US city size distribution revisited: Theory and empirical evidence

Ramos, Arturo and Sanz-Gracia, Fernando

Universidad de Zaragoza, Universidad de Zaragoza

30 April 2015

US city size distribution revisited: Theory and empirical evidence

ARTURO RAMOS[a] FERNANDO SANZ-GRACIA[b]

May 6, 2016

Abstract

We have developed an urban economic model in which a social planner maximizes the net output of the whole system of cities in a country in such a way that agents locate themselves in cities of different sizes. From this model we derive the new "threshold double Pareto Generalized Beta of the second kind". In order to test the theory empirically, we have analysed the US urban system and have considered two types of data (incorporated places from 1900 to 2000 and all places in 2000 and 2010). The results are encouraging because the new distribution always outperforms the lognormal and the double Pareto lognormal. The results are robust to a number of different criteria. Thus, the new density function describes accurately the US city size distribution and, therefore, tends to support the validity of the theoretical model.

Keywords: human capital; congestion costs; lower tail, body, and upper tail; Pareto and Generalized Beta of the second kind distributions

JEL: C46, D39, R11, R12

[a]Department of Economic Analysis, *Universidad de Zaragoza* (SPAIN) aramos@unizar.es. We would like to thank Rafael González-Val and Fernando Pueyo for useful comments on previous versions of this paper. This work is supported by the Spanish *Ministerio de Economía y Competitividad* (ECO2013-45969-P) and *Gobierno de Aragón*, ADETRE Consolidated Group.

[b]Department of Economic Analysis, *Universidad de Zaragoza* (SPAIN) fsanz@unizar.es

1 Introduction

Cities are complex systems, which tend to self-organize, and where everything is interconnected (Batty, 2013; Bettencourt and West, 2010). Therefore, their study can and must be addressed from different points of view. This work aims to shed some light, both from a theoretical and an empirical perspective, to the analysis of city size distributions.

The literature on city size distributions is ample. Without pretending to be exhaustive, and citing only contributions of this century, we have Overman and Ioannides (2001); Black and Henderson (2003); Ioannides and Overman (2004); Eeckhout (2004); Resende (2004); Soo (2005); Bosker et al. (2008); Xu and Zhu (2009); Giesen et al. (2010); Berry and Okulicz-Kozaryn (2012); Ioannides and Skouras (2013); Luckstead and Devadoss (2014); González-Val et al. (2015); Berliant and Watanabe (2015) and Fazio and Modica (2015). We will point out in the next paragraph the main characteristics of this body of literature.

First, the most studied geographic area is that of the United States. Second, the two most studied distributions are the Pareto I or power law (a particular case of this is the so-called Zipf's law) and the lognormal. Third, the definition of what is considered a city is not neutral to the results obtained finally. Indeed, researchers in this field usually have to take two decisions: the consideration or not of a truncation point of the population variable (and, if affirmative, of what size) and the specific definition of the objects of study. Fourth, there is some consensus (Desmet and Rappaport (2015) take it for granted as the basis of their article), that the overall US city size distribution is lognormal and approximately Zipf at the upper tail, or at least Pareto.[1] Lastly, a number of recent papers argue that for an excellent fit to the data for the whole range of possible sizes, it is necessary to consider more than a single functional form, since the different parts of the distribution behave differently.

[1] An exception to this consensus can be found in Bee et al. (2013).

The theory will be addressed in Section 4, where we will present an urban model, based on Parker (1999), in which a social planner distributes the population so as to maximize the net output of the system of cities of the country. From this model the new density function proposed in this paper, called the "threshold double Pareto Generalized Beta of the second kind", (tdPGB2), can be deduced.

In the empirical part of the paper, with US city data from 1900 to 2010, we will compare the tdPGB2 with two well-known distributions previously considered when studying city size, namely the lognormal (lgn) and the double Pareto lognormal (dPln).[2] The new density function improves on the performance of the distributions used up to now, at least for the US.

The rest of the paper is structured as follows. In Section 2 we will detail the principles underlying our approach. Section 3 defines the densities that are estimated later. Section 4 develops the theoretical model, yielding the new distribution as a result. Section 5 describes the data sets used in the empirical application. Section 6 gives an account of the empirical results. Lastly, we will give some conclusions.

2 The motivation for our approach

This paper is based on the following principles (one in each paragraph), many of them standard in the Urban Economics literature.

From our point of view, to find statistical distributions that fit the data better than the ones known in the literature is an interesting contribution by itself. But it is even more interesting if these new distributions are derived from a theoretical economic model, in which functions that have clear economic meaning are defined. See, in this regard, Section 4.

[2] We do not specifically present a separate study for the Pareto distribution as it is encoded in the tails of the new tdPGB2. Moreover, the new density is better than a single power law for all the range of city sizes. It is also better than the Generalized Beta of the second kind (GB2) distribution.

The study of city size distributions should be, to the extent possible, a long-term analysis (Parr, 1985; Gabaix and Ioannides, 2004). In particular, we have used US data from 1900 to 2010.

It seems that there is no single density function capable of providing an adequate description of the distribution *for all values of the city size population variable*. This is an accepted statement in the literature on income size distribution.[3] We consider this idea to also be applicable to the study of city size distributions. Consequently, in our approach, we have divided the overall distribution into three parts: the lower tail, the body, and the upper tail.

Therefore, large urban nuclei (the upper tail) do matter and require special attention. This is a generally accepted fact in the Urban Economics literature, where the largest cities are often considered to be outliers with respect to the the hypothesized distribution.

