Spatial Economics Research Centre

SERC DISCUSSION PAPER 122

The Size Distribution Across All "Cities": A Unifying Approach

Kristian Giesen (Mercator School of Management, University of Duisburg-Essen)
Jens Suedekum (Mercator School of Management, University of Duisburg-Essen; Ruhr Graduate School of Economics, CESifo and IZA)

December 2012

This work is part of the research programme of the independent UK Spatial Economics Research Centre funded by a grant from the Economic and Social Research Council (ESRC), Department for Business, Innovation & Skills (BIS) and the Welsh Assembly Government. The support of the funders is acknowledged. The views expressed are those of the authors and do not represent the views of the funders.

The Size Distribution Across All "Cities": A Unifying Approach

Kristian Giesen* and Jens Suedekum**

December 2012

* Mercator School of Management, University of Duisburg-Essen
** Mercator School of Management, University of Duisburg-Essen; Ruhr Graduate School of Economics, CESifo and IZA

Acknowledgements
We thank Xavier Gabaix, Jacques Thisse, Henry Overman, Kurt Schmidheiny, Rafael Gonzáles-Val, Hernan Makse, Diego Rybski, Gary Brand, and seminar participants at the London School of Economics, the 2nd Workshop of Urban Economics in Barcelona and the meetings of the Urban Economics Association in Ottawa and Bratislava for very helpful comments. All errors are solely our responsibility. Parts of this paper were written while Suedekum was visiting SERC/CEP at the London School of Economics. He gratefully acknowledges the hospitality of this institution.

Abstract
Older cities in the US tend to be larger than younger ones. The distribution of city sizes is, therefore, systematically related to the country's city age distribution. We introduce endogenous city creation into a dynamic economic model of an urban system. All cities exhibit the same long-run growth rate (Gibrat's law), but young cities initially grow faster. The double Pareto lognormal (DPLN) emerges as the city size distribution in our model. The DPLN unifies the lognormal and the Pareto distribution (Zipf's law), and closely fits US city size data. This evidence can potentially resolve several debates from the recent literature.

Keywords: Zipf's law, Gibrat's law, city size distributions, city age, DPLN distribution
JEL Classifications: R11; R12

1 Introduction

Ever since the seminal works by Felix Auerbach (1913) and George Zipf (1949), there has been vast interest in the distribution of city sizes in an economy. This research has largely neglected, however, that cities also differ along another fundamental dimension: age. Using novel data on the foundation dates of almost 7,000 American cities, we show that age heterogeneity is a salient empirical fact. The average US city in our sample is 140 years old today, but there are strong differences. Boston was founded around 383 years ago, while places like Laguna Woods (CA) not even had their 13^{th} birthday yet. Importantly, we find that age and size are positively correlated: Doubling the age of a city is – on average – associated with an increase of the city's current population size by 65%. The country's city size distribution and the city age distribution, therefore, have a systematic relationship that we explore in this paper.

We introduce endogenous city creation into a dynamic economic model of an urban system. Our starting point is the influential approach by Xavier Gabaix (1999) and Jan Eeckhout (2004) where the distribution of city sizes in the economy emerges from *Gibrat's law*, that is, from a scale-independent urban growth process where all cities grow with the same expected rate irrespective of their current size. In Eeckhout (2004) there is a fixed population that is freely mobile across a fixed number of equally old cities. City sizes then – in fact, only then – converge to a lognormal (LN) distribution, as cities face random and permanent productivity shocks and thus obey to the "pure" Gibrat's law. The famous *Zipf's law* for city sizes emerges instead of the LN when an "impurity" is added, and cities are prevented from becoming too small (Gabaix 1999).[1]

Our model extends Eeckhout's (2004) approach by assuming that the country's total population is growing. If the number of cities were fixed, this would lead to rising congestion and decreasing equilibrium utility over time, as more people have to be squeezed into the urban system. We hence consider endogenous city creation, which allows the population to spread across more cities and leads to age differences between cities. When a new city is founded, it starts from a randomly drawn initial productivity and accordingly adjusts to its equilibrium starting size through population inflows from the established cities. The new city exhibits strong growth during this transition. Afterwards, all cities are subject to random productivity shocks which affect the evolution of their equilibrium sizes. Gibrat's law is therefore at work in the long run, but there are also deviations as young cities grow faster than established cities early on, while reverting to the economy-wide average later. Such a pattern is consistent with recent empirical evidence for US urban growth over the last two centuries.[2] Moreover, as expected city growth is positive, our model predicts – in line with the aforementioned facts – that older cities tend to be larger than younger ones.

[1]Zipf's law states that city sizes follow a Pareto distribution with tail exponent close to one. The country's largest city is then twice as large as the second-largest, three times as large as the third-largest city, and so on.

[2]See, in particular, Klaus Desmet and Jordan Rappaport (2012) and Rafael Gonzáles-Val, Maria Sánchez-Vidal and Elisabet Viladecans-Marsal (2012).

From this urban system model with its empirically relevant new features – city age heterogeneity, a positive correlation of age and size, and Gibrat's law with stronger initial growth of young cities – we then derive the city size distribution (CSD) that emerges in our economy. This turns out to be the so-called *double Pareto lognormal* (DPLN) distribution. The functional form of the DPLN is characterized by a lognormal body and power laws in the tails, which are fatter the stronger the age differences between cities are. It thus unifies the LN suggested by Eeckhout (2004), and the Pareto distribution (Zipf's law) advocated by Gabaix (1999) and by Hernán Rozenfeld, Diego Rybski, Gabaix and Hernán Makse (2011) in a single model for the overall CSD.

Taking this DPLN distribution to US city size data, we find that it delivers an excellent and robust fit, and it (easily) outperforms the LN, Zipf's law and also other functional forms that have been suggested. In other words, our urban system model which takes the nexus of city age and city size into account, is more successful in matching contemporaneous city size data than alternative theoretical frameworks that disregard this relationship. This evidence of the superior fit of the DPLN distribution has important implications, because it can potentially settle (at least) two controversial issues from the recent empirical literature on CSDs.

