CESifo Working Papers

www.cesifo.org/wp

The Size Distribution across all "Cities": A Unifying Approach

Kristian Giesen

Jens Suedekum

CESifo Working Paper No. 3730

Category 8: Trade Policy

February 2012

CESifo *Working Paper No. 3730*

The Size Distribution across all "Cities": A Unifying Approach

Abstract

In this paper we show that the double Pareto lognormal (DPLN) parameterization provides an excellent fit to the overall US city size distribution, regardless of whether "cities" are administratively defined Census places or economically defined area clusters. We then consider an economic model that combines scale-independent urban growth (Gibrat's law) with endogenous city creation. City sizes converge to a DPLN distribution in this model, which is much better in line with the data than previous urban growth frameworks that predict a lognormal or a Pareto city size distribution (Zipf's law).

JEL-Code: R110, R120, O400.

Keywords: Zipf's law, Gibrat's law, city size distributions, double Pareto-Lognormal.

Kristian Giesen
Mercator School of Management
University of Duisburg-Essen
Lotharstraße 65
Germany - 47057 Duisburg
kristian.giesen@uni-due.de

Jens Suedekum
Mercator School of Management
University of Duisburg-Essen
Lotharstraße 65
Germany - 47057 Duisburg
jens.suedekum@uni-due.de

February 6, 2012

1. Introduction

What is the most appropriate parameterization for the distribution of city sizes within a country? This question, which has important bearing for theories of urban growth, has attracted vast attention ever since the seminal works by Auerbach (1913) and Zipf (1949).

Recently, two major contributions have been published which use different approaches of how a "city" is defined to begin with. These papers come to divergent conclusions about which distribution provides an appropriate approximation of the data. Using administrative US "Census places", Eeckhout (2004) shows that the lognormal (LN) distribution closely fits the population sizes of those cities. Rozenfeld *et al.* (2011), in contrast, use a *bottom up* approach of constructing cities as "area clusters" from high resolution data on population density in the US. They find that the LN poorly fits the size distribution across those cities. At the same time they find that the sizes of clusters with at least 12,000 inhabitants closely follow a Pareto distribution with tail exponent equal to minus one. That is, Rozenfeld *et al.* (2011) emphasize that the famous *Zipf's law* holds in the upper tail, but they do not suggest a parameterization for the *overall* size distribution across *all* clusters. [1]

Our aim in this paper is to provide a unifying view. We suggest a parameterization, the so-called *double Pareto lognormal* (DPLN) distribution, that provides an excellent data fit to the overall size distribution for both definitions of "cities". The DPLN, which was initially developed by Reed (2002), is characterized by a lognormal body and Paretian power laws in the tails so that it can encapsulate Zipf's law as an upper tail feature. Our findings can thus potentially resolve several issues regarding city size distributions. First, our results are consistent with, but go beyond those of Rozenfeld *et al.* (2011): the DPLN is a parameterization for the size distribution across all area clusters that is fully consistent with the validity of Zipf's law among the large clusters. Second, this paper may reconcile a recent debate related to Eeckhout's (2004, 2009) claim that the sizes of Census places follow a LN distribution.[2] Finally, our results suggest that the precise definition of a "city" does not crucially affect the fundamentals of the size distribution, because the data can be robustly approximated by the same functional form.

We then turn to the foundations of the DPLN and clarify the implications of our empirical findings for theories of urban growth. As is well known from Gabaix (1999) and Eeckhout (2004), both Zipf's law and the LN have their origin in Gibrat's law, which is a growth process where

[1] The traditional literature on city sizes has consistently found support for Zipf's law in most countries and time periods, see Soo (2005) or Nitsch (2004). That literature was, however, constrained to using truncated data sets on city sizes above a certain threshold (e.g., *metropolitan statistical areas* in the US). Untruncated settlement size data became available only recently, which then facilitated the focus on the features of the *overall* city size distribution.

[2] Some authors (e.g., Levy 2009; Ioannides and Skouras, 2009; Malevergne *et al.*, 2011) have argued that the large Census places actually tend to follow a Zipfian power law pattern that is only imperfectly captured by the LN parameterization, even though the LN fits well outside the upper tail. The features of the DPLN are precisely in line with that evidence. The debate between Levy (2009) and Eeckhout (2009) may thus be settled by the observation that the sizes of Census places are best approximated by a DPLN rather than a LN distribution. Indeed, Giesen *et al.* (2010) have shown that the DPLN performs significantly better than the LN in fitting the size distribution across Census places, even after being penalized for having four instead of two parameters. The additional empirical contribution of this paper is to show that the superior data fit of the DPLN remains robust when defining cities according to the recently developed area clusters data, where the LN actually delivers a poor fit.

cities grow randomly irrespective of their current size. In Eeckhout (2004) there is a fixed number of equally old cities that are hit by random productivity shocks. The "pure" form of Gibrat's law holds in his model, and city sizes then converge to a LN distribution. Zipf's law emerges instead when an "impurity" is added to Gibrat's law, and cities are prevented from becoming too small by an arbitrary lower bound (Gabaix, 1999).[3] Evidence shows, however, that neither of these models is fully in line with the data, since neither the LN nor Zipf's law provide a robust and close approximation of the overall US city size distribution.

