975	(0.010)
17	0.983	(0.008)	0.981	(0.009)	0.976	(0.010)
18	0.984	(0.008)	0.981	(0.009)	0.977	(0.010)
19	0.984	(0.008)	0.982	(0.009)	0.977	(0.010)
20	0.985	(0.008)	0.982	(0.009)	0.978	(0.009)
Panel C: Top 940 Cities						
6	0.961	(0.007)	0.950	(0.008)	0.933	(0.008)
7	0.965	(0.007)	0.956	(0.008)	0.941	(0.008)
8	0.969	(0.007)	0.960	(0.008)	0.947	(0.008)
9	0.971	(0.007)	0.964	(0.007)	0.952	(0.008)
10	0.973	(0.007)	0.966	(0.007)	0.956	(0.008)
11	0.974	(0.007)	0.968	(0.007)	0.958	(0.008)
12	0.975	(0.007)	0.970	(0.007)	0.961	(0.007)
13	0.977	(0.007)	0.972	(0.007)	0.963	(0.007)
14	0.978	(0.007)	0.973	(0.007)	0.965	(0.007)
15	0.978	(0.006)	0.974	(0.007)	0.967	(0.007)
16	0.979	(0.006)	0.975	(0.007)	0.968	(0.007)
17	0.979	(0.006)	0.976	(0.007)	0.969	(0.007)
18	0.980	(0.006)	0.976	(0.007)	0.970	(0.007)
19	0.980	(0.006)	0.977	(0.007)	0.971	(0.007)
20	0.981	(0.006)	0.978	(0.006)	0.972	(0.007)

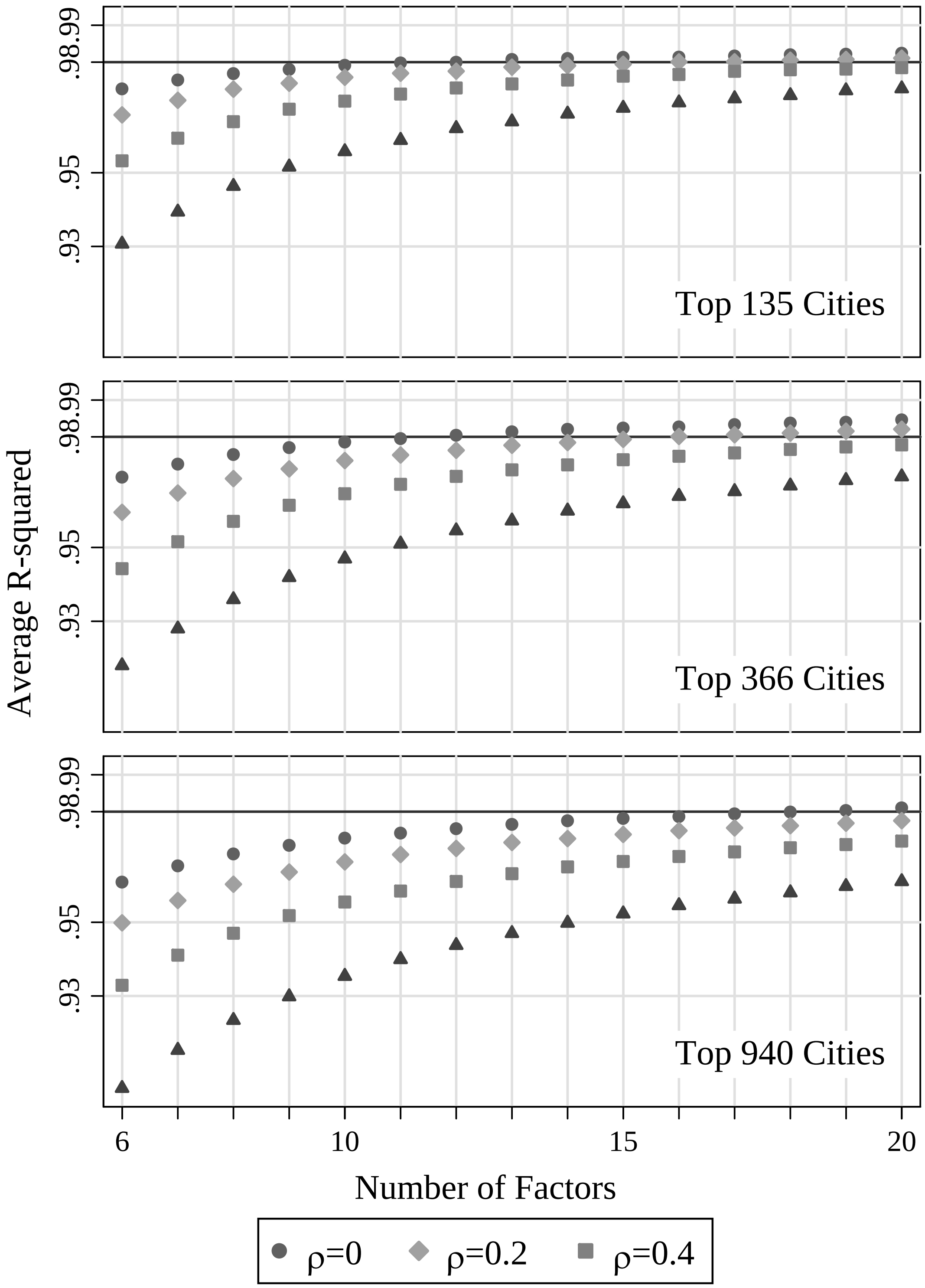
Note: "S.D." reports the standard deviation of the R-Squared.