The first one deals with the question how to define a *city*. In fact, the influential contributions by Eeckhout (2004) and by Rozenfeld *et al.* (2011) use different city size data, and come to divergent conclusions about the appropriate parameterization of the CSD. Using administratively defined US *Census places*, Eeckhout (2004) shows that the LN closely fits the data, thus providing empirical support for his model. Rozenfeld *et al.* (2011), in contrast, use a bottom-up approach of constructing *area clusters* from high resolution data on population density in the US, independently of administrative boundaries. They emphasize that the sizes of area clusters with at least 13,000 inhabitants (together accounting for 63 % of the US population) closely obey to Zipf's law. Yet, when analyzing the distribution of the entire US population across space, that is, the *overall* CSD across *all* clusters, it turns out that Zipf's law breaks down. Importantly, the LN also provides a poor fit to this distribution across all clusters, as we show in Figure 1 below. The LN thus seems to approximate the overall CSD fairly well for one definition of US cities (Census places), but not for the other (area clusters). Unfortunately, Rozenfeld *et al.* (2011) do not suggest a better alternative for their data.

We show that the DPLN distribution closely fits the empirical CSD across all settlements for both definitions of US cities (see Figure 1). It also performs well for other countries. Our findings thus suggest that the CSD can be robustly approximated by the same functional form regardless of which city size data is used. This evidence is, furthermore, in line with, but goes beyond the findings of Rozenfeld *et al.* (2011): The DPLN is a parameterization for the overall CSD across all clusters that is fully consistent with their claim that Zipf's law holds among the large clusters.

Second, our paper may reconcile another recent debate. Some authors (most notably Moshe Levy 2009, Yannis Ioannides and Spyros Skouras 2013, and Yannik Malevergne, Vladilen Pisarenko and Didier Sornette 2011) have argued that the large Census places also 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 our finding that the sizes of Census places are better approximated by a DPLN, rather than by a LN distribution. More generally, the DPLN builds a bridge between the “old” and the “new” literature on CSDs. Dozens of older studies have found support for Pareto distributed city sizes across different countries and time periods.[3] Eeckhout’s findings have challenged this conclusion, since the LN does not actually feature a Zipfian power law in the upper tail. The DPLN, on the other hand, is fully consistent with Zipf’s law and incorporates it into a model for the overall size distribution across all cities.

The rest of this paper is organized as follows. In Section 2 we present our evidence on the distribution of city sizes and show that the DPLN fits the empirical data better than other parameterizations. Section 3 turns to our theoretical model of an urban system with endogenous city creation. There we show that age heterogeneity across cities, together with Gibrat’s law, is key to understanding why the DPLN distribution of city sizes emerges. Section 4 presents our novel empirical evidence on the nexus of city age and city size in the US. Finally, Section 5 concludes.

2 City size distributions: The evidence

2.1 Data

For our empirical analysis of the city size distribution (CSD) we utilize two different definitions of US “cities”: Census places and area clusters. The former dataset refers to the year 2000 and includes administratively defined settlements according to legal boundaries. It contains 25,359 cities with sizes ranging from one to about 8 million inhabitants (New York City). 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 quite arbitrary, as two Census places may be considered as separate cities even though they are essentially part of the same city.[4]

The second dataset has been constructed by (and is explained in detail in) Rozenfeld *et al.* (2008, 2011). Here, cities are defined by using a clustering 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

[3]Because of data limitations, those older studies were forced to use truncated data sets which only include cities above a certain threshold size. Volker Nitsch (2004) summarizes this first wave of research on the CSD.

[4]More details about the widely used Census places data can be found in the Geographic Areas Reference Manual available online under http://www.census.gov/geo/www/garm.html. A further problem with this data is that it only represents 74 % of the total US population who reside in incorporated or Census designated places.

million inhabitants (the New York cluster). The advantage of this data is that cities are defined as genuine agglomerations ignoring administrative boundaries.[5] 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 empirical CSDs for both definitions of cities, see the black solid lines. As can be seen, the mean size of area clusters is higher than for the Census places, while the the variance is lower. These distributional features result from the fact that the clustering algorithm tends to connect adjacent places into one agglomeration (the same area cluster), as is explained in detail by Rozenfeld *et al.* (2008).

2.2 Parameterization and comparison of data fit

We first fit the LN distribution to the data by using maximum likelihood estimation (see Table 1 for the results). Figure 1 depicts the fitted LN distributions as the grey solid lines. For the Census places, the figure corroborates Eeckhout's (2004) finding: the LN indeed provides a good fit. However, when using the area clusters, 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, outside the upper tail Zipf's law eventually breaks down and, hence, it is not a useful parameterization for the overall CSD.

Our suggested functional form for the overall CSD is the DPLN distribution, which has been first introduced by William J. Reed (2002) and is further discussed by Reed and Murray Jorgensen (2005). It has the following density function for city sizes S:

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

In (1), the 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 if $S \to 0$ then $f(S) \sim S^{\beta-1}$. The slope parameters of the Pareto tails are thus α and β, 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, that is, it can be more peaked (leptokurtic) or more flat (platykurtic) than the LN.

[5] In the top range these area clusters are often coincident with metropolitan statistical areas (MSAs). However, unlike the MSAs, the area clusters data is not subject to a minimum threshold size but gives a comprehensive portray how the US population spreads across space.

[6] We have verified the result by Rozenfeld *et al.* (2011). Using only area clusters that are larger than 13,000 inhabitants, a standard rank-size regression yields a highly significant tail exponent of 0.994 with a R^2 level of 0.99.