In this paper, we therefore consider an economic model of an urban system that is consistent with the US evidence. The model extends the framework by Eeckhout (2004) and allows for endogenous city creation, and thus for age heterogeneity across cities. We show that the optimal *number* of cities grows at a constant rate in this model, namely the country's population growth rate, which in turn implies an exponential age distribution across cities. This together with pure scale-independent urban growth resulting from random productivity shocks for existing cities leads to DPLN distributed city sizes, which is what we observe in the data.[4] Our main conclusion is, hence, that the overall US city size distribution can be robustly matched by an urban growth framework that combines Gibrat's law with the realistic feature of a growing number of cities.

2. Data and parameterization

We utilize two different definitions of US "cities" in this paper: Census places and area clusters. The former dataset refers to the year 2000 and includes administratively defined settlements according to legal boundaries ("Census designated places"). It contains 25,359 cities covering 74% of the total US population. The sizes of the places range from one to about 8 million inhabitants (New York City).[5] Comparable data sets on the sizes of administratively defined settlements (not subject to a threshold size) are by now available for many countries. This is a clear advantage. However, a disadvantage is that the boundaries between those units are sometimes defined quite arbitrarily. The second dataset has been constructed by (and is explained in detail in) Rozenfeld *et al.* (2011). Here, cities are defined as area clusters by using an algorithm on high resolution data on population densities in the US. We use their benchmark clusters with ℓ=3 km, which leads to 23,499 cities covering about 96% of the US population in 2001 and range from one to about 16 million inhabitants (the New York cluster). The advantage of this data is that cities are defined as genuine agglomerations ignoring administrative boundaries. A current disadvantage is that such data is not (yet) available for many countries.

Figure 1 shows kernel density estimates (in logarithmic scale) of the city size distribution for both definitions. The area clusters data is depicted by the solid, and the Census places by the dashed black line. As can be seen, the mean size of area clusters is higher than for the Census places, while the the variance is lower. Furthermore, the size distribution across area clusters

[3]This growth process considered in Gabaix (1999) also leads to Zipf's law when the number of cities is growing over time, as long as the city creation rate is not too large.

[4]We also consider an alternative assumption on the dynamics of city creation, but we find that the resulting city size distribution does not come closer to reality than the DPLN (though it still outperforms the LN).

[5]More details can be found in the Geographic Areas Reference Manual available online under http://www.census.gov/geo/www/garm.html.

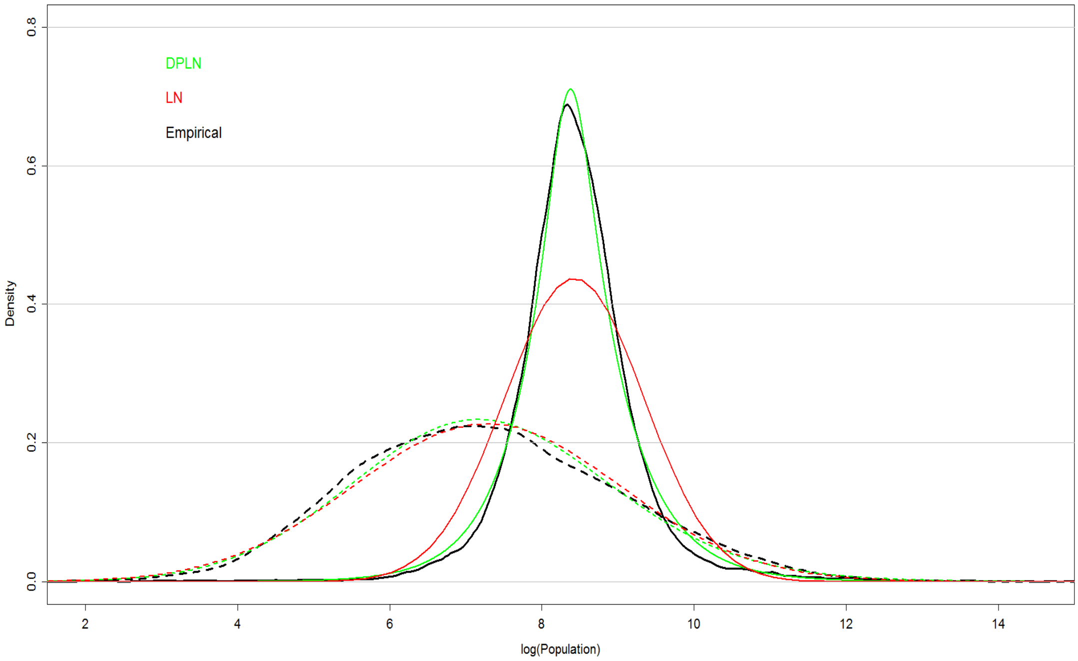

Figure 1: Kernel density estimates of US city sizes along with fitted distributions

has a thicker upper tail and a thinner lower tail than the size distribution across places. These distributional features result from the fact that the clustering algorithm by Rozenfeld *et al* (2011) tends to connect adjacent places into one agglomeration, i.e., the same area cluster.

