European Economic Review 71 (2014) 193–208

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

City age and city size

Kristian Giesen [a], Jens Suedekum [b,*]

[a] Mercator School of Management, University of Duisburg-Essen, Germany
[b] Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine-University Düsseldorf, CESifo and IZA. Universitätsstrasse 1, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 25 February 2014
Accepted 30 July 2014
Available online 28 August 2014

JEL classification:
R11
R12
O40

Keywords:
Zipf's law
Gibrat's law
City size distributions
City age
Urban growth

ABSTRACT

Using novel data on the foundation dates of more than 10,000 American Census places, we show that older cities in the US tend to be larger than younger ones. To take this nexus between city age and city size into account, we introduce endogenous city creation into a dynamic economic model of an urban system. Our model predicts a pattern of age-dependent urban growth that is in line with recently established empirical evidence. The size distribution that emerges in our economy delivers a close fit to different types of US data and it outperforms other suggested parametrizations. This evidence can resolve several debates and build a bridge between different views in the literature on city size distributions.

1. Introduction

Ever since the seminal works by Auerbach (1913) and 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 in another fundamental dimension: age. Using novel data on the foundation dates of more than 10,000 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 384 years ago, while places like Laguna Woods (CA) not even had their 14th 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 57%. 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 Gabaix (1999) and Eeckhout (2004) who consider urban systems where *Gibrat's law* is satisfied, that is, 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 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]

* Corresponding author. Tel.: +49 211 81 11622.
E-mail addresses: kristian.giesen@uni-due.de (K. Giesen), suedekum@dice.hhu.de (J. Suedekum).

[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.

http://dx.doi.org/10.1016/j.euroecorev.2014.07.006
0014-2921/

In our model, the country's total population is assumed to grow (at least initially) over time. If the number of cities were fixed, this would lead to rising congestion and decreasing equilibrium utility, as more and more people have to be squeezed into the urban system. We hence allow for the creation of new cities, which enables the population to spread across more and leads to age differences between cities. When a new city is founded, it starts from a randomly drawn initial productivity which may reflect some deep fundamentals of the city's location. Given this initial draw, a new city accordingly adjusts to its equilibrium starting size through population inflows from the established cities. Afterwards, all cities are subject to random shocks which affect the evolution of their population sizes. Since expected city growth is positive, our model then predicts – in line with the aforementioned facts – that older cities tend to be larger than younger ones.

As for the distribution of growth rates across cities, Gibrat's law is at work in our model as all cities grow with the same expected rate in the long run. Yet, there are also deviations: new cities (which tend to be relatively small) exhibit strong population growth rates during the transition towards spatial equilibrium, much higher than in established cities. Young cities thus initially grow faster, but revert to the economy-wide average later on. Such a pattern is consistent with recent empirical evidence on US urban growth over the last two centuries. In particular, the studies by Desmet and Rappaport (2013) and González-Val et al. (2014) 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 tend to grow with the same rate.[2]

From this urban system model with its empirically relevant new features, we are able to derive a closed-form solution for the city size distribution (CSD) that emerges endogenously in our economy. This turns out to be the so-called *double Pareto lognormal* (DPLN) distribution (see Reed, 2002). 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 Rozenfeld et al. (2011) in a single model for the overall CSD. The DPLN distribution delivers a close fit to empirical city size data, both in the US and in various other countries, and it (easily) outperforms the LN, Zipf's law and also other functional forms that have been suggested. It also does so in terms of “adjusted fit”, that is, when penalizing the DPLN for having a more flexible functional form with more free parameters (see Giesen et al., 2010).

The main contribution of this paper is then twofold. First, we derive a micro-founded economic theory to explain *why* DPLN distributed city sizes may emerge.[3] Second, we show with novel city age data that the main new buildings blocks of our model – city age heterogeneity and the positive correlation of age and size – are empirically relevant. Ultimately, we therefore argue that 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. Furthermore, our model predicts a pattern of urban growth that is consistent with recent evidence, namely Gibrat's law with stronger initial growth of young cities.

Finally, another contribution of this paper is that it can potentially settle a controversy from the recent literature on CSDs. That debate deals with the question on how to define a *city* in the first place. 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 parametrization 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 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, when fitting the LN to the area clusters data, one also obtains a very poor fit as is shown in Fig. 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). By contrast, we show that the DPLN distribution closely fits the empirical CSD across all settlements for *both* definitions of US cities (see Fig. 1). 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 also fully in line with, but goes beyond the findings of Rozenfeld et al. (2011): The DPLN is a parametrization for the overall CSD across all clusters that is consistent with their claim that Zipf's law holds among the large clusters.[4]

More generally, the DPLN builds a bridge between the “old” and the “new” literature on city size distributions. It is fully consistent with Zipf's law for large cities, and incorporates this 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 parametrizations. 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,

[2] Michaels et al. (2012) report that, in a comprehensive dataset comprising not only large metropolitan areas, small cities tend to grow faster. These authors do not consider city age, but bearing in mind that small cities are on average younger, our model is in line with that evidence as well.

[3] The stochastic foundations of the DPLN distribution are discussed in Reed and Jorgensen (2005), who show that it emerges by combining a scale-free growth process with a Yule process for the birth of new units. That model is statistical in nature, however, and does not have economic micro-foundations. We provide an *economic* theory for the DPLN distribution of city sizes by extending the seminal approach by Eeckhout (2004) to incorporate endogenous city creation and age heterogeneity across cities.

[4] Relatedly, some authors (most notably Levy, 2009; Ioannides and Skouras, 2013; Malevergne et al., 2011) have argued that the large Census places also follow a Zipfian power law pattern that is only imperfectly captured by the LN parametrization, 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 also be settled by our finding that the sizes of Census places are better approximated by a DPLN distribution.