Figure 1: Kernel density estimates and fitted LN + DPLN distributions

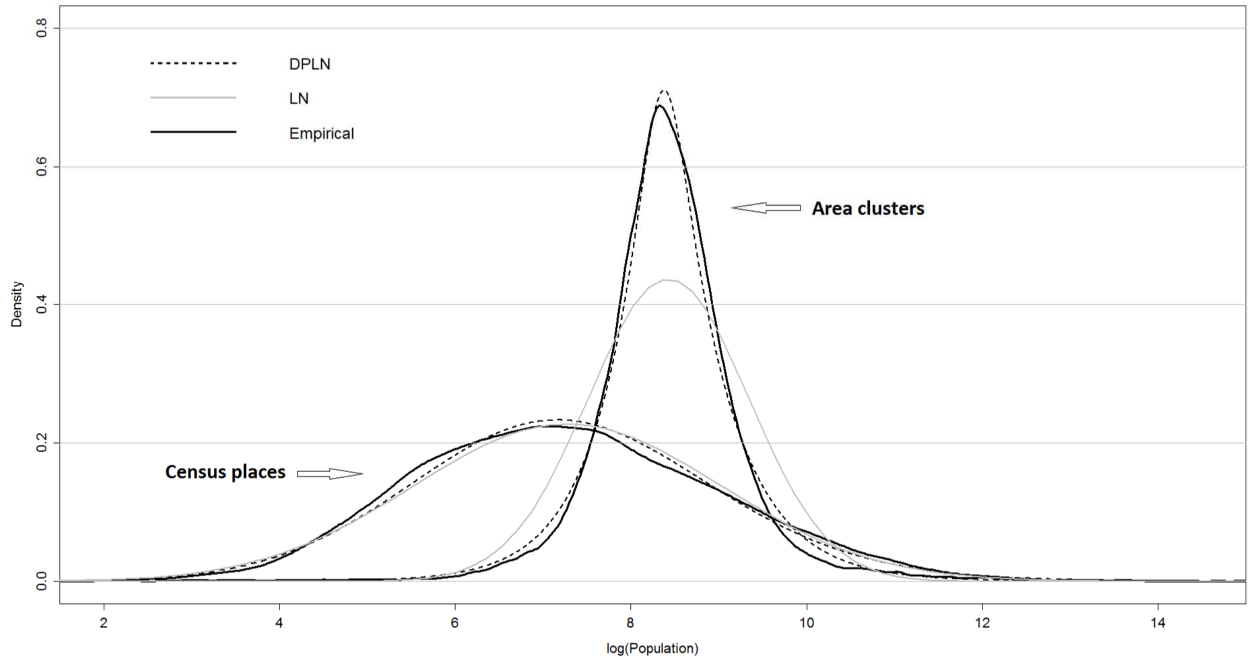


Table 1: Data and estimated parameters

	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
$ln(L_j^i)$	-225,493.9	-229,171.3	-234,711.2	-234,773.1

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. $ln(L_j^i)$ is 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 $AIC_j^i = 2k_j - 2ln(L_j^i)$, and the Bayesian information criterion as $BIC_j^i = k_j \cdot ln(N^i) - 2ln(L_j^i)$, with k_j denoting the number of parameters of distribution j. Both criteria favor the distribution j that yields the lower value.

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 as the dashed black lines. As can be seen, the DPLN provides a very close fit to the empirical CSD for *both* definitions of cities. Certainly the DPLN does a better job than the LN. For the area clusters this is self-evident by visual inspection. For the Census places, the performance difference is less pronounced. Still, the DPLN clearly fits better than the LN, even when taking into account that there are two more parameters that need to be estimated. This improvement in the adjusted performance can be seen from the Akaike (AIC) and the Bayesian information criterion (BIC), which are also reported in Table 1. Standard statistical specification tests convey the same message: for both data sets, the LN is rejected much earlier than the DPLN.[8]

2.3 Discussion

The better performance of the DPLN also holds for other countries. Kristian Giesen, Jens Suedekum and Arndt Zimmermann (2010) analyze the CSDs of seven other economies, using data on administratively defined cities comparable to the Census places. Using various model selection tests, they show that the DPLN outperforms the LN in terms of adjusted fit for almost all countries (the only exception is Switzerland) but they do not explain the theory underlying the DPLN. An additional empirical contribution of this paper is to show that the superior fit of the DPLN also holds for the recently developed US area clusters data. Rozenfeld *et al.* (2011) also provide similarly defined area clusters data for Great Britain. We have used that data as well, and obtained the consistent result that the DPLN provides a very good fit while the LN fits poorly.

Finally, the DPLN also outperforms other parameterizations that have been suggested. In particular, Ioannides and Skouras (2013) suggest a mixture of LN and Pareto as the appropriate functional form for the overall CSD, and estimate several versions of it using the US Census places and area clusters data. However, while their ad hoc parameterizations fit better than the LN, they deliver a worse fit for both data sets than the DPLN.[9] In addition, Rafael Gonzáles-Val, Fernando Sanz and Arturo Ramos (2012) compare the DPLN and three other parameterizations for the overall CSD, using data from Italy, Spain and the US. They find that the DPLN consistently delivers a better fit than the competing distributions in all three countries.

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

[8] 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 (Census 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.

[9] This can be immediately seen by comparing their Table 1 with our Table 1 above. For the LN distribution we obtain exactly the same results as they do, since we have used the same data. Comparing the log-likelihood, the AIC and the BIC for their parameterizations with our values presented above, it follows that the DPLN is more successful in matching the empirical CSDs than their ad hoc mixture model.

3 The model

The last section has shown that the DPLN distribution is a very useful parameterization for the empirical CSD that delivers a robust data fit across different data sets. We now turn to the theory and explain *why* the DPLN distribution of city sizes may emerge. Before we describe our dynamic economic model in Section 3.2., it is useful to first discuss some background about the stochastic foundations of CSDs in an urban system where Gibrat's law holds.

3.1 Background

Gibrat's law states that the growth rate of a city is independent of its current size. In this subsection, we first describe what this implies for the stochastic evolution of the size of a single city, and then turn to the overall city size distribution in the economy.

Let $S(i,t)$ be the size of city i at time t, and let $dS(i,t)/S(i,t) = \epsilon(i,t)$ denote the population growth rate between t and $t + \mathrm{d}t$. The "pure" Gibrat's law is satisfied in continuous time when $\epsilon(i,t)$ follows a geometric Brownian motion of the following form:

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

where $B(i,t)$ is a Wiener process, $\gamma \geq 0$ is the positive drift, and $\varsigma > 0$ is the variability of this stochastic urban growth process.

Assume that the initial size of city i in logarithmic scale at the time of birth, $\ln S(i,0)$, is drawn from some distribution with finite mean $s_0 > 0$ and variance $\sigma_0^2 \geq 0$. Now move ahead in time and consider the probability distribution for the size of that city at time T. It follows from the central limit theorem and standard Itô calculus that the (log) size of that city in T can be described by the following size probability distribution:

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

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

The expected size of a city, conditional on its age T, is thus $E[S(i,T)] = exp(s_0 + \sigma_0^2/2 + \gamma \cdot T)$. Provided $\gamma > 0$, this shows that older cities are larger on average since they had longer time to grow under the process specified in (2). The conditional variance of city sizes is also larger for older cities, since they were exposed to random shocks for a longer time.