The next step is to parameterize these city size distributions. We firstly fit the LN distribution to the data by using maximum likelihood estimation (see Table 1 for the results). For the Census places, Figure 1 corroborates Eeckhout's finding: the LN indeed provides a good fit (see the red dashed line). However, when using the area clusters (solid red line), the LN plainly fails to match the data. Turning to Zipf's law, it can be easily verified that it closely fits the data when focusing only on large cities (in either definition).[6] However, as is clear from Figure 1, a Pareto does not hold outside the upper tail and, hence, it is not a useful parameterization for the overall city size distribution. Our suggested functional form for the overall city size distribution is the DPLN distribution which has the following density function for city sizes S:

$$f(S) = \frac{\alpha\beta}{\alpha+\beta}\left[S^{\beta-1}e^{\left(\beta\mu+\frac{\beta^2\sigma^2}{2}\right)}\Phi^c\left(\frac{\ln(S)-\mu+\beta\sigma^2}{\sigma}\right)+S^{-\alpha-1}e^{\left(\alpha\mu+\frac{\alpha^2\sigma^2}{2}\right)}\Phi\left(\frac{\ln(S)-\mu-\alpha\sigma^2}{\sigma}\right)\right]. \quad (1)$$

Here, Φ is the cumulative and Φ^c the complementary-cumulative standard normal distribution. The genesis of the DPLN is discussed in detail in the next section. For the moment, it suffices to note some basic properties. It is a four-parameter distribution (α, β, μ and σ) featuring a lognormal shape in the body and power laws in the tails. More specifically, if $S \to \infty$ then $f(S) \sim S^{-\alpha-1}$, and

[6]We have verified the result by Rozenfeld *et al.* (2011). Using only area clusters that are larger than 12,000 inhabitants, a standard rank-size regression yields a highly significant tail exponent of 0.994 with a R^2 level of 0.99.

if $S \to 0$ then $f(S) \sim S^{\beta-1}$. The parameters α and β are thus the slope parameters of the Pareto tails, while the parameters μ and σ pertain to the location and scale of the LN body. In logarithmic scale, the DPLN can be skewed and its kurtosis can have positive or negative excess, i.e., it can be more peaked (leptokurtic) or more flat (platykurtic) than the LN.

It is straightforward to estimate the parameters of the DPLN as given in (1) by maximum likelihood (see Table 1 for the estimation results).[7] We depict the fitted DPLN distributions in Figure 1. The green solid line is for the area clusters, and the green dashed line is for the Census places. As can be seen, the DPLN provides a very close fit to the empirical size distribution in both cases. Certainly the DPLN does a better job than the LN. For the area clusters this is self-evident by visual inspection. For the places, the performance difference is less pronounced. Still, the DPLN fits clearly better than the LN, even when taking into account that there are two more parameters that need to be estimated.[8] Standard statistical specification tests convey the same message: the LN is rejected much earlier than the DPLN.[9] In sum, we thus conclude that regardless of whether "cities" are defined administratively as Census places or economically as area clusters, the overall size distribution is very closely matched by the DPLN parameterization.[10]

Data and estimated parameters

	Area clusters		Places	
N	23,499		25,359	
coverage	0.96		0.74	
Min	1		1	
Max	15,594,627		8,008,278	
	DPLN	LN	DPLN	LN
α	1.659	-	1.221	-
β	1.830	-	2.821	-
μ	8.370	8.427	6.813	7.277
σ	0.155	0.911	1.514	1.753
AIC	450,996	458,347	469,430	469,550
BIC	451,028	458,363	469,463	469,566
$ln(L^i_j)$	225493.9	229171.3	234711.2	234773.1

[7]We utilize the log-likelihood function and the corresponding starting values as proposed by Reed (2002).

[8]The better adjusted performance can be seen from the Akaike (AIC) and the Bayesian information criterion (BIC), see Table 1. Further model selection tests are presented in Giesen *et al.* (2010), all of which indicate that the DPLN outperforms the LN in fitting the size distribution across US Census places. There we also provide consistent evidence for seven other countries, using comparable data sets on administratively defined settlements.

[9]We have performed Kolmogorov-Smirnov tests by drawing 1000 random samples of size 1000 from both datasets, and for the two hypothesized parameterizations. Using the area cluster (places) data we obtain an average p-value of 0.34 (0.41) for the null that the data follows the DPLN. For the null that the data follows the LN we get a p-value much below 0.001 for both datasets. We hence cannot reject the DPLN, while the LN is strongly rejected.

[10]Rozenfeld *et al.* (2011) also provide settlement size data for similarly defined area clusters in Great Britain. We have worked with that data as well, and obtained the consistent result that the DPLN provides a very good data fit for the British case while the LN fits poorly. Details are available upon request from the authors.

Legend: N is the number of data points (cities), *coverage* is the percentage of the total US population represented by the data set. *Min* and *Max* are the population size of the smallest and the largest settlement. Parameters are estimated with maximum likelihood. $ln(L^i_j)$ is the absolute value for the log-likelihood of distribution $j = LN; DPLN$ for the respective dataset. The Akaike information criterion for dataset i and distribution j is computed as $AIC^i_j = 2k_j - 2ln(L^i_j)$, and the Bayesian information criterion as $BIC^i_j = k_j \cdot ln(N^i) - 2ln(L^i_j)$, with k_j denoting the number of parameters of distribution j. Both criteria favor the distribution j that yields the lower value.