In addition, we have the certainty that small nuclei (the lower tail) do matter and also require a specific treatment (something confirmed in our empirical application). This approach is fairly overlooked, with the possible exception of Reed (2002, 2003) and, theoretically, of Blank and Solomon (2000) and Lee and Li (2013). Therefore, we have considered all entities of population, without any truncation point. From an empirical perspective, small urban nuclei are not relevant for the percentage of the population that they represent, but this is not the case with regard to the total number of nuclei.

The parsimony in terms of the number of parameters of the distribution to be estimated is always a goal to be pursued. This is one of the reasons for the success of power laws and Zipf's law. However, the new distribution that we have proposed in this paper seems not to be particularly parsimonious. But we can defend this option based on two arguments. First, the information criteria used in Section 6 in order to discriminate

[3] See Dagum (1979) and the citations included therein.

between the studied distributions, namely the Akaike Information Criterion (AIC) and the Bayesian or Schwarz Information Criterion (BIC), explicitly penalize the number of parameters of an hypothesized distribution. Second, there already exist examples in the literature where a mere increase in the number of parameters of the distribution does not always lead to a better fit in information-theoretic terms.[4]

The results are more valid and powerful as they are robust to different alternatives. In the first place, we have used two definitions of US cities: incorporated places and all places. Second, we will consider a number of different criteria in order to assess the quality of the empirical fits. Indeed, we will use three different statistical tests which are very powerful for the large sample sizes at hand (Razali and Wah, 2011), and for which the non-rejections occur only if the deviations (statistics) are really small. They are the Kolmogorov–Smirnov (KS) test, the Crámer–von Mises (CM) test, and the Anderson–Darling (AD) test. Also, we will use the AIC and BIC information criteria. Third, both in the theoretical model and in the empirical analysis, we have divided the support of the distribution into three parts: the lower tail, the body and the upper tail. There are two main reasons for the three-parts option. On the one hand, the literature supports this alternative; on the other hand, we have empirically explored other possibilities and the best results are obtained with the option reported here.

3 Description of the distributions used

3.1 The lognormal (lgn)

The well-known lognormal distribution for the population of cities was proposed in the field of Urban Economics by Parr and Suzuki (1973) and afterwards by Eeckhout

[4] See, for example, the case of Switzerland in Giesen et al. (2010), where the lognormal (two parameters) outperforms the double Pareto lognormal (four parameters), and other examples in González-Val et al. (2015) where the log-logistic (two parameters) also outperforms the dPln.

(2004) when considering *all* cities. The corresponding density is simply

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

where $\mu, \sigma > 0$ are respectively the mean and the standard deviation of $\ln x$, and x is the population of the urban units under study.

3.2 The double Pareto lognormal (dPln)

The second distribution in our study will be the double Pareto lognormal distribution, introduced by (Reed, 2002, 2003; Reed and Jorgensen, 2004):

$$\begin{aligned}
&f_{\mathrm{dPln}}(x,\alpha,\beta,\mu,\sigma)\\
&\quad=\frac{\alpha\beta}{2x(\alpha+\beta)}\exp\left(\alpha\mu+\frac{\alpha^2\sigma^2}{2}\right)x^{-\alpha}\left(1+\mathrm{erf}\left(\frac{\ln x-\mu-\alpha\sigma^2}{\sqrt{2}\sigma}\right)\right)\\
&\quad-\frac{\alpha\beta}{2x(\alpha+\beta)}\exp\left(-\beta\mu+\frac{\beta^2\sigma^2}{2}\right)x^{\beta}\left(\mathrm{erf}\left(\frac{\ln x-\mu+\beta\sigma^2}{\sqrt{2}\sigma}\right)-1\right)
\end{aligned}$$

where erf is the error function associated to the normal distribution and $\alpha, \beta, \mu, \sigma > 0$ are the four parameters of the distribution. It has the property that it approximates different power laws in each of its two tails: $f_{\mathrm{dPln}}(x) \approx x^{-\alpha-1}$ when $x \to \infty$ and $f_{\mathrm{dPln}}(x) \approx x^{\beta-1}$ when $x \to 0$, hence the name "double Pareto". The body is approximately lognormal, although it is not possible to exactly delineate the switch between the lognormal and the Pareto behaviour. Several references show that the dPln offers a good fit for different countries (Giesen et al., 2010; Giesen and Suedekum, 2014; González-Val et al., 2015).

3.3 The threshold double Pareto Generalized Beta of the second kind (tdPGB2)

We introduce here a new distribution. By construction, the tdPGB2 has a Generalized Beta of the second kind (GB2) body and Pareto tails, the three regions exactly delineated by two thresholds: $\epsilon > 0$ separates the Pareto lower tail from the GB2 body, and $\tau > \epsilon$ separates the body from the Pareto upper tail.

The specific description is as follows. We first define the building block functions, setting

$$f_{\mathrm{GB2}}(x, a, b, p, q) = \frac{a x^{ap-1}}{b^{ap} B(p, q) \left(1 + (x/b)^a\right)^{p+q}} \tag{1}$$

$$\mathrm{cdf}_{\mathrm{GB2}}(x, a, b, p, q) = \frac{1}{B(p, q)} B\left(\frac{(x/b)^a}{1 + (x/b)^a}, p, q\right) \tag{2}$$

$$l(x, \rho) = x^{\rho - 1} \tag{3}$$

$$u(x, \zeta) = \frac{1}{x^{1+\zeta}} \tag{4}$$

The f_{GB2} ($\mathrm{cdf}_{\mathrm{GB2}}$) is the Generalized Beta of the second kind density (resp., cumulative distribution function, cdf) (McDonald and Xu, 1995; Kleiber and Kotz, 2003), $B(z, p, q) = \int_0^z t^{p-1}(1 - t)^{q-1}\, dt$ with $z \in [0, 1]$ is the incomplete Beta function and $B(p, q) = B(1, p, q)$ is the Beta function. The four parameters a, b, p, q are positive, b is a scale parameter, and a, p, q are shape parameters. The functions $l(x, \rho)$ and $u(x, \zeta)$ will model, except for a multiplicative positive constant, the Pareto lower (l) and upper (u) tails of our distribution, where $\rho > 0$ and $\zeta > 0$ are the respective Pareto exponents.