Table 2: Average R^2's from Monte-Carlo Simulation B

# of Subfactors	J=1		J=2		J=3	
# of Factors	Mean	S.D.	Mean	S.D.	Mean	S.D.
Panel A: Top 135 Cities						
2	0.913	(0.022)	0.964	(0.016)	0.977	(0.014)
3	0.947	(0.019)	0.976	(0.014)	0.981	(0.011)
4	0.962	(0.017)	0.980	(0.012)	0.982	(0.011)
5	0.969	(0.016)	0.982	(0.011)	0.983	(0.011)
6	0.973	(0.015)	0.982	(0.011)	0.983	(0.011)
7	0.975	(0.014)	0.982	(0.011)	0.983	(0.011)
8	0.977	(0.013)	0.983	(0.011)	0.983	(0.011)
9	0.978	(0.013)	0.983	(0.011)	0.983	(0.011)
10	0.979	(0.012)	0.983	(0.011)	0.983	(0.011)
Panel B: Top 366 Cities						
2	0.899	(0.014)	0.959	(0.011)	0.976	(0.010)
3	0.938	(0.013)	0.975	(0.010)	0.982	(0.009)
4	0.955	(0.012)	0.980	(0.009)	0.984	(0.008)
5	0.964	(0.011)	0.983	(0.008)	0.985	(0.008)
6	0.969	(0.011)	0.984	(0.008)	0.986	(0.008)
7	0.973	(0.010)	0.985	(0.008)	0.986	(0.007)
8	0.975	(0.010)	0.985	(0.008)	0.987	(0.007)
9	0.977	(0.010)	0.986	(0.008)	0.987	(0.007)
10	0.979	(0.009)	0.986	(0.008)	0.987	(0.007)
Panel C: Top 940 Cities						
2	0.886	(0.009)	0.952	(0.008)	0.971	(0.007)
3	0.926	(0.008)	0.970	(0.007)	0.978	(0.006)
4	0.945	(0.008)	0.976	(0.007)	0.981	(0.006)
5	0.955	(0.008)	0.978	(0.006)	0.982	(0.006)
6	0.961	(0.007)	0.980	(0.006)	0.983	(0.006)
7	0.965	(0.007)	0.981	(0.006)	0.984	(0.006)
8	0.969	(0.007)	0.982	(0.006)	0.984	(0.006)
9	0.971	(0.007)	0.982	(0.006)	0.985	(0.006)
10	0.973	(0.007)	0.983	(0.006)	0.985	(0.006)