Turning to the country's overall CSD in a given point in time, this is the *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 $T = \overline{T}$. 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 $s_0 + \mu_t(\overline{T})$ and $\sigma_0^2 + \sigma_t^2(\overline{T})$. The overall CSD that results from a mixture of these identical distributions is then itself also LN with parameters $s_0 + \mu_t(\overline{T})$ and $\sigma_0^2 + \sigma_t^2(\overline{T})$. For the more general case with age heterogeneity across cities and strictly positive drift in the stochastic growth process, however, the overall CSD is *not* a LN but a mixture of *different* LNs with parameters dependent on the city's age.[10]

In fact, the overall CSD in a given point in time, $f(S)$, can be written as the Riemann-Stieltjes integral of the LN with respect to the distribution of the mixing parameter T. Let this distribution, which in our context is the city age distribution, be denoted by $g(T)$. We then have

$$f(S) = \int LN\left(S; s_0 + \mu_t(T), \sigma_0^2 + \sigma_t^2(T)\right) dg(T). \tag{5}$$

For particular cases of the distribution $g(T)$ this integral in (5) can be solved analytically. In particular, assume that the mixing parameter T is exponentially distributed with shape parameter λ, that is, $g(T) = exp(T; \lambda)$. As is shown by Reed (2002, 2003), the DPLN as given in (1) is then the solution for this density function $f(S)$ (see the appendix for details). For other age distributions $g(T)$ the function $f(S)$ may still exist, but often it cannot be solved in closed form.

Summing up, there are two important insights: i) *an urban system where the pure Gibrat's law (with positive drift) holds only converges to an overall CSD with LN distributed city sizes if all cities have the same age, ii) if cities differ by age, such that the age distribution across cities is exponential, then Gibrat's law implies an overall CSD where city sizes follow a DPLN distribution.*

An exponential age distribution across cities arises dynamically if the mass (the "*number*") of cities is increasing at a constant rate λ over time, where λ is the parameter of the exponential distribution (see Giesen 2012). It can be shown (see the appendix) that the slope parameters of the DPLN (α and β) are increasing in λ, so that the CSD has fatter tails the lower λ is. Intuitively, if λ is very low, the upper tail of the CSD is dominated by a small number of 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 across cities is lower.[11]

Notice further that an exponential city age distribution does not require sustained growth in the mass of cities. Consider, for example, 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. The city age distribution $g(T)$ is then still a *shifted* exponential distribution,[12] and the mixing of the city-specific size probability distributions works

[10]Stated differently, the *conditional* CSD across all cities with the same age T is a LN when urban growth follows Gibrat's law as in (2). However, the *unconditional* CSD across all cities is in general not a LN.

[11]In the limit with $\lambda \to \infty$, all cities have the same age and the DPLN turns to a LN. The scenario studied by Eeckhout (2004) corresponds to this case with a degenerate age distribution $g(T) = \overline{T}$. In addition, $\gamma = 0$ is assumed in his framework. In that case, even if there were age heterogeneity, there would be no positive correlation between city age and expected city size although older cities would have a higher variance in their size probability distributions. In Section 4 we provide empirical evidence that age and size are positively correlated across US cities.

[12]There are no cities younger than $\hat{T} = (t - \hat{t})$ at t, while age is exponentially distributed for cities older than $\hat{T}$.

analogously in that case. City sizes thus still converge to a DPLN distribution, although absolute size differences between cities fan out by the variance of the growth process in (2).

3.2 An urban system with endogenous city creation

We now develop an *economic* model of an urban system where the pure Gibrat's law holds and where city sizes converge to a DPLN distribution. Our starting point is a continuous time version of the urban growth framework by Eeckhout (2004). We extend that model to incorporate exogenous population growth and technological progress, as well as endogenous city creation.

Basic setup Consider an economy with a total population $S(t)$ 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 on the current city size $S(i,t)$. The positive effect of $S(i,t)$ on $w(i,t)$ represents a localized size externality: The wage is higher in larger cities because of agglomeration effects such as knowledge spillovers. On the other hand, within each city, agents consume land and have to commute to work, thereby losing effective working time. This represents a negative size externality from congestion: land prices are higher, and more time is lost for commuting in larger cities. Ultimately, as in Eeckhout (2004), we assume that the utility of a city resident in city i at time t, $V(i,t)$, is monotonically decreasing in the local population size $S(i,t)$. More specifically, considering for simplicity the particular functional forms for the localized externalities as used in that paper, indirect utility in city i at time t can be written as

$$V(i,t) = \Phi \left(A(i,t) \cdot S(i,t)^{-\Theta} \right)^{\alpha}, \tag{6}$$

where α, Θ, and Φ are positive parameters that are the same across cities and time. Notice that $V(i,t)$ is decreasing in $S(i,t)$. That is, the negative size externality dominates at the city level.

With respect to productivity, we assume that locations are hit by idiosyncratic and permanent i.i.d. shocks. More specifically, we assume a Brownian motion $dA(i,t)/A(i,t) = \epsilon^A(i,t)$ where $\epsilon^A(i,t) = g_A \cdot dt + \varsigma_A \cdot \mathrm{d}B(i,t)$. The positive drift $g_A > 0$ thus captures the expected productivity growth in the economy, while $\varsigma_A > 0$ is the variability of this stochastic growth process. The term $A(i,t)$ in (6) then reflects the history of productivity shocks in city i up to time t, and $V(i,t)$ is increasing in $A(i,t)$. That is, utility is higher in cities with higher accumulated productivity.[13]

[13] As Gabaix (1999), Eeckhout (2004), Esteban Rossi-Hansberg and Mark Wright (2007), and others, we do not explicitly model the nature of the random shocks. Our specification may provide a short-cut for a variety of micro-foundations, however, such as changes in localized production amenities, technological innovations causing relocation of firms, city-specific productivity realizations for particular matches of firms and workers, etc. Furthermore, as in those and most other models from the urban growth literature, we also do not explicitly analyze *where* in space the cities are located. See Wen-Tai Hsu (2012) for a recent model that addresses the spatial dimension of the CSD.