3. Implications for theories of urban growth

In this section we suggest a random urban growth model that (asymptotically) implies DPLN distributed city sizes. The model is along the lines of Eeckhout (2004) and Gabaix (1999), yet with one important difference: It combines the pure Gibrat's law with age heterogeneity across cities, which results because cities are created at different points in time.

3.1. Background

Let $S(i,t)$ be the population size of some city i where t indexes time, and let $\dot{S}(i,t)/S(i,t) = \epsilon(i,t)$, where $\dot{S}(i,t)$ refers to the time derivative. Here, $\epsilon(i,t)$ is the population growth rate of that city between t and $t + \mathrm{d}t$. Gibrat's law states that the growth rate of a city is independent of its current size. This is satisfied when $\epsilon(i,t)$ follows a geometric Brownian motion:

$$\epsilon(i,t) = \gamma \cdot dt + \varsigma \cdot \mathrm{d}B(i,t), \tag{2}$$

where $\gamma > 0$ is the drift and $\varsigma > 0$ is the variability of this stochastic growth process.

Let T denote the city's age, i.e., the time that has elapsed since that city was created. Assuming that the initial size (in logarithmic scale) at the time of birth, $\ln S(i,0)$, is drawn from a normal distribution with mean s_0 and variance σ_0^2, it follows from the central limit theorem and standard Itô calculus that the probability distribution for the (log) size of a city with age T is given by:

$$\ln S(i,T) \sim N\left(s_0 + \mu_t(T), \sigma_0^2 + \sigma_t^2(T)\right), \tag{3}$$

with

$$\mu_t(T) = \left((\gamma - \varsigma^2/2) \cdot T\right) \qquad \text{and} \qquad \sigma_t^2(T) = \varsigma^2 \cdot T. \tag{4}$$

Assuming that $\gamma > \varsigma^2/2$, expressions (3) and (4) thus show that older cities are larger on average, since they had longer time to grow.

Turning to the country's overall city size distribution in a given point in time, this is simply a mixture of the size probability distributions of all cities that exist at that time. Suppose for the moment that all cities have the same age $T = \bar{T}$ and the same initial size s_0. In that case, it is easy to see from (3) and (4) that all city-specific size probability distributions are LN with the same parameters $s_0 + \mu_t(\bar{T})$ and $\sigma_t^2(\bar{T})$. The overall city size distribution that results from a mixture of these identical distributions is then itself also LN with parameters $s_0 + \mu_t(\bar{T})$ and $\sigma_t^2(\bar{T})$. In general, however, the overall city size distribution is *not* a LN but a mixture of *different* city-specific LNs with parameters dependent on the city's age. Put differently, *an urban system where the pure Gibrat's law holds only converges to a LN city size distribution if all cities have the same age.*

To account for age heterogeneity, assume that the mixing parameter T is exponentially distributed across cities with shape parameter λ. Under this assumption, the resulting city size distribution $f(S)$ at a given point in time is the Riemann-Stieltjes integral of the LN with respect to the exponential distribution:

$$f(S) = \int LN\left(S\,; \mu_t(T), \sigma_t^2(T)\right) d\,exp(T; \lambda). \tag{5}$$

As is shown by Reed (2002, 2003), the DPLN as given in (1) is the closed form solution for this density function (see the appendix for more details).

The foundation of the DPLN is, hence, a model with random (scale-independent) urban growth and an exponential age distribution across cities. Such an age distribution, in which there are more young than old cities, arises dynamically if the mass (the "*number*") of cities is increasing at a constant rate λ over time. Initial city sizes are drawn from a common LN distribution, and after its creation, each city then grows stochastically according to (2). It can be shown (see the appendix) that the slope parameters of the DPLN (α and β) are increasing in λ, so that the city size distribution has fatter tails the lower λ is. Intuitively, if λ is very low, the upper tail of the size distribution is dominated by some very old cities which tend to be very large. Vice versa, the higher λ is, the thinner is the upper tail of the DPLN since the age heterogeneity is lower and there are more young cities. In the limit with $\lambda \to \infty$, all cities have the same age and the DPLN degenerates to a LN.[11]

It is also possible to consider alternative assumptions on the dynamics of city creation and, hence, on the age distribution across cities. Suppose, for example, that the number of cities is growing over time but at a decreasing rate. In that case, there are more old than young cities. In the appendix, we show the resulting approximate city size distribution for this alternative assumption on the distribution of the mixing parameter T. As it turns out, that distribution performs better than the LN but worse than the DPLN when fitted to US data (both for places and area clusters). In other words, the worst empirical performance is delivered by a random urban growth model *without* age heterogeneity across cities (the LN). Models that allow for city creation and age heterogeneity perform better, and the DPLN is not outperformed by the alternative model.[12]

3.2. An economic model of random urban growth and endogenous city creation

Our aim in this subsection is to describe an economic model that introduces endogenous city creation and age heterogeneity across cities into an urban growth framework where the pure Gibrat's law holds. The model is an extension of Eeckhout (2004) and is analyzed in further detail in Giesen (2012). In this short paper we focus on the main economic mechanisms and results.