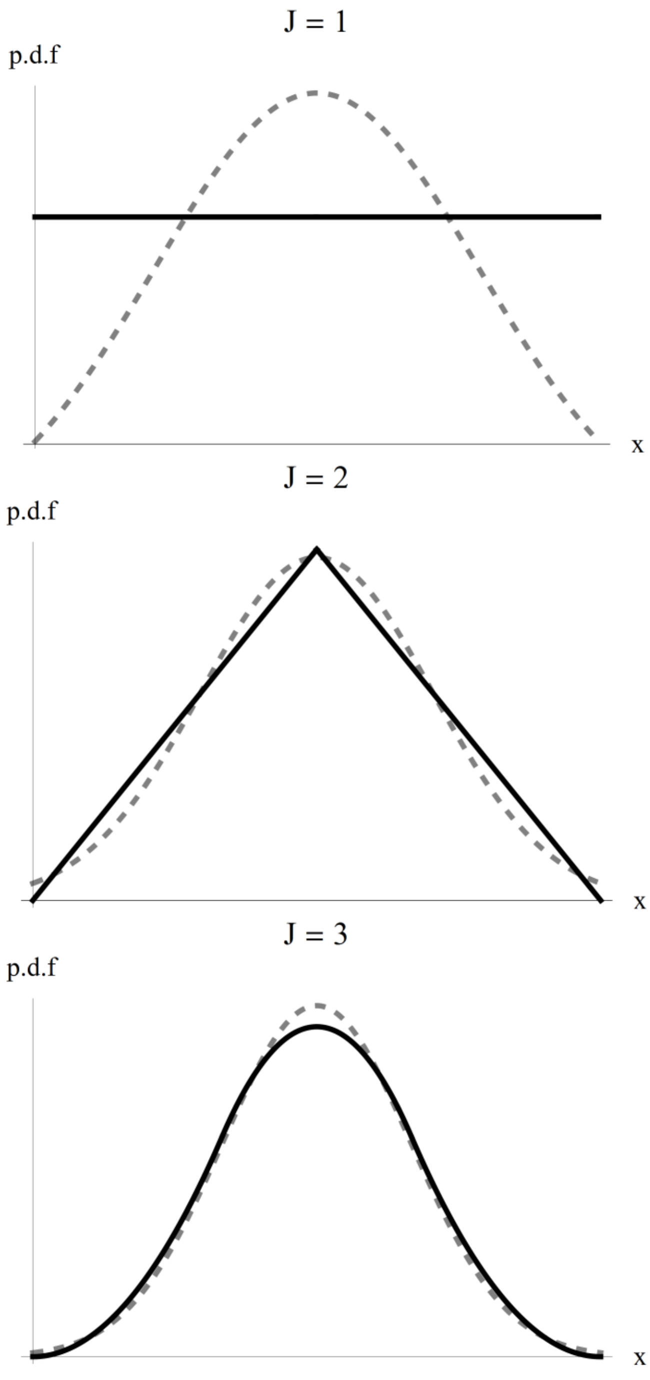
Note: "S.D." reports the standard deviation of the R-Squared.