Spatial equilibrium Workers are freely mobile so that indirect utility is equalized across all cities at each point in time. Using the property that $V(i,t) = V(j,t)$ for all i and j, it can be shown (see Giesen 2012) that the economy-wide indirect utility level in the spatial equilibrium is:

$$V^*(t) = \Phi \left(A(t) \cdot S(t)^{-\Theta}\right)^{\alpha}, \qquad \text{where} \quad A(t) = \left(\int_{i=0}^{N(t)} A(i,t)^{1/\Theta}\right)^{\Theta} \mathrm{d}i \tag{7}$$

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 (7) 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 pervasive negative size externality. Since the total population grows at the rate $g_S > 0$, welfare would thus decrease over time, ceteris paribus. Vice versa, $V(t)^*$ is increasing in $A(t)$. Expected productivity growth $g_A > 0$ thus raises welfare over time, ceteris paribus, since it increases wages everywhere.

Endogenous city creation and growth in new cities The formation of new cities in an urban system has been analyzed ever since the classical contributions by Vernon Henderson (1974) and Masahisa Fujita (1978). This literature has shown that decentralized market allocations are typically characterized by an inefficient number of cities with inefficient sizes, and analyzed different arrangements how the involved externalities can be internalized.[14] Those aspects are not the focus of this paper, but our main interest is the city age distribution that arises endogenously from the dynamics of city formation. We therefore take the simplest possible approach, and consider a forward-looking social planner who creates the efficient number of cities over time.

In particular, assume there is a large amount of featureless land where the planner can form cities. The creation 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.[15] Whenever the planner creates a new city, its initial productivity $A_{i,0}$ is drawn from some distribution with finite mean $A_0 > 0$ and variance $\sigma^2_{A0} > 0$. Afterwards, productivity in those new cities evolves just as in any other city, namely, according to the Brownian motion with positive drift described above.

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 city's equilibrium starting

[14]See Henderson and Anthony Venables (2009) for a recent analysis of those issues in the context of a dynamic urban system model with a growing population.

[15]If city formation were costless, the planner would create an infinite number of infinitely small cities. This results from the fact that our model does not feature a U-shaped net agglomeration curve á la Henderson (1974) which would greatly complicate the analysis, but a negative size externality at the city level á la Eeckhout (2004).

size $S_{i,0}$ reflects its initial productivity draw, and the new city exhibits strong growth during the transition towards this starting size. In the model, this transition works instantaneously because of free mobility. If the transition would require some time, which is likely to be the case in reality, young cities would then initially exhibit very high growth rates. Eventually though, they revert to the growth rate of the established cities. Such a pattern, where Gibrat's law holds in the long run but where young cities (which tend to be relatively small) initially grow faster, is consistent with recent empirical evidence on US urban growth over the last two centuries.[16]

Social planner's problem Let $x(t)$ denote the mass of cities that the planner creates between t and $t+\mathrm{d}t$, which adds to the stock of existing cities $N(t)$. The formation of every new city raises the country's normed productivity $A(t)$ and equilibrium utility $V(t)$, since the population can spread across more cities. Specifically, using (7) equilibrium utility can be rewritten as $V(t)^* = \Phi \cdot \Omega(t)^{\alpha\Theta}$ where

$$\Omega(t) = \frac{\int_{i=0}^{N(t)} A(i,t)^{1/\Theta} \mathrm{d}i}{S(t)}. \tag{8}$$

This state variable evolves according to

$$\dot{\Omega}(t) = \left(\frac{(1+g_A)^{1/\Theta}}{1+g_S} - 1 \right) \Omega(t) + \frac{x(t) \cdot A_0^{1/\Theta}}{S(t)}. \tag{9}$$

The first term in (9) 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 forward-looking 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,

$$H(t) = e^{-(\rho - g_S)\cdot t} \left(V(t)^* - \frac{x(t) \cdot \chi F}{S(t)} \right) + \lambda(t) \cdot \dot{\Omega}(t), \tag{10}$$

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 (10) subject to the transition equation (9) and $x(t) \geq 0$. This is a standard optimal control problem, and it can be easily shown that the planner creates cities so as to smooth utility over time. It becomes $V^* = \Phi \cdot \Omega^{*\,\alpha\,\Theta}$, where

$$\Omega^* = \left(\frac{\alpha\Theta\Phi \cdot A_0^{1/\Theta}}{\chi F} \cdot \frac{1+g_S}{(1+\rho-g_S)(1+g_S) - (1+g_A)^{1/\Theta}} \right)^{\frac{1}{1-\alpha\Theta}} \tag{11}$$

[16]Desmet and Rappaport (2012) and Gonzáles-Val *et al.* (2012) find that, among young US cities, small ones initially grow faster than the rest of the economy. Among old cities, however, small and large ones grow with the same rate. This pattern may also be described as one of "sequential city growth" (David Cuberes 2009, 2012).

The time path of city creation is then given by

$$x^*(t) = e^{g_S \cdot t} \cdot \left[\left(1 - \frac{(1+g_A)^{1/\Theta}}{1+g_S} \right) \cdot \frac{S_0}{A_0^{1/\Theta}} \right] \cdot \Omega^* \tag{12}$$

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. We assume that this is the case. It then follows from (11) and (12) 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 (12) that $\dot{x}(t)/x(t) = g_S$.

In other words, the planner creates cities at a constant rate, namely the country's population growth rate. Productivity growth g_A positively affects the *level* of city creation, but not its *growth rate*. Finally, when the mass of new born cities increases at a constant rate, so does the total number of cities. Specifically, we have $\dot{N}(t)/N(t) = x(t)/N(t)$ which becomes $\frac{e^{(t \cdot g_S)}}{e^{(t \cdot g_S)}-1} \cdot g_S$ and thus (quickly) converges to g_S. This is chosen by the planner in view of the constant growth of the economy's total population, as the creation of new cities avoids crowding in the established cities and thereby smoothes equilibrium utility over time.

City age and city size distribution Summing up, in our framework: i) the mass of cities grows at a constant rate, which in turn leads (endogenously) to an exponential city age distribution, and ii) growth among established cities obeys to the pure Gibrat's law, as they are hit by idiosyncratic productivity shocks. City sizes will thus converge to a DPLN distribution: The city-specific size probability distributions follow a LN because of Gibrat's law, with mean and variance increasing by the city's age T. These city-specific distributions are then mixed according to the exponential age distribution, which in turn leads to the DPLN distribution for city sizes (see Section 3.1.).

Recall that the DPLN would also emerge if the city age distribution were a *shifted* exponential. That age distribution would result in our model if the population grows at the rate $g_S > 0$ for $t < \hat{t}$, but when growth unexpectedly stops at $\hat{t}$ and the overall population remains constant afterwards. Then, at $\hat{t}$, the planner stops creating cities so that their total mass remains fixed from there on.