Consider an economy with a total population size of $S(t)$ infinitely lived agents that is growing at the exogenous rate $g_S > 0$. The economy consists of a continuum of $N(t)$ locations/cities at time t. Firms produce a perfectly tradeable commodity using labor only, and operate under perfect competition. The wage $w(i,t)$ is equal to the marginal product of labor in location i and time t and depends positively on the city's overall productivity $A(i,t)$ and population size $S(i,t)$. With

[11]The model by Eeckhout (2004) corresponds to this knife-edge case since all cities have the same age.

[12]One could consider yet other distributions of T, but closed form solutions for the resulting density function of city sizes are then often not attainable.

respect to productivity, we assume a Brownian motion $\dot{A}(i,t)/A(i,t) = \epsilon^A(i,t)$ where $\epsilon^A(i,t) = g_A \cdot dt + \varsigma_A \cdot \mathrm{d}B(i,t)$. That is, locations are hit by idiosyncratic and permanent *i.i.d.* shocks with positive drift $g_A \geq 0$ and variability ς_A, so that $A(i,t)$ reflects the history of productivity shocks in city i up to time t. The positive effect of $S(i,t)$ on $w(i,t)$ represents a positive localized externality: The wage is higher in larger cities because of agglomeration effects such as knowledge spillovers. Finally, within each city, agents consume land and have to commute to work, thereby losing effective working time. This represents a competing negative localized externality: land prices are higher and more time is lost for commuting in larger cities.

Overall, it is assumed that the utility of a city resident in city i at time t, $V(i,t)$, is decreasing in $S(i,t)$ because the negative external effect dominates the positive one. Considering the particular functional forms for the localized externalities as used by Eeckhout (2004), indirect utility can be written as $V(i,t) = \Phi\left(A(i,t) \cdot S(i,t)^{-\Theta}\right)^{\alpha}$, where $\{\alpha, \Theta, \Phi\}$ are parameters that are the same across cities and time.[13] Workers are freely mobile so that indirect utility is equalized across cities. Using the property that $V(i,t) = V(j,t)$ for all cities i and j and at each point in time, it can be shown that the indirect utility level in the spatial equilibrium satisfies (see Giesen, 2012):

$$V^*(t) = \Phi\left(A(t) \cdot S(t)^{-\Theta}\right)^{\alpha}, \qquad \text{where} \quad A(t) = \left(\int_{i=0}^{N(t)} A(i,t)^{1/\Theta}\right)^{\Theta} \mathrm{d}i \tag{6}$$

The equilibrium size of a single city then reflects its relative productivity level, $S(i,t)^*/S(t) = (A(i,t)/A(t))^{1/\Theta}$, and it immediately follows from this relationship that Gibrat's law holds since $A(i,t)$ evolves randomly around the common trend g_A. Furthermore, it follows from (6) that $V(t)^*$ is decreasing in $S(t)$. If more workers have to be fitted into a fixed set of cities, city sizes would rise proportionally and all individuals end up worse off because of the prevalent negative localized externality. In other words, since the population grows at the rate $g_S > 0$ welfare would decrease over time, ceteris paribus. Vice versa, $V(t)^*$ is increasing in $A(t)$. Expected productivity growth g_A thus raises welfare over time, ceteris paribus, since it increases wages everywhere.

Now consider the creation of new cities. In particular, we assume that there is a large amount of featureless land where cities can be formed by a social planner.[14] Let $x(t)$ denote the mass of newly created cities between t and $t+\mathrm{d}t$, so that $N(\bar{t}) = \int_{t=0}^{\bar{t}} x(t)\mathrm{d}t$ is the total mass of cities existing at time $\bar{t}$. The formation of every new city imposes sunk resource costs F for developing infrastructure, the housing stock, and so on, that are borne by the currently alive population. Whenever the planner creates a new city, its initial productivity is drawn from a common LN distribution with mean A_0 and variance σ^2_{A0}.[15] Afterwards, productivity in those new cities evolves just as in any other city, i.e., according to the Brownian motion with positive drift described above.

[13]Our parameter $\Theta > 0$ refers to $|\Theta|$ in Eeckhout (2004), and our $\Phi > 0$ refers to $\alpha^{\alpha} H^{\beta}(1-\alpha-\beta)^{(1-\alpha-\beta)}$, where α and β are parameters of the utility function.

[14]If there were decentralized city creation by the workers, this would lead to the well-known coordination failures analyzed by Henderson (1974). Those externalities are not the focus of our paper. We therefore consider a social planner who chooses the efficient number of cities. Since workers are identical and freely mobile across space, the resulting spatial equilibrium allocation is also efficient.

[15]We could allow this distribution of initial productivity draws to be time-varying without affecting our main result that cities are created at a constant rate.

At the time of creation, a new city is initially empty and, hence, offers very high utility. There is inflow of population from the other cities until a new spatial equilibrium is reached. This induced inflow is stronger, the higher is the realization of $A_{i,0}$. That is, the initial city size $S_{i,0}$ reflects the initial productivity draw. The country's normed productivity $A(t)$ and equilibrium utility $V(t)$ increase, since a new productive site has opened up and the population can spread across more cities. Specifically, using (6) equilibrium utility can be rewritten as $V(t)^* = \Phi \cdot \Omega(t)^{\alpha\Theta}$ where

$$\Omega(t) = \frac{\int_{i=0}^{N(t)} A(i,t)^{1/\Theta} \mathrm{d}i}{S(t)}. \tag{7}$$

This state variable evolves according to

$$\dot{\Omega}(t) = \left(\frac{(1+g_A)^{1/\Theta}}{1+g_S} - 1\right)\Omega(t) + \frac{x(t) \cdot A_0^{1/\Theta}}{S(t)}. \tag{8}$$

The first term in (8) entails the exogenous growth rate of the (transformed) equilibrium utility for a fixed set of cities, which is increasing in g_A and decreasing in g_S. The (positive) second term is the expected benefit from developing new cities.