We have imposed the continuity of the composite density function at the two threshold points and an overall normalization of the former to unity. The resulting density is

$$f_{\mathrm{tdPGB2}}(x,\rho,\epsilon,a,b,p,q,\tau,\zeta)=\begin{cases} b_3\, e_3\, l(x,\rho) & 0<x\le\epsilon \\ b_3\, f_{\mathrm{GB2}}(x,a,b,p,q) & \epsilon\le x\le\tau \\ b_3\, a_3\, u(x,\zeta) & \tau\le x \end{cases}$$

where the constants are given by

$$e_3 = \frac{f_{\mathrm{GB2}}(\epsilon,a,b,p,q)}{l(\epsilon,\rho)} \tag{5}$$

$$a_3 = \frac{f_{\mathrm{GB2}}(\tau,a,b,p,q)}{u(\tau,\zeta)} \tag{6}$$

$$b_3^{-1} = e_3\frac{\epsilon^\rho}{\rho} + \mathrm{cdf}_{\mathrm{GB2}}(\tau,a,b,p,q) - \mathrm{cdf}_{\mathrm{GB2}}(\epsilon,a,b,p,q) + \frac{a_3}{\zeta\,\tau^\zeta} \tag{7}$$

This distribution depends on eight parameters $(\rho,\epsilon,a,b,p,q,\tau,\zeta)$ to be estimated.

4 The theoretical model generating the new distribution

The most common functions used to describe city size distributions all have an underlying theoretical model from which they are derived. Thus, Gabaix (1999) and Córdoba (2008) deduce power laws and, more specifically, Zipf's law. The same law, although in a very different setting, is also obtained by Hsu (2012), while Eeckhout (2004) proposes a model for the lognormal. The more recent double Pareto lognormal comes from the theoretical models proposed by Reed (2002), Reed and Jorgensen (2004) and Giesen and Suedekum (2014).

Our model is not of a statistical nature as are those just mentioned and, to some extent, the productivity random shocks model of Eeckhout (2004) and, especially, the random growth Gibrat model of Gabaix (1999).

In Parker (1999), within a neoclassical labour market model where firms maximize profits, the GB2 density is exactly deduced. What is interesting about that model is that it allows, *mutatis mutandis*, being applied to the case of urban nuclei to get our new distribution of Section 3: the tdPGB2.

We have separated the study of city size distributions according to three different regions: the lower tail ($x \in (0, \epsilon]$), the body ($x \in [\epsilon, \tau]$), and the upper tail ($x \in [\tau, \infty)$). The quantities ϵ and τ are the thresholds. Clearly, $\epsilon < \tau$.

We will denote the number of cities within the three intervals of population values as $n_i(x)$, $i = 1, 2, 3$, respectively. The corresponding cumulative numbers of cities are $N_i(x),\ i = 1, 2, 3$. The total number of cities, $N_3(\infty)$, is obviously a constant and it is assumed to have a finite upper bound Θ, so that $N_3(\infty) < \Theta$. If the total number of cities is finite, the total population to be allocated will be finite as well.

If we want to obtain an overall continuous probability density function we have to:

i) Assume, as is usual in the field, x to be a continuous variable, and obtain the continuity of $n_i(x)$, $i = 1, 2, 3$ on the respective intervals where they are defined.

ii) Impose the continuity of the previous functions at the threshold points, namely

$$n_1(\epsilon) = n_2(\epsilon)\,, \quad n_2(\tau) = n_3(\tau) \tag{8}$$

iii) Divide the number of cities $n_i(x),\ i = 1, 2, 3$ by the *total number of cities* $N_3(\infty)$, so that $n_i(x)/N_3(\infty),\ i = 1, 2, 3$ give the correct *densities* of cities of population x on the respective intervals and also at the threshold points ϵ and τ.

At the end, this process will lead to the generation of the previously defined tdPGB2.

We will develop the model considering the three regions separately; then we will consider the joint results.

4.1 Model for the lower tail (variables and parameters with index 1) of the tdPGB2

The model consists of maximizing the net output function in monetary units of the whole urban system of a country at a given time.

The human capital level of each city depends on the population x according to the function $\psi_1(x)$, with $\psi_1(0) = 0$. We assume it to be positive and increasing. Each inhabitant supplies one unit of labour inelastically. The gross output of the cities of population x is $F_1[n_1(x), \psi_1(x)]$. There are diminishing returns to the number of cities, i.e.,

$$\frac{\partial F_1}{\partial n_1} > 0\,, \quad \frac{\partial^2 F_1}{\partial n_1^2} < 0 \tag{9}$$

at all population levels. There are also monetary congestion costs $c_1(x)$ associated to a city of population x. These costs reduce the gross output of each urban settlement. We assume that $c_1(0) = 0$, $c_1(x) > 0$ and $c_1'(x) > 0$.

Thus, the net output of the cities of population $x \in (0, \epsilon]$ is $F_1[n_1(x), \psi_1(x)] - c_1(x)n_1(x)$, and the net output of all cities with populations between 0 and ϵ (the lower tail) is the corresponding definite integral of this last quantity. To specify the problem more, we assume further that $F_1[n_1(x), \psi_1(x)] = \psi_1(x)n_1(x)^\beta$, where $\beta \in (0, 1)$ in order to arrange that the signs of the derivatives behave as stated in (9).