4 City age and city size in the US urban system

The key difference between our model and the baseline framework by Eeckhout (2004) is that we consider an urban system where cities differ by their age since they are created at different points in time. In this final section we provide novel empirical evidence on age heterogeneity across American cities. Afterwards, we discuss this evidence in the light of our theoretical approach.

Although little is known so far about the number or the age structure of cities in an economy, we are not the first to analyze those issues. Among the few existing papers are Linda Dobkins and Ioannides (2001), Henderson and Hyong Gun Wang (2007), Gonzáles-Val *et al.* (2012) and

Desmet and Rappaport (2012). These studies clearly show that the number of cities has grown over time, which implies that cities differ by age.[17] However, Dobkins and Ioannides (2001) and Henderson and Wang (2007) only include cities in their analysis that are larger than a certain threshold size. Their information thus refers to the date when the city's size has crossed the threshold, but not to the city's actual creation. Desmet and Rappaport (2012) and Gonzáles-Val *et al.* (2012) comprehensively count the number of all US counties or, respectively, Census places that exist in a given time, thereby giving a more comprehensive picture of city ages in the US. However, they focus on age-dependent patterns of urban growth as discussed before, but do not address the correlation of city age and the current city size which is one of our main aims.

4.1 Data

Our data traces the actual foundation dates of American cities, which correspond in their definition to the US Census places, so that the city age data is compatible with the previously used city size data for the year 2000.[18] Among historians, there is no agreement on the precise meaning of the term "foundation date" for a city. Some claim it to be the date when the first settlers arrived at the site, when the deed for the land was granted, or when the first building was completed. However, such dates are typically unknown. As the birth date of the respective US Census place, we therefore consider the foundation date of the administrative municipality, that is, the earliest date of self-government or incorporation.

We use data from the commercially sold *Cities Databank*^{TM} that has extensively collected this information, drawing on official sources including legal citation of a law, court order, county commission order, or city charter, as well as by examining standard library citations for an original published source, or a legally designated source or repository, and the laws of each legislative session of each territory, colony and State. The data base, in total, includes 6,999 US Census places, together accounting for 138 million citizens, roughly half of the total US population in 2000.[19] For the vast majority of cases ($> 99\%$), the foundation date refers to the earliest incorporation or self-governing date of the respective place.[20] In a few instances, a merger or a consolidation of

[17] Henderson and Wang (2007) report that the worldwide number of cities with more than 100,000 inhabitants has increased from 1,220 to 2,684 between the years 1960 and 2000. Dobkins and Ioannides (2001) focus on the US urban system and find that the number of metropolitan areas with more than 50,000 inhabitants has increased from 1 in 1760 to 334 in 1990. Desmet and Rappaport use data for US counties and find that their number has increased from about 300 in 1800 to more than 3,000 in the year 2000. Finally, Gonzáles-Val *et al.* report that the number of incorporated Census places has increased from 10,496 to 19,211 in the period 1900–2000.

[18] Explicit information about the history of the area clusters is, unfortunately, not available.

[19] This sample is largely representative for the universe of all US Census places considered above. The sample's smallest city has 9 inhabitants in 2000, so even the very small places are included. Furthermore, the sample includes cities from all 51 US States, and the CSD among the cities looks similar as the overall CSD across all Census places.

[20] Incorporations under colonial law were often limited by the terms of the royal charter of the colony. Some cities

several Census places to a single one occurred, in which case the consolidation date of the new municipality is used as the birth date.[21] We exclude 17 cities where the foundation date is given by a later than the first incorporation, or where an unclear merger happened. Thereby we end up with 6,982 Census places for which we have reliable information on their foundation dates.

4.2 Empirical analysis

Table 2 gives a first overview and reports the birth dates of some selected US cities. The oldest ones – including Boston – were 370 years old in the year 2000. The youngest cities have just been founded at that time. Among the largest US cities, Detroit is relatively old with a birth date in 1802, while Las Vegas has been founded more than 100 years later. Chicago, Los Angeles and New York are in the middle of the spectrum, with ages between 102 and 167 years in 2000.[22]

In table 3 we summarize some features of the US city age distribution. The data clearly shows the development of the country from the East to the West. The average age of US cities in 2000 was 127.6 years. Cities in the "frontier" States in the South-West and West are on average much younger than that, however, while cities in the more traditional States in the Mid-West and along the East Coast (particularly in New England) are older. This is shown in the second and third row, where we divide the US into two parts. Table 3 also shows differences in the shape of the city age distribution across those two groups of US Federal States. The distribution is positively skewed among the cities in the traditional States, while it has negative skewness in the frontier States. This is also shown in Figure 2, where we graphically illustrate the city age distributions. In the traditional US States, only few cities are younger than 100 years old in the year 2000. The distribution exhibits a peak in the range between 110-120 years, and then has some very old cities to the far right in the upper tail. In the frontier States, on the other hand, the bulk of cities is younger than 100 years, and only few are older than 150 years. The shape of the age distribution for the US as a whole resembles the one in the traditional States, with the young cities from the frontier States showing up in the lower tail.

---

were also initially chartered by the British or Dutch crowns, by the royal governor of the colony, or by the colonial legislature. Such charter dates are used if they provided self-government for the city. Furthermore, for cities that were self-governing prior to the American Revolution, the earliest incorporation may not be the appropriate date of self-determination. An example is Boston, which has been self-governing since 1630 but was incorporated only in 1822. In such cases, the earlier date was used as the foundation date of the city.

[21] There are two further data issues concerning border modifications over time. First, small annexations that did not significantly change the appearance of the annexing place are not recorded in the data. Specifically, the initial founding date of the surviving place is then kept. Second, in the rare event where a place is divided into several ones, the date of the division is used as the birth date of the resulting places.

[22] The foundation date of New York refers to the consolidation of the boroughs Manhattan, Brooklyn, Richmond (Staten Island), Queens and Bronx to the Greater New York area, which is incorporated as one US Census place. Alternatively, we have assigned the year 1688 as the foundation date, where New York City joined the New England confederation. Our later results do not change depending on that choice.