The planner chooses the time-path of city creation $x(t)$ in order to maximize overall welfare, taking into account the real resource costs of city creation. The present-value Hamiltonian of this dynamic problem can be written as follows,

$$\max H(t) = e^{-(\rho - g_S)\cdot t}\left(V(t)^* - \frac{x(t) \cdot \chi F}{S(t)}\right) + \lambda(t) \cdot \dot{\Omega}(t), \tag{9}$$

where $\rho > g_S > 0$ is the time discount rate, χ is the marginal utility of income that is assumed fixed, and $\lambda(t)$ is the costate variable. The planner maximizes (9) subject to the transition equation (8) and $x(t) \geq 0$. This is a standard optimal control problem reminiscent of problems where a social planner invests into a stock of public capital. It can be shown that the planner chooses the time path of city creation so as to smooth utility over time. It becomes $V^* = \Phi \cdot \Omega^{*\,\alpha\,\phi_1}$, where

$$\Omega^* = \left(\frac{\alpha\Theta\Phi \cdot A_0^{1/\Theta}}{\chi F} \cdot \frac{1+g_S}{(1+\rho-g_S)(1+g_S) - (1+g_A)^{1/\Theta}}\right)^{\frac{1}{1-\alpha\Theta}} \tag{10}$$

The time path of city creation is then given by

$$x^*(t) = e^{g_S \cdot t} \cdot \left[\left(1 - \frac{(1+g_A)^{1/\Theta}}{1+g_S}\right) \cdot \frac{S_0}{A_0^{1/\Theta}}\right] \cdot \Omega^* \tag{11}$$

The condition $x(t) \geq 0$ requires that $(1+g_A)^{1/\Theta} < (1+g_S)$, i.e., population growth must be sufficiently strong relative to exogenous productivity growth. It then follows from (10) and (11) that the mass of created cities is higher at every point in time the higher is g_A and the lower is F. Most importantly, it follows from (11) that $\hat{x} = \dot{x}(t)/x(t) = g_S$. In other words, the mass of newly created cities increases at a constant rate over time, namely the population growth rate. Productivity growth g_A positively affects the *level* of city creation, but not its *growth rate*.

When the mass of new born cities increases at a constant rate, so does the total number of cities. Specifically, we have $\hat{N} = \dot{N}(t)/N(t) = x(t)/N(t)$ which becomes $\hat{N} = \frac{e^{g_S t}}{e^{g_S t}-1} \cdot g_S$ and thus (quickly) converges to g_S. With this, we thus have a framework where: i) the number of cities grows at a constant rate over time, ii) initial sizes of newly born cities follow a LN distribution, reflecting the initial productivities of those new cities, and iii) growth among existing cities obeys the pure Gibrat's law as they are hit by idiosyncratic productivity shocks. As described in the previous subsection, the overall city size distribution thus asymptotically follows the DPLN parameterization, as it is a mixture of city-specific LN size probability distributions where the mixing parameter T (the city age) is exponentially distributed.

3.3. Discussion

The key difference between the baseline model by Eeckhout (2004) and the above extension is that the number of cities grows at a constant rate in our approach, while Eeckhout considers a fixed number of equally old cities. Empirical evidence on the evolution of the number of cities in a country is still scant, particularly when small settlements ought to be included in the analysis. However, it seems fair to say that city creation is a realistic assumption at least over the longer course of history. Henderson and Wang (2007) report, for example, that the worldwide number of cities with more than 100,000 inhabitants increased from 1220 to 2684 between the years 1960 and 2000. Another well-known dataset that traces cities over a very long time period is due to Bairoch (1988), who shows that the number of European cities (except Russia) with more than 20,000 inhabitants increased from 39 to 130 between the years 1000 and 1760. These datasets are not perfectly suited for our purpose, because they only include cities that are larger than a certain threshold. We therefore do not know the cities' actual creation date, but only when they have crossed the threshold. Still, those datasets clearly suggest that the number of cities is not fixed but growing over time, which in turn implies that cities in reality differ by age. As for the universe of US settlements, there is consistent evidence that their number has increased. Gonzáles-Val (2010) reports that the number of US Census places has risen from slightly more than 10,000 in the year 1900 to about 20,000 in the year 2000.

If evidence suggests that the number of cities has indeed increased, what still might seem empirically implausible is the constant growth rate over time. In the model presented above, the planner chooses constant growth in the number of cities in order to smooth utility in the light of constant growth of population and productivity. With a bit more informal approach, it is also possible to consider different dynamics of city creation without crucially affecting our main results for the resulting city size distribution. First, we may consider a scenario where the number of cities first grows exponentially in an early phase of history (say, for $t < \hat{T}$), but city creation then stops at $t = \hat{T}$ and the number of cities stays fixed afterwards. Such a scenario may roughly match the experience of some mature European countries where settlement creation activity was strong in former times but rather low recently. The city age distribution in that modified model is still a (shifted) exponential, however, and the previously described mixing of the city-specific size probability distributions works analogously as in the baseline case. City sizes thus still converge to a DPLN distribution. Second, as discussed above, we may also consider entirely different dynamics, e.g., a case where the number of cities is growing at a decreasing rate over time (see appendix). The implied city size distribution is still closer to the US data than the LN, but there is no evidence that this alternative model leads to an improved data fit relative to the DPLN.

