Journal of Urban Economics 78 (2013) 19–29

Contents lists available at SciVerse ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Uneven landscapes and city size distributions

Sanghoon Lee [a,*], Qiang Li [b]

[a] Sauder School of Business, The University of British Columbia, 2053 Main Mall, Vancouver, BC, Canada V6T 1Z2
[b] Department of Real Estate, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore

ARTICLE INFO

Article history:
Received 14 August 2012
Revised 25 April 2013
Available online 16 May 2013

JEL classification:
D39
R12

Keywords:
City size distribution
Zipf's law
Rank-size rule
Log-normal distribution

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.

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, 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 US 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

* Corresponding author.
E-mail addresses: sanghoon.lee@sauder.ubc.ca (S. Lee), rstlq@nus.edu.sg (Q. Li).

[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.

0094-1190/$ - see front matter
http://dx.doi.org/10.1016/j.jue.2013.05.001

Rozenfeld et al. (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; Duranton, 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 et al. (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 et al. (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 need and how much correlation is allowed among the factors. Section 5 shows how well our model can fit the US 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 et al., 2006; Rappaport, 2008; Glaeser and Gottlieb, 2009, and Desmet and Rossi-Hansberg, forthcoming). 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, and zoning. 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 $L \in \mathbb{R}$.

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

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.

$\overline{N}$ 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:

$$\max_s V(r_s, w_s; A_s)$$

where

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

[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.