Table 2: Age and size of some selected US cities

	City (Census Place)	Foundation date	Population size in 2000
1	Watertown, MA	9/17/1630	32,986
2	Boston, MA	10/29/1630	589,141
3	Stratford, CT	4/19/1640	49,976
4	Braintree, MA	5/23/1640	33,828
5	Dover, NH	11/1/1640	26,884
...	...	...	...
477	Detroit, MI	2/1/1802	951,270
...	...	...	...
1,146	Chicago, IL	8/12/1833	2,896,016
...	...	...	...
2,091	Los Angeles, CA	4/4/1850	3,694,820
...	...	...	...
4,957	New York City, NY (consolidation)	1/1/1898	8,008,278
...	...	...	...
5,762	Las Vegas, NV	3/16/1911	478,434
...	...	...	...
6,980	Laguna Woods, CA	3/24/1999	16,507
6,981	Sammamish, WA	8/31/1999	34,104
6,982	Palm Coast, FL	12/31/1999	32,732

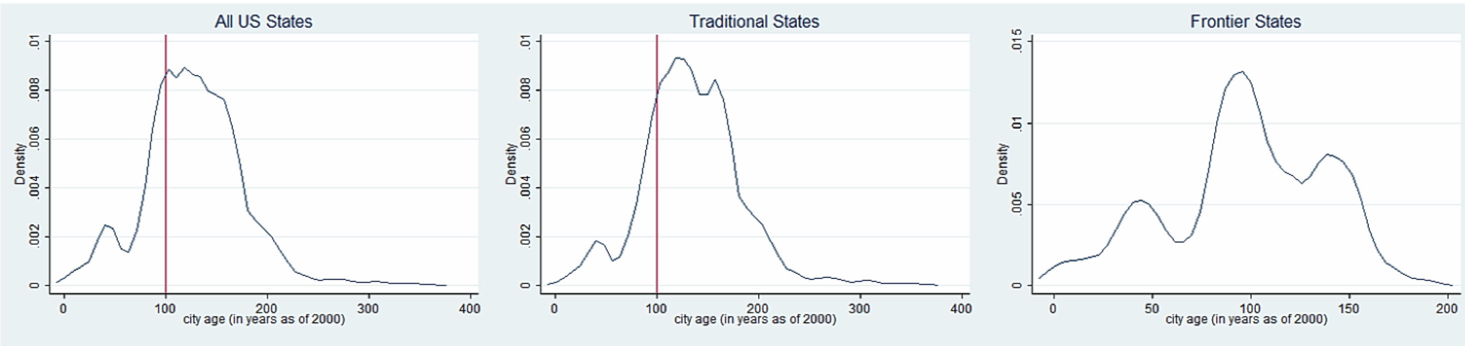
Legend: Table reports the foundation date and the population size in 2000 of selected US Census Places. Foundation dates are taken from the *Cities Databank*^{TM} and refer to the earliest date of incorporation or self-government of the municipality. Data collection ends as of 12/31/1999.

Table 3: City age distribution and correlation between age and size

	number of cities	mean age	std. deviation	skewness	age-size correl.
ALL	6,982	127.6	49.54	.492	**0.652*****
traditional States (East Coast & Mid-West)	5,677	134.5	49.06	.567	**0.710*****
frontier States (West & South-West)	1,305	97.8	39.54	-.352	**0.483*****

Legend: Table reports the number cities, mean age, standard deviation and skewness of the age distribution of US Census places. Foundation dates are taken from the *Cities Databank*^{TM} and refer to the earliest date of incorporation or self-government of the municipality. The last column reports the correlation between log age (in years as of 2000) and log population size in 2000, controlling for Federal State fixed effects to account for State-specific differences in incorporation laws. The second row refer to cities from the following States: AL, CT, DE, FL, GA, IA, IL, IN, KY, MA, MD, ME, MI, MN, MO, MS, NC, NH, NJ, NY, OH, PA, RI, SC, TN, VA, VT, WI, WV. The third row refer to cities from the following States: AK, AR, AZ, CA, CO, HI, ID, KS, LA, MT, ND, NE, NM, NV, OK, OR, SD, TX, UT, WA, WY.

Figure 2: US city age distribution (kernel density)



Finally, recall from the previous section that a positive correlation between city age and city size plays an important role in the genesis of the DPLN city size distribution. As can be seen in Table 3, this positive correlation is strongly empirically supported both for the traditional and the frontier States.[23] The elasticity of current city size with respect to city age is estimated to be 0.651 for the US urban system as a whole, and is highly statistically significant. That is, doubling the age of a city is – on average – associated with an increase of the city's current population size by about 65%. In the traditional States, the elasticity is a bit higher (0.710) and in the frontier States it is a bit lower (0.483), but in both cases there is a notably positive and highly significant relationship between city age and city size in the data.

4.3 Discussion

Summing up, using the novel US city age data, we find empirical support for several features and predictions of our theoretical model. There is vast age heterogeneity across American cities, and older cities tend to be larger than younger ones. The country's overall distribution of city sizes is, therefore, also affected by the city age profile in the economy. Evidence is more mixed when it comes to the precise functional form of the city age distribution. Our model shows that the DPLN distribution for city sizes emerges when city ages follow a (possibly shifted) exponential distribution. As can be seen in the left panel of Figure 2, such an exponential shape may not perfectly fit the city age data, mainly because of the mass of young cities in the lower tail. However, it can be argued that a shifted exponential still yields a reasonable approximation. The empirical age distribution roughly starts at a minimum city age of around 100 years, and is then clearly right-tailed as indicated in Table 3. Those features are in line with the parameterization of a shifted exponential, which delivers a decent approximation of the city age data particularly for the cities in the mature part of the US urban system (see middle panel of Figure 2).

[23]In the log size-log age regression we have controlled for Federal State fixed effects, in order to take into account State-specific differences in historical incorporation legislations which affect the measured city foundation dates.

One possible strategy could be to simulate a distribution $\widetilde{g}(T)$ that closely matches the city age data, and then to mix age-specific LN size probability distributions as in (5), under the assumption that the mixing parameter T is distributed according to this function $\widetilde{g}(T)$. The disadvantage of such an approach, however, is that an analytical expression for the resulting size distribution $\widetilde{f}(S)$ is then, in general, no longer available and $\widetilde{f}(S)$ can only be obtained via simulation.