Therefore, the cities' optimal control problem for the lower tail, where the output price has been normalized to unity, can be stated as

$$\begin{aligned}
&\max_{n_1} && \int_0^\epsilon (\psi_1(x)n_1(x)^\beta - c_1(x)n_1(x))\,dx \\
&\text{subject to :} && \frac{dN_1(x)}{dx} = n_1(x) \\
& && N_1(0) = 0 \\
& && N_1(\epsilon) = \int_0^\epsilon n_1(x)\,dx < \Theta \\
& && n_1(x) \in (0, \infty)
\end{aligned}$$

where the state variable is $N_1(x)$ and the control is $n_1(x)$. The associated Hamiltonian function is simply

$$H_1(x, N_1, n_1, \lambda_1) = \psi_1(x) n_1(x)^{\beta} - c_1(x) n_1(x) - \lambda_1(x) n_1(x) \tag{10}$$

The state and costate equations are the following[5]

$$\begin{aligned} \frac{dN_1(x)}{dx} &= -\frac{\partial H_1}{\partial \lambda_1} = n_1(x) \\ \frac{d\lambda_1(x)}{dx} &= \frac{\partial H_1}{\partial N_1} = 0 \end{aligned}$$

and thus $\lambda_1(x) = \lambda_1 = \mathrm{Constant}$. The control $n_1(x)$ to be chosen is the one which maximizes the Hamiltonian and belongs to an open interval, so no corner solutions may arise. The first order condition is just

$$\frac{\partial H_1}{\partial n_1} = \psi_1(x) \beta n_1(x)^{\beta-1} - c_1(x) - \lambda_1 = 0 \tag{11}$$

The second order derivative is

$$\frac{\partial^2 H_1}{\partial n_1^2} = \psi_1(x) \beta(\beta - 1) n_1(x)^{\beta-2} < 0\,, \quad x \in (0, \epsilon]$$

and therefore the first order condition becomes necessary and sufficient for a strict global maximum. From equation (11) we can solve for $n_1(x)$ as follows

$$n_1(x) = \left(\frac{\beta \psi_1(x)}{c_1(x) + \lambda_1} \right)^{1/(1-\beta)} \quad , \quad x \in (0, \epsilon]$$

It is time now to define specific functional forms for the human capital and cost functions: $\psi_1(x) = A_1 x^{\gamma_1}$, $c_1(x) = k_1 x^{b_1}$, where $A_1 > 0$, $\gamma_1 > 0$, $k_1 > 0$ and $b_1 > 0$. The microfoundations for these specific functional forms are as follows. We want both

[5]It is not necessary to impose the transversality conditions because $N_1(\epsilon) < \Theta$.

$\psi_1(x)$ and $c_1(x)$ to be positive and increasing functions (larger cities are associated to higher levels of human capital and higher congestion costs) and $\psi_1(0) = c_1(0) = 0$ (obviously, cities with no people do not have either human capital or costs). Whether concave ($\gamma_1, b_1 < 1$), linear ($\gamma_1 = b_1 = 1$) or convex ($\gamma_1, b_1 > 1$) will be discussed in subsection 4.5.

Consequently, we have

$$n_1(x) = \left(\frac{\beta\, A_1\, x^{\gamma_1}}{k_1 x^{b_1} + \lambda_1} \right)^{1/(1-\beta)}, \quad x \in (0, \epsilon]$$

We want $n_1(x)$ to be a pure Pareto power law, that is, to be proportional to a power function. For this, it is necessary and sufficient that $\lambda_1 = 0$.[6] Then, with $\lambda_1 = 0$ we simply have $n_1(x) = \left(\frac{\beta\, A_1}{k_1} \right)^{1/(1-\beta)} x^{\frac{\gamma_1 - b_1}{1-\beta}}$ so in order to have a pure Pareto lower tail we require that the corresponding Pareto exponent ρ satisfies $\rho = \frac{\gamma_1 - b_1}{1-\beta} + 1 > 0$. The assumptions made so far about the values of the parameters β, b_1 and γ_1 are compatible with the validity of this equation and the empirical analysis confirms that the estimations of ρ are always positive.

4.2 Model for the body (variables and parameters with index 2) of the tdPGB2

In the body of the tdPGB2 distribution, we have assumed a similar model as for the lower tail on the corresponding interval $[\epsilon, \tau]$. Therefore, the number of cities in the body can be found to be

$$n_2(x) \quad = \quad \left(\frac{\beta\, A_2 x^{\gamma_2}}{k_2 x^{b_2} + \lambda_2} \right)^{1/(1-\beta)}, \quad x \in [\epsilon, \tau]$$

[6]As already indicated, the proper probability density function on the interval $x \in (0, \epsilon]$ is $n_1(x)/N_3(\infty)$ (see also Parker (1999)). Since $N_3(\infty)$ is a finite positive constant, we have a Pareto distribution in the lower tail if and only if, as stated in the text, $\lambda_1 = 0$. This footnote also applies to the power law in the upper tail with the corresponding Lagrange multiplier, see subsection 4.3.