Finally, notice that many other urban models also consider exponential growth in the number of cities. For example, Gabaix (1999) analyzes an extension of his framework where cities are exogenously created at a constant rate and then grow according to the "impure" Gibrat's law subject to a lower bound. City sizes adhere to Zipf's law in that case, and Gabaix shows that the upper tail of the size distribution is dominated by old cities. Our model is similar in that respect. We, however, explicitly consider *endogenous* city creation and we do not impose a lower bound for city sizes. This then leads to DPLN distributed city sizes, which is fully consistent with a Zipfian Pareto shape for the sizes of the large cities. Another important urban model with endogenous city creation is due to Rossi-Hansberg and Wright (2007). In their model, cities specialize in particular industries and productivity shocks are industry- rather than city-specific. They allow for both city creation and destruction, and show that this adjustment at the extensive margin allows reconciling increasing returns at the local level with constant returns (balanced growth) at the aggregate level. In contrast to our approach, they also focus on the upper tail of the size distribution and on the Zipfian power law among large cities.

4. Conclusions

In this paper we have shown that the DPLN distribution provides an excellent fit to the US city size distribution, regardless of whether cities are economically or administratively defined. The key feature of the DPLN is age heterogeneity across cities resulting from the fact that cities are created at different points in time. Once this feature is taken into account, the DPLN is the natural outcome of an otherwise standard scale-independent urban growth process.

The DPLN is useful as it can settle several controversies on city size distributions. First, there is a discussion how a "city" should be defined. Our results suggest that this may be a second-order problem, at least when it comes to the size distribution, because the same functional form closely approximates the data both for Census places and area clusters. Second, even if one agrees on a particular definition, there is still a lively debate about the parameterization of the city size distribution, and especially about the relationship of Zipf's law and the LN distribution. In particular, several authors have noticed that the sizes of large cities follow a distinctive power law pattern that is not too well captured by the LN. Outside the upper tail, however, Zipf's law breaks down and the LN starts to fit well (see, e.g., Levy 2009). With the DPLN parameterization this controversy can be resolved, because it combines a lognormal body with a power law (Zipf's law) in the upper tail. Third, the DPLN is not an ad-hoc functional form that is chosen purely on the basis of data fit, but it has an explicit theoretical foundation. While the LN follows under the "pure" Gibrat's law with a fixed number of cities, and Zipf's law under an "impure" version with a lower bound on city sizes , the DPLN follows when the "pure" Gibrat's law is combined with a growing number of cities. We have shown that this growing number of cities is the natural outcome of a model where a growing population allocates over an endogenously determined set of locations.

References

[1] Auerbach F., (1913). Das Gesetz der Bevölkerungskonzentration. Petermanns Geographische Mitteilungen 59, 74-76.

[2] Bairoch, P., Braider, C., 1988. Cities and Economic Development: From Dawn of History to the Present. Univ. of Chicago Press.

[3] Eeckhout, J., 2004. Gibrat's law for all cities. American Economic Review 94, 1429-1451.

[4] Eeckhout, J., 2009. Gibrat's law for (all) cities: reply. American Economic Review 99, 1676-1683.

[5] Gabaix, X., 1999. Zipf's law for cities: an explanation. Quarterly Journal of Economics 114, 739-767.

[6] Gibrat, R., 1931. Les inégalités économiques. Paris: Librairie du Recueil Sirey.

[7] Giesen, K., Suedekum, J., Zimmermann, A., 2010. The size distribution across all cities - double Pareto lognormal Strikes. Journal of Urban Economics 68, 129-137.

[8] Giesen, K., 2012. Random urban growth and endogenous city formation. mimeo.

[9] Gonzáles-Val, R., 2010. The evolution of U.S. city size distribution from a long term perspective, Journal of Regional Science 50, 952 - 972.

[10] Henderson, V., 1974. The sizes and types of cities. American Economic Review 64, 640 - 656.

[11] Henderson, J., Wang, H., 2007. Urbanization and city growth: The role of institutions. Regional Science and Urban Economics 37, 283-313.

[12] Ioannides, Y., Skouras, S., 2009. Gibrat's law for (all) cities: a rejoinder. Tufts University, Economics Department Discussion Paper.

[13] Levy, M., 2009. Gibrat's law for (all) cities: comment. American Economic Review 99, 1672-1675.

[14] Malevergne, Y., Pisarenko, V., Sornette, D., 2011. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. Physical Review E 83, 036111.

[15] Nitsch V., 2004. Zipf zipped. Journal of Urban Economics 57, 86-100.

[16] Reed, W., 2002. On the rank-size distribution for human settlements. Journal of Regional Science 42, 1-17.

[17] Reed, W., 2003. The Pareto law of incomes - an explanation and an extension. Physica A 319, 469 - 486.