The DPLN, by contrast, can be solved in closed form, and it can be readily taken to the data by using standard methods. As shown in Section 2, it achieves a considerable edge over the LN and other parameterizations in terms of data fit to the empirical CSD, yet without being computationally much more difficult to handle. This advantage would disappear when both the age distribution $\widetilde{g}(T)$ and the resulting asymptotic city size distribution $\widetilde{f}(S)$ have to be simulated. Furthermore, the exponential city age distribution has economic foundations: it arises, in a natural way, in an economy where the overall population and the (optimal) number of cities grow at the same rate, as it is the case in our economic model. We therefore believe that our theory-based approach to derive the DPLN distribution for city sizes is more attractive than a pure simulation approach, even if the exponential city age distribution does not fit the empirical age data perfectly.

5 Conclusions

Recently, there has been a lively discussion about city size distributions. Our research can potentially resolve several controversial issues from this literature. First, our results show that the same functional form – the DPLN distribution – closely approximates the empirical city size data, regardless of whether cities are economically or administratively defined. Second, the DPLN unifies the lognormal distribution suggested by Eeckhout (2004) and the Pareto distribution (Zipf's law) advocated by Gabaix (1999), Rozenfeld *et al.* (2011), and many others, in a single framework of an urban system, thereby building a bridge between those two views.

The main aim of this paper was to provide economic foundations where this DPLN distribution of city sizes comes from. One crucial building block is age heterogeneity across cities, which emerges in our model as a growing population allocates over an endogenously determined set of locations. A second important feature is the positive correlation of city age and city size. Finally, the model predicts that cities grow with the same expected rate in the long run (Gibrat's law), but that young cities may grow faster in the beginning. As we show in this paper, these building blocks of the DPLN size distribution are empirically relevant. In particular, using novel data on the foundation dates of American cities, we indeed find strong age differences, and that older cities in the US tend to be larger than younger ones.

References

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

[2] Cuberes, D., 2009. A Model of Sequential City Growth. The B.E. Journal of Macroeconomics 9, Article 18.

[3] Cuberes, D., 2012. Sequential City Growth: Empirical Evidence. Journal of Urban Economics 69, 229-239.

[4] Desmet, K., Rappaport, J., 2012. The Settlement of the United States, 1800 to 2000: The Long Transition Towards Gibrat's Law, Mimeo.

[5] Dobkins, L., Ioannides, Y., 2001. Spatial Interactions among U.S. Cities: 1900-1990. Regional Science and Urban Economics 31, 701-731.

[6] Eeckhout, J., 2004. Gibrat's Law for (All) cities. American Economic Review 94, 1429-1451.

[7] Eeckhout, J., 2009. Gibrat's Law for (All) cities: Reply. American Economic Review 99, 1676-1683.

[8] Fujita, M ., 1978. Spatial Development Planning, Amsterdam: North Holland.

[9] Gabaix, X., 1999. Zipf's Law for Cities: An Explanation. Quarterly Journal of Economics 114, 739-767.

[10] Giesen, K., 2012. Zipf's Law for Cities and the Double Pareto Lognormal Distribution. PhD thesis, University of Duisburg-Essen.

[11] Giesen, K., Suedekum, J., Zimmermann, A., 2010. The Size Distribution Across All Cities - Double Pareto Lognormal Strikes. Journal of Urban Economics 68, 129-137.

[12] Gonzáles-Val, R., Sanz, F., Ramos, A., 2012. Size Distributions for All Cities: Which One is Best?, Working Paper, Institut d'Economia de Barcelona (IEB).

[13] Gonzáles-Val, R., Sánchez-Vidal, M., Viladecans-Marsal, E., 2012. Sequential City Growth in the US: Does Age Matter?, Working Paper,Institut d'Economia de Barcelona (IEB).

[14] Henderson, V., 1974. The Sizes and Types of Cities. American Economic Review 64, 640-656.

[15] Henderson, V., Venables, A., 2009. Dynamics of City Formation. Review of Economic Dynamics 2, 233-254.

[16] Henderson, V., Wang, H., 2007. Urbanization and City Growth: The Role of Institutions. Regional Science and Urban Economics 37, 283-313.

[17] Hsu, Wen-Tai (2012) Central Place Theory and City Size Distribution. Economic Journal 122, 903-932.

[18] Ioannides, Y., Skouras, S., 2013. US City Size Distribution: Robustly Pareto, but Only in the Tail. Journal of Urban Economics 73, 18-29.

[19] Levy, M., 2009. Gibrat's Law for (All) Cities: Comment. American Economic Review 99, 1672-1675.

[20] 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.

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

[22] Reed, W., 2002. On the Rank-Size Distribution for Human Settlements. Journal of Regional Science 42, 1-17.

[23] Reed, W., 2003. The Pareto Law of Incomes - An Explanation and an Extension. Physica A 319, 469 - 486.

[24] Reed, W., Jorgensen, M., 2005. The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distributions. Communications in Statistics 34, 1733-1753.

[25] Rossi-Hansberg, E., Wright, M., 2007. Urban Structure and Growth. Review of Economic Studies 74, 597-624.

[26] Rozenfeld, H., Rybski, D., Gabaix, X., Makse, H., 2011. The Area and Population of Cities: New Insights from a Different Perspective on Cities. American Economic Review 101, 2205-2225.

[27] Rozenfeld, H., Rybski, D., Andrade, J., Batty, M., Stanley, E., Makse, H., 2008. Laws of population growth, Proceedings of the National Academy of Science USA 105, 18702-18707.

[28] 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

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

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

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

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

$$M_{log(S)}(\theta) = \frac{exp\left(s_0\theta + \frac{\sigma_0^2\theta^2}{2}\right)}{\lambda^{-1}\left(\lambda - \left(\gamma - \frac{\varsigma^2}{2}\right)\theta - \frac{\varsigma^2}{2}\theta^2\right)}, \tag{15}$$

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

$$M_{log(S)}(\theta) = exp\left(s_0\theta + \frac{\sigma_0^2\theta^2}{2}\right)\cdot\frac{\alpha\beta}{(\alpha-\theta)(\beta+\theta)}. \tag{16}$$

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 (15) to (16). 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 (16) converges to the mgf of a normal distribution.

BIS Department for Business Innovation & Skills

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Spatial Economics Research Centre (SERC)
London School of Economics
Houghton Street
London WC2A 2AE

Tel: 020 7852 3565
Fax: 020 7955 6848
Web: www.spatialeconomics.ac.uk

SERC is an independent research centre funded by the Economic and Social Research Council (ESRC), Department for Business Innovation and Skills (BIS) and the Welsh Assembly Government.