where now we expect to have $\lambda_2 > 0$. Comparing this last expression with the definition of the GB2 distribution (see Eq. (1)) both functions can be properly related, so that $n_2(x)/N_3(\infty)$ is, up to a positive multiplicative constant, the expression of $f_{\mathrm{GB2}}(x,a,b,p,q)$ of (1). Indeed, we simply have

$$n_2(x) = \left(\frac{\beta A_2}{\lambda_2}\right)^{1/(1-\beta)} B(p,q)\frac{b^{ap}}{a} f_{\mathrm{GB2}}(x,a,b,p,q)\,, \quad x \in [\epsilon,\tau]$$

with the identifications of the parameters

$$\begin{aligned}
a &= b_2 \\
b &= \left(\frac{\lambda_2}{k_2}\right)^{1/b_2} \\
p &= \frac{1}{b_2}\left(1+\frac{\gamma_2}{1-\beta}\right) \\
q &= \frac{1}{1-\beta} - \frac{1}{b_2}\left(1+\frac{\gamma_2}{1-\beta}\right)
\end{aligned}$$

4.3 Model for the upper tail (variables and parameters with index 3) of the tdPGB2

The corresponding number of cities in the upper tail can be found to be

$$n_3(x) = \left(\frac{\beta\, A_3 x^{\gamma_3}}{k_3 x^{b_3} + \lambda_3}\right)^{1/(1-\beta)} , \quad x \in [\tau,\infty)$$

In this case, we want to obtain again a pure Pareto upper tail. Thus, we require that $\lambda_3 = 0$. Then, we have $n_3(x) = \left(\frac{\beta\, A_3}{k_3}\right)^{1/(1-\beta)} x^{\frac{\gamma_3 - b_3}{1-\beta}}$ and the Pareto exponent ζ must satisfy $-\zeta = \frac{\gamma_3 - b_3}{1-\beta} + 1 < 0$. The assumptions made so far about the values of the parameters β, b_3 and γ_3 are compatible with the validity of this equation and the empirical analysis confirms that the estimations of ζ are always positive.

4.4 The overall distribution

Lastly, we can impose the following natural conditions, namely, continuity of human capital ($\psi(x)$) and *effective* cost functions ($c(x) + \lambda$)[7] at the threshold values ϵ and τ:

$$A_1\epsilon^{\gamma_1} = A_2\epsilon^{\gamma_2}\,, \quad A_2\tau^{\gamma_2} = A_3\tau^{\gamma_3} \tag{12}$$

$$k_1\epsilon^{b_1} = k_2\epsilon^{b_2} + \lambda_2\,, \quad k_2\tau^{b_2} + \lambda_2 = k_3\tau^{b_3} \tag{13}$$

These conditions have (8) as an immediate consequence. Now, as stated previously, dividing $n_i(x)$, $i = 1, 2, 3$ by the total number of cities $N_3(\infty)$ provides the exact probability density function on each interval and the threshold values corresponding to the tdPGB2. Let us remark that in subsection 3.3, the definition of the quantities e_3, a_3, b_3 by Eqs. (5), (6) and (7) reflects exactly the conditions for the overall probability density function to be continuous at the threshold values ϵ and τ and to be normalized to unity.

As a final outcome, we have demonstrated that we can obtain the tdPGB2 probability distribution from a theoretical economic model.

4.5 Economic explanations for the shape of the functions used in the model

There are two functions that define the most important characteristics of our model: $\psi(x)$ and $c(x)$. With regard to the first, the proposed functional form and values of the γ parameters in the previous subsections make it an increasing function ($\gamma > 0$). It can be convex ($\gamma > 1$), linear ($\gamma = 1$), or concave ($\gamma < 1$). In the first (third) case, if a city has a size that is, for example, twice the size of another city, its human capital stock will be larger (smaller) than twice that of the smaller city. Our theoretical model is compatible with these three options and, therefore, it is an empirical question, outside

[7] See Eq. (10) to notice that the effective cost function derived from the Hamiltonian is just $c(x) + \lambda$.

the scope of this paper, to determine which of them holds.

Of course, from an economic point of view, the most interesting scenario is that in which human capital accumulates at rates that are increasing with respect to the size of the urban settlements. We can justify this behaviour with two arguments.

First, "there is some evidence suggesting that human capital accumulates more quickly in urban areas" (Glaeser and Resseger, 2010). This empirical evidence is also corroborated in Moretti (2004) and Rauch (1993). Second, there is the existence of agglomeration economies. Estimates of their magnitude imply that doubling the size or density of an urban area increases its productivity between 2% to 8% (see the surveys of this topic in Melo et al. (2009), Puga (2010) and Combes and Gobillon (2015)). Human capital is one of the inputs in our production function, and therefore causes the productivity per worker to be higher in larger cities.

It is time now to justify the shape adopted for the cost function $c(x)$, which is increasing with respect to city size ($b > 0$). We can label these costs, without loss of generality, as congestion costs. Now, we are thinking about the factors related to all the "bad" things that are traditionally associated with bigger cities. This is why these costs reduce the gross output of each urban nucleus and affect the net output. Some of these factors are crime, traffic, diseases, pollution, and housing prices. All of them tend to increase in absolute terms with city size. In per capita or relative terms things are not so clear. Bettencourt and West (2010) report that the magnitude of crime, traffic, and certain diseases is multiplied by 2.3 if the population of a city is doubled; in this case $b > 1$. Regarding pollution, there is a certain consensus about the fact that larger cities are, on average, greener (Glaeser, 2011); in this case $b < 1$. Lastly, the connection between city size and housing prices is a complex topic which depends on local geography, regulatory policies, and the internal spatial structure of the city: see Saiz (2010), Glaeser et al. (2012), and the references therein for an overview; in this case $b > 1$ or $b < 1$. Again, as for the $\psi(x)$ function, our theoretical model is compatible with all the possibilities and, therefore, is an open empirical question to determine whether $c(x)$ increases with city size at a more than proportional rate or not.