[18] Reed, W., Jorgensen, M., 2005. The double Pareto-lognormal distribution - A new parametric model for size distributions. Communications in Statistics 34, 1733-1753.

[19] Rozenfeld, H.D., Rybski, Makse, H.A., Gabaix, X., 2011. The area and population of cities: New insights from a different perspective on cities. American Economic Review 101, 2205-2225.

[20] Rossi-Hansberg, E., Wright, M., 2007. Urban structure and growth. Review of Economic Studies 74, 597-624.

[21] Soo, K., 2005. Zipf's law for cities: a cross-country investigation. Regional Science and Urban Economics 35, 239-263.

[22] Zipf, G. K., 1949. Human Behavior and the Principle of Least Effort. Cambridge, MA: Addison-Wesley.

Appendix A: Genesis of the DPLN

Instead of directly deriving the density function of the DPLN by solving the Riemann-Stieltjes integral given in (5), one can make use of the respective moment generating function (mgf). Reed (2002) shows the mgf of a city with distribution according to equation (3) and age T is given by

$$M_{log(S_T)}(\theta) = exp\left(s_0\theta + \sigma_0^2\theta^2 + \left(\left(\gamma - \frac{\varsigma^2}{2}\right)\theta + \frac{\theta^2\varsigma^2}{2}\right)\cdot T\right) \tag{12}$$

and the corresponding mgf of the overall distribution, under which T is also a random variable, is

$$M_{log(S)}(\theta) = exp\left(s_0\theta + \frac{\sigma_0^2\theta}{2}\right)\cdot M_T\left(\left(\gamma - \frac{\varsigma^2}{2}\right)\theta + \frac{\theta^2\varsigma^2}{2}\right). \tag{13}$$

Under the assumption that T follows an exponential distribution, the mgf of time becomes $M_T(\theta) = \frac{\lambda}{\lambda-\theta}$ and therefore

$$M_{log(S)}(\theta) = \frac{exp\left(s_0\theta + \frac{\sigma_0^2\theta^2}{2}\right)}{\lambda^{-1}\left(\lambda - \left(\gamma - \frac{\varsigma^2}{2}\right)\theta - \frac{\varsigma^2}{2}\theta^2\right)}, \tag{14}$$

which can be simplified by using a partial decomposition (see Appendix B) to

$$M_{log(S)}(\theta) = exp\left(s_0\theta + \frac{\sigma_0^2\theta^2}{2}\right)\cdot\frac{\alpha\beta}{(\alpha-\theta)(\beta+\theta)}. \tag{15}$$

This shows that the distribution of $log(S)$ is the convolution of a normal distribution with an asymmetric Laplace distribution, since $exp\left(s_0\theta + \frac{\varsigma^2\theta^2}{2}\right)$ is the mgf of a normal distribution and $\frac{\alpha\beta}{(\alpha-\theta)(\beta+\theta)}$ is the mgf of an asymmetric Laplace distribution. The respective distribution of S, as represented in equation (1), is then obtained by transforming log city sizes to levels.

Appendix B: Specifics of α and β

The parameters α and β are time constant collections of the parameters γ, ς and λ, which govern the growth process. They are determined in the above partial decomposition of the mgf of the DPLN, which reduces equation (14) to (15). Therein, the parameters α and $-\beta$ are the roots of the characteristic equation

$$\left(\gamma - \frac{\sigma^2}{2}\right)\theta + \frac{\sigma^2}{2}\theta^2 - \lambda = 0$$

given by

$$\alpha = \frac{-2\gamma + \varsigma^2 + \sqrt{(-2\gamma+\varsigma^2)^2 + 8\varsigma^2\lambda}}{2\varsigma^2} \quad \text{and} \quad \beta = \frac{2\gamma - \varsigma^2 + \sqrt{(-2\gamma+\varsigma^2)^2 + 8\varsigma^2\lambda}}{2\varsigma^2}.$$

As can be seen, α and β are increasing in λ. Therefore, in the limit where $\lambda \to \infty$ this translates into $\alpha \to \infty$ and $\beta \to \infty$ and the DPLN turns to a LN, as the mgf of the DPLN in equation (15) converges to the mgf of a normal distribution.

Appendix C: Decreasing growth in the number of cities

Now consider the alternative specification, where cities are created at a decreasing rate over time. Here, the mixing parameter is distributed according to $g(T) = -\rho e^{\rho T}$, and Reed (2003) shows that the density of $log(S)$ is the convolution of a normal distribution with a distribution with density $f(S) = \frac{\alpha\beta}{(\alpha-\beta)}(exp(\beta S) - exp(\alpha S))$ where $0 < exp(S) < 1$. In this case, the resulting density of S is approximately given by:

$$\frac{\alpha\beta}{(\alpha-\beta)}\left(S^{\beta-1}e^{\left(\frac{\beta^2\sigma^2}{2}-\beta\mu\right)}\Phi\left(\frac{\mu-\beta\sigma^2-log(S)}{\sigma}\right)-S^{\alpha-1}e^{\left(\frac{\alpha^2\sigma^2}{2}-\alpha\mu\right)}\Phi\left(\frac{\mu-\alpha\sigma^2-log(S)}{\sigma}\right)\right)$$

This density function can be estimated using maximum likelihood. The respective Mathematica file is available upon request.