Figure 1. Monte-Carlo Simulation A



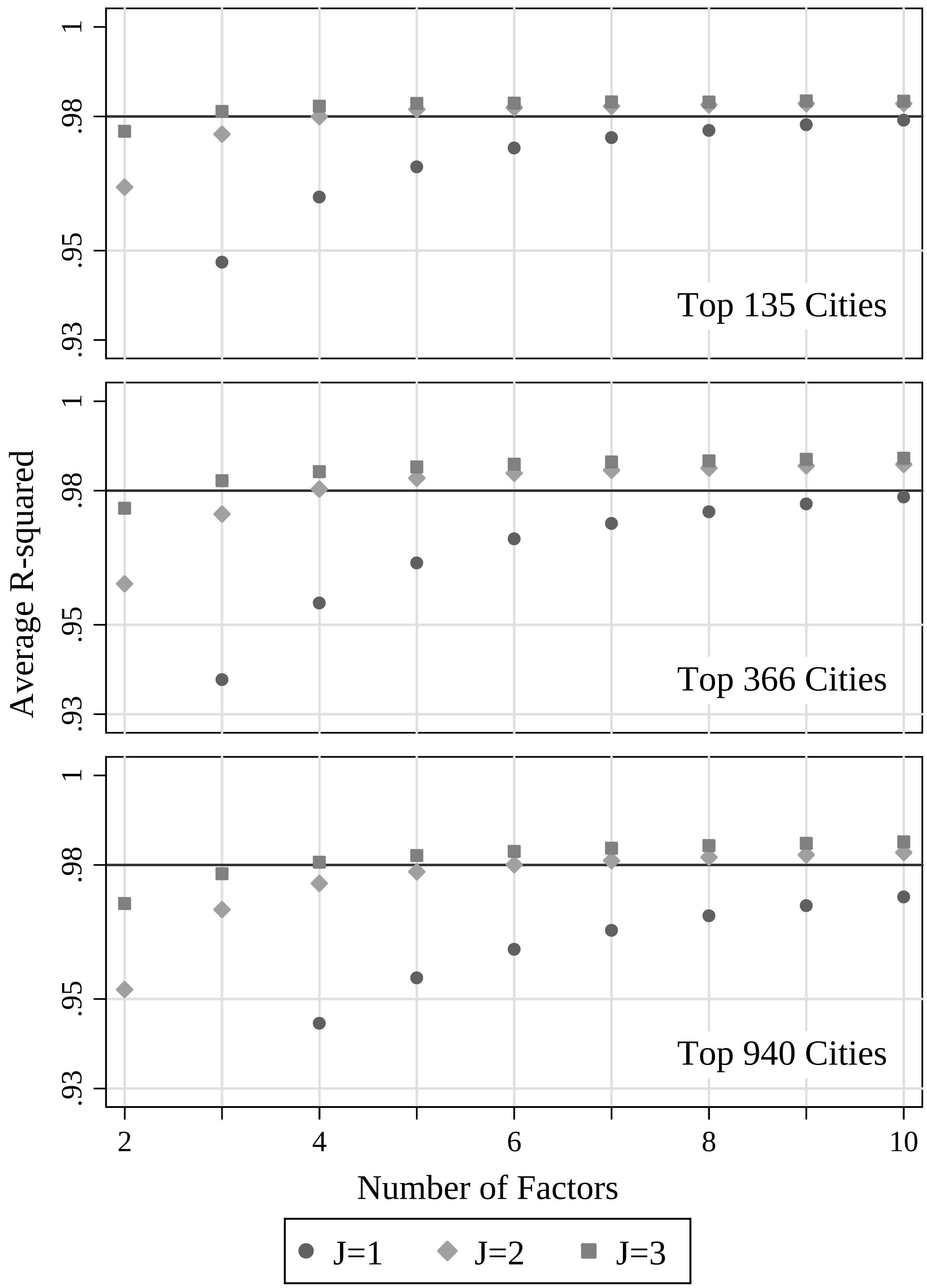
* Correlation between random factors i and j is $\rho^{|i-j|}$. Each factor follows Uniform distribution. See Table 1 for precise values.

Figure 2. The Distribution of Random Factor x_i with J Sub-factors



* Dashed lines are the p.d.f. of the normal distribution.

Figure 3. Monte-Carlo Simulation B



* Each factor is a sum of J sub-factors. Each sub-factor follows i.i.d. Uniform distribution. See Table 2 for precise values.