4.6 A reflection on the practical implications of our model

The parameters of the overall distribution so obtained, the tdPGB2, depend on the elasticities β, γ_i, $i = 1, 2, 3$ and b_i, $i = 1, 2, 3$. In particular, the Pareto exponents at the lower and upper tails are also related to β, γ_i and b_i. These elasticities might vary over time, mainly for economic reasons, so we obtain that the city size distribution is explained at a given time by the economic conditions that determine it. Therefore, this model may help in explaining the observed persistence of the city size distribution in the short term (Kim, 2000; Beeson et al., 2001), because the previously mentioned elasticities probably have a slow time variation. However, in the long term the variations can be quite remarkable, as Batty (2006) points out.

The content of the last paragraph leads us to two important outcomes. First, the urban policy implications of the previous discussion are, in our opinion, important. Second, the interpretation of the city size distribution as a steady state (by definition, with no time changes at all) of a stochastic process is not the only possible approach. In our framework the city size distribution can be interpreted as an equilibrium given the economic conditions at a given time. Our model is static in nature but also explains the evolution of the city size distribution, in the sense that if the elasticities of the model do change with time, the distribution will change as well.

5 The databases

In this article, we have used data about US urban centres from two sources. The first is the decennial data of the US Census Bureau of "incorporated places" without any size restriction from 1900 to 2000. These include governmental units classified under state laws as cities, towns, boroughs, or villages. Alaska, Hawaii and Puerto Rico have not been considered due to data limitations. The data has been collected from the original documents of the annual census published by the US Census Bureau.[8] This data was first introduced in González-Val (2010), see therein for details.

The second source consists of all US urban places, unincorporated and incorporated, and without size restrictions, also provided by the US Census Bureau for the years 2000 and 2010. The data for the year 2000 was first used in Eeckhout (2004) and later in Giesen et al. (2010), Ioannides and Skouras (2013) and Giesen and Suedekum (2014). The two samples were also used in González-Val et al. (2015).

We do not consider, on the other hand, data like Economic Areas, Core Based Statistical Areas (CBSA), or MSAs. These three types violate our principle that the small nuclei do matter and that there should be no truncation point: there are only 366 MSAs, 940 CBSAs, and less than 200 Economic Areas in 2010.

The descriptive statistics of the data sets used in this paper can be seen in Table 1.[9]

6 Results

We will show briefly in this section how our new distribution, the tdPGB2, performs in fitting the size of US places (incorporated and all), compared to well-known distributions of city size as the lognormal (lgn) and the double Pareto lognormal (dPln).

First, Table 2 shows the maximum likelihood estimation (MLE)[10] results for the used distributions. We can observe that the estimations are rather precise in all cases.

We have shown in Table 3 the results of the Kolmogorov–Smirnov (KS), Crámer–

[8] `http://www.census.gov/prod/www/decennial.html` Last accessed: May 6$^{\mathrm{th}}$, 2016.

[9] The results for the remaining years of incorporated places in the period 1900–2000 are similar and are not shown for the sake of brevity. They are avaliable from the authors upon request. The previous statement also applies to all the tables in Section 6.

[10] We have performed all the estimations using the numerical `mle` command of MATLAB®, on an equal footing for all the parameters. The standard errors were computed independently using the software MATHEMATICA®.

von Mises (CM) and Anderson–Darling (AD) tests for the studied samples and density functions used. The AD test is very appropriate when one wants to assess the adequacy of the distributions at the tails, see, e.g., Cirillo (2013). The first remarkable result is that the lognormal (lgn) is always strongly rejected, so this specification seems not to be as good as a parametric description in practice, at least for US places.[11] The second observation is that a similar thing happens for the double Pareto lognormal (dPln): it is rejected almost always, with the only exception being the sample of incorporated places in 1900.

In addition, at the same time and with the same techniques, the proposed tdPGB2 is never rejected by any of the three tests. In this respect, the differences in the statistics of the used tests are relevant when going from the lognormal to the dPln and then to the tdPGB2. This means that the tdPGB2 is a good parametric specification for the size of US places.

Lastly, we have shown in Table 4 the results of the Akaike Information Criterion (AIC) and the Schwarz or Bayesian Information Criterion (BIC), which are standard in the literature, in order to choose between the proposed distributions. We can see that the selected specification is the tdPGB2 by both AIC and BIC criteria.

7 Conclusions

This paper has tried to contribute, both from a theoretical perspective and from an empirical approach, to the literature on city size distributions.

To summarize, the contributions from the theoretical point of view are the following. The main result is that the new statistical distribution introduced in this paper, namely the "threshold double Pareto Generalized Beta of the second kind" (tdPGB2) is deduced using a simple model in which a social planner allocates the population of a

[11] This fact has been previously highlighted by Giesen and Suedekum (2014).

country in cities of different sizes so as to maximize the net output of the whole urban system. There are four basic features of this model. First, that it is built up piecewise, taking into account the specific particularities of the lower tail, the body and the upper tail. Second, the production function is increasing and concave in the number of cities, so that it complies with the law of diminishing returns. Third, the human capital stock of a city is increasing with respect to city size. And fourth, the congestion costs that lessen the gross output of each urban unit are also increasing with respect to cities population.

The theoretical parameters of the overall distribution are given explicitly, at any given time, in terms of the elasticities of the gross output with respect the number of cities, of human capital stock with respect to city size, and of costs with respect to city size. Economic conditions may change and accordingly the associated elasticities, thus determining the resulting city size distribution. This fact opens the door for urban economic policy recipes trying to govern the economic conditions previously mentioned. Therefore, our approach is rooted in economic modelling, rather than in pure statistical reasoning.

Empirically, the data sets we have shown are those of the US incorporated places in 1900, 1950, and 2000. Also, all US places in 2000 and 2010. As mentioned, we have introduced the tdPGB2 distribution. It is pure Pareto at both tails and Generalized Beta of the second kind (GB2) on the body. This new density function outperforms the most widely used ones in the literature, namely, the Pareto, the lognormal and the double Pareto lognormal (dPln). In fact, the tdPGB2 is the distribution chosen to describe US places, incorporated and all. These results are robust to a battery of different independent criteria: Kolmogorov–Smirnov, Crámer–von Mises, Anderson–Darling tests; Akaike Information Criterion and Bayesian Information Criterion.

From an empirical point of view, the main contributions of this paper are the following:

i) A classical distribution, with underlying theoretical model (Eeckhout, 2004), the lognormal, is surpassed by the tdPGB2.

ii) A newer distribution, the dPln, also with underlying theoretical foundations (Giesen and Suedekum, 2014), is outperformed by the new tdPGB2.

iii) The new distribution confirms something that has been known for a long time: that the upper tail can be taken as pure Pareto. Moreover, also the lower tail can be taken as pure Pareto.

iv) The Generalized Beta of the second kind distribution improves the performance compared to the lognormal for the body, and this distinction has an economic theoretical origin.

The empirical results are in good agreement with the theoretical model developed, based on economic foundations. Both theory and empirical support may lead to a new way of looking at city size distributions.

References

Batty, M. (2006). Rank clocks. *Nature*, 444(7119):592–596.

Batty, M. (2013). *The New Science of Cities*. The MIT Press.

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

Beeson, P. E., DeJong, D. N., and Troesken, W. (2001). Population growth in US counties, 1840–1990. *Regional Science and Urban Economics*, 31(6):669–699.

Berliant, M. and Watanabe, H. (2015). Explaining the size distribution of cities: Extreme economies. *Quantitative Economics*, 6:153–187.

Berry, B. J. L. and Okulicz-Kozaryn, A. (2012). The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, 29:517–523.

Bettencourt, L. and West, G. (2010). A unified theory of urban living. *Nature*, 4667:912–913.

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

Blank, A. and Solomon, S. (2000). Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A*, 289:279–288.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2008). A century of shocks: The evolution of the German city size distribution 1925–1999. *Regional Science and Urban Economics*, 38(4):330–347.

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A*, 392:5947–5962.

Combes, P. P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics, forthcoming*, volume 5. Elsevier.

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

Dagum, C. (1979). Generating systems and properties of income distribution models. In *Conference of Canadian Incomes, May 10–12*.

Desmet, K. and Rappaport, J. (2015). The settlement of the United States, 1800–2000: The long transition towards Gibrat's Law. *Journal of Urban Economics. DOI 10.1016/j.jue.2015.03.004*.

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

Fazio, G. and Modica, M. (2015). Pareto or log-normal? Best fit and truncation in the distribution of all cities. *Journal of Regional Science*, 55(5):736–756.

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

Gabaix, X. and Ioannides, Y. (2004). The evolution of city size distributions. In Henderson, V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 53, pages 2341–2378. Elsevier.

Giesen, K. and Suedekum, J. (2014). City age and city size. *European Economic Review*, 71:193–208.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities–Double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

Glaeser, E. L. (2011). *Triumph of the City*. The Penguin Press.

Glaeser, E. L., Gottlieb, J. D., and Tobio, K. (2012). Housing booms and city centers. *NBER Working Paper No. 17914*.

Glaeser, E. L. and Resseger, M. G. (2010). The complementarity between cities and skills. *Journal of Regional Science*, 50(1):221–244.

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

González-Val, R., Ramos, A., Sanz-Gracia, F., and Vera-Cabello, M. (2015). Size distribution for all cities: Which one is best? *Papers in Regional Science*, 94(1):177–197.

Hsu, W.-T. (2012). Central place theory and city size distribution. *The Economic Journal*, 122:903–932.

Ioannides, Y. M. and Overman, H. G. (2004). Spatial evolution of the US urban system. *Journal of Economic Geography*, 4(2):131–156.

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

Kim, S. (2000). Urban development in the United States, 1690–1990. *Southern Economic Journal*, 66(4):855–880.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in Economics and actuarial sciences*. Wiley-Interscience.

Lee, S. and Li, Q. (2013). Uneven landscapes and city size distributions. *Journal of Urban Economics*, 78:19–29.

Luckstead, J. and Devadoss, S. (2014). Do the world's largest cities follow Zipf's and Gibrat's laws? *Economics Letters*, 125:182–186.

McDonald, J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152.

Melo, P., Graham, D., and Noland, R. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342.

Moretti, E. (2004). Human capital externalities in cities. In Henderson, V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 51, pages 2243–2291. Elsevier.

Overman, H. G. and Ioannides, Y. M. (2001). Cross-Sectional evolution of the US city size distribution. *Journal of Urban Economics*, 49(3):543–566.

Parker, S. C. (1999). The generalised beta as a model for the distribution of earnings. *Economics Letters*, 62:197–200.

Parr, J. B. (1985). A note on the size distribution of cities over time. *Journal of Urban Economics*, 18:199–212.

Parr, J. B. and Suzuki, K. (1973). Settlement populations and the lognormal distribution. *Urban Studies*, 10(3):335–352.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics*, 34(3):380–400.

Razali, N. M. and Wah, Y. B. (2011). Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33.

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.

Reed, W. J. (2003). The Pareto law of incomes–An explanation and an extension. *Physica A*, 319:469–486.

Reed, W. J. and Jorgensen, M. (2004). The double Pareto-lognormal distribution–A new parametric model for size distributions. *Communications in Statistics–Theory and Methods*, 33(8):1733–1753.

Resende, M. (2004). Gibrat's Law and the Growth of Cities in Brazil: A Panel Data Investigation. *Urban Studies*, 41(8):1537–1549.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3):1253–1296.