Figure 4. The Goodness of Fit to CBSAs

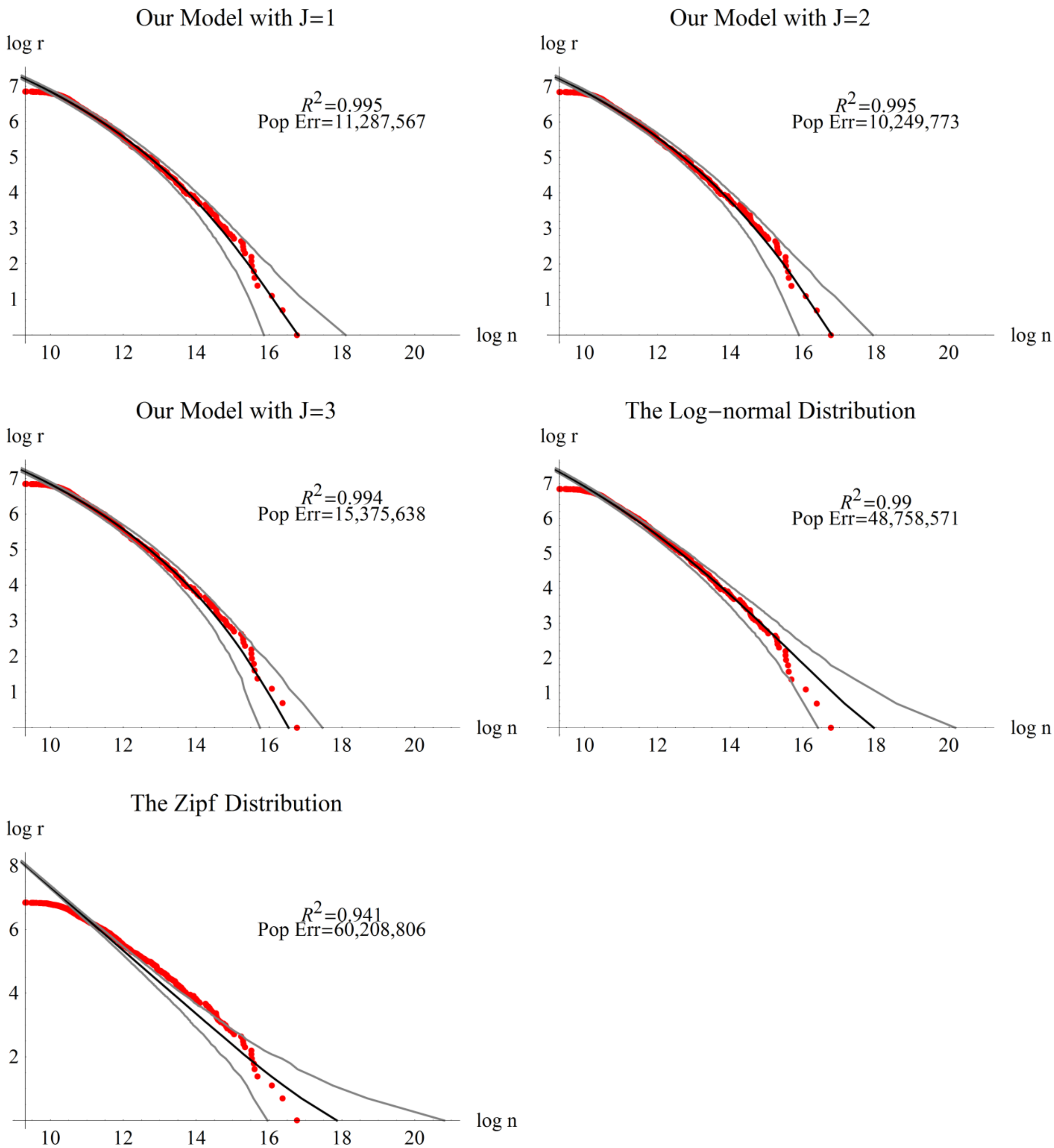


Figure 5. The Goodness of Fit of Our Model (J=2) to Census Places and the Area Clusters