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

Xu, Z. and Zhu, N. (2009). City size distribution in China: Are large cities dominant? *Urban Studies*, 46(10):2159–2185.

Table 1: Descriptive statistics of the US data samples used

Sample	Obs.	% of US pop.	Mean	SD	Mean (log scale)	SD (log scale)	Min.	Max.
Inc. places 1900	10,596	46.99	3,376	42,324	6.65	1.26	7	3,437,202
Inc. places 1950	17,113	63.48	5,613	76,064	6.84	1.50	1	7,891,957
Inc. places 2000	19,296	61.49	8,968	78,015	7.18	1.78	1	8,008,278
All places 2000	25,358	73.98	8,232	68,390	7.28	1.75	1	8,008,278
All places 2010	29,461	74.31	7,826	65,494	7.11	1.82	1	8,175,133

Table 2: ML estimators and standard errors (SE) of the parameters of the dPln and tdPGB2 for the US places samples. The estimators for the lognormal are the mean and the standard deviation of the logarithm of population data, see Table 1

	dPln			
	α (SE)	β (SE)	μ (SE)	σ (SE)
Inc. places 1900	0.92 (0.01)	2.64 (0.06)	5.95 (0.01)	0.58 (0.01)
Inc. places 1950	0.80 (0.01)	2.15 (0.04)	6.06 (0.01)	0.78 (0.01)
Inc. places 2000	0.87 (0.01)	3.62 (0.09)	6.31 (0.01)	1.36 (0.01)
All places 2000	1.22 (0.01)	3.15 (0.08)	6.78 (0.01)	1.52 (0.01)
All places 2010	1.12 (0.01)	3.03 (0.07)	6.54 (0.01)	1.55 (0.01)
	tdPGB2			
	ρ (SE)	ϵ (SE)	a (SE)	b (SE)
Inc. places 1900	1.88 (0.17)	58 (7)	2.350 (0.030)	242 (3)
Inc. places 1950	1.43 (0.07)	62 (2)	1.063 (0.005)	49 (1)
Inc. places 2000	1.51 (0.07)	43 (4)	0.699 (0.003)	36 (1)
All places 2000	1.52 (0.03)	102 (4)	0.473 (0.002)	223 (3)
All places 2010	1.32 (0.02)	132 (3)	0.497 (0.003)	318 (5)
	tdPGB2			
	p (SE)	q (SE)	τ (SE)	ζ (SE)
Inc. places 1900	1.271 (0.018)	0.305 (0.005)	2,472 (227)	1.001 (0.018)
Inc. places 1950	7.882 (0.068)	0.679 (0.006)	15,968 (1,164)	1.082 (0.028)
Inc. places 2000	7.201 (0.055)	0.895 (0.007)	55,595 (2,537)	1.380 (0.043)
All places 2000	4.592 (0.027)	2.182 (0.011)	54,016 (3,094)	1.445 (0.042)
All places 2010	3.068 (0.022)	1.878 (0.010)	56,703 (3,238)	1.430 (0.040)

Table 3: p-values (statistics) of the Kolmogorov–Smirnov (KS), Cramér–Von Mises (CM) and Anderson–Darling (AD) tests for US places and the density functions used. Non-rejections at the 5% significance level are in bold

	lgn			dPln		
	KS	CM	AD	KS	CM	AD
Inc. places 1900	0 (0.07)	0 (17.22)	0 (100.47)	**0.17 (0.01)**	**0.11 (0.34)**	**0.10 (1.97)**
Inc. places 1950	0 (0.06)	0 (17.56)	0 (104.90)	0 (0.02)	0 (1.40)	0 (10.48)
Inc. places 2000	0 (0.04)	0 (9.40)	0 (53.66)	0 (0.02)	0 (1.95)	0 (12.63)
All places 2000	0 (0.02)	0 (3.03)	0 (19.12)	0 (0.02)	0 (1.45)	0 (8.98)
All places 2010	0 (0.02)	0 (4.57)	0 (29.48)	0 (0.02)	0 (1.73)	0 (11.73)
	tdPGB2					
	KS	CM	AD			
Inc. places 1900	**0.978 (0.005)**	**0.971 (0.032)**	**0.989 (0.205)**			
Inc. places 1950	**0.994 (0.003)**	**0.990 (0.025)**	**0.989 (0.206)**			
Inc. places 2000	**0.986 (0.004)**	**0.974 (0.031)**	**0.986 (0.214)**			
All places 2000	**0.969 (0.003)**	**0.936 (0.039)**	**0.971 (0.249)**			
All places 2010	**0.899 (0.004)**	**0.814 (0.060)**	**0.769 (0.478)**			

Table 4: Maximum log-likelihoods, AIC and BIC for US places and the density functions used. The lowest values of AIC and BIC for each sample are in bold

	lgn			dPln		
	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
Inc. places 1900	-87,943	175,891	175,905	-87,254	174,516	174,545
Inc. places 1950	-148,254	296,512	296,528	-147,593	295,194	295,225
Inc. places 2000	-177,127	354,258	354,274	-176,931	353,870	353,901
All places 2000	-234,773	469,550	469,566	-234,710	469,428	469,461
All places 2010	-268,748	537,499	537,516	-268,657	537,323	537,356
	tdPGB2					
	log-likelihood	AIC	BIC			
Inc. places 1900	-87,230	**174,476**	**174,535**			
Inc. places 1950	-147,471	**294,958**	**295,020**			
Inc. places 2000	-176,770	**353,556**	**353,619**			
All places 2000	-234,628	**469,272**	**469,337**			
All places 2010	-268,520	**537,056**	**537,122**			