Mercator School of Management

Lehrstuhl für Mikroökonomik und Außenwirtschaft

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Zipf's Law for Cities and the Double Pareto Lognormal Distribution

MERCATOR SCHOOL OF MANAGEMENT

In Urban Economics and Regional Science there is a long tradition to propose economic models to explain the existence of cities.

Unfortunately, existing models are not in line with reality and there is no common agreement upon what "reality" looks like, especially with respect to the distribution of city sizes. Some researchers argue, that city sizes are distributed according to Zipf's law (see box below) while others argue that city sizes are distributed according to the Lognormal distribution (see other box below). There are several further controversies in that literature and researchers in that field are confused. The contribution of this thesis is to solve those controversies.

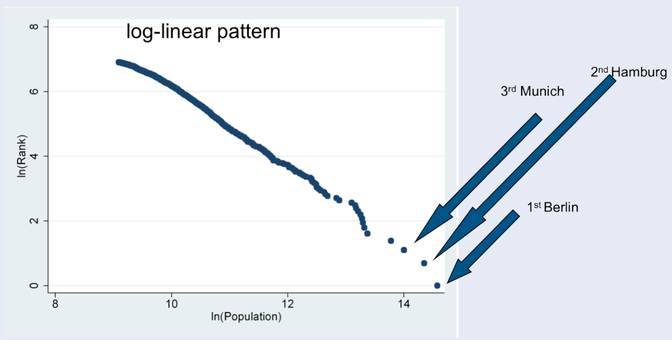
The central questions are:
- How to solve the above mentioned controversies?
- What are the economic forces that shape huge cities like New York City, London or the German „Ruhrgebiet"?
- What is the true city size distribution?
- Why do we have age heterogeneity among cities?

Zipf's Law

Nobel laureate Paul Krugman (1996): "We are unused to seeing regularities this exact in economics. It is so exact that I find it spooky"

Fujita et al. (1999) note: "We must acknowledge that it poses a real intellectual challenge to our understanding of cities [...] nobody has come up with a plausible story about the process that generates Zipf 's law [...]".

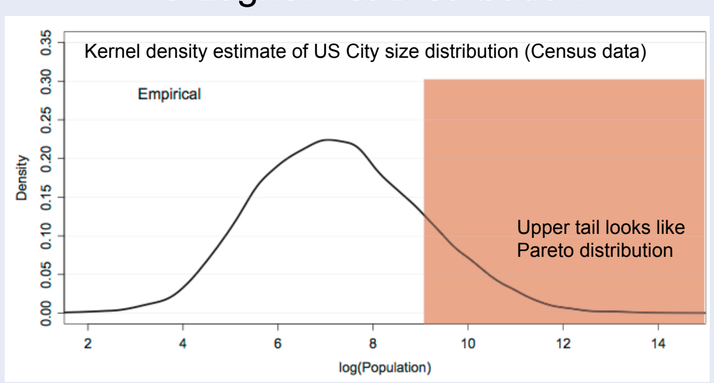
City Name	Population	Rank
Berlin	3 395 189	1
Hamburg	1 743 627	2
München	1 259 677	3
Köln	983 347	4
Frankfurt am Main	651 899	5
Stuttgart	592 569	6
Dortmund	588 168	7
Essen	585 430	8
Düsseldorf	574 514	9



In 1913, the German geographer Felix Auerbach made a considerable observation. He discovered that for the largest cities of a country, the product of their size and their rank in the urban hierarchy roughly equals a constant. This means that a city's size S_i is determined by its rank R_i in the urban hierarchy according to $S_i / S_j = R_j / R_i$. This implies the country's largest city being approximately twice as large as the second largest city, or the third largest city being 5/3 times as large as the fifth largest city etc. Using the insights of probability theory, the rank-size rule means that the largest cities follow a Pareto distribution with a slope coefficient of unity.

Those two statements about city sizes, being Pareto and the unity slope coefficient, have attracted the attention of generations of scientists. Today, this phenomenon is one of the best-studied topics in Urban Economics and Regional Science. The empirical evidence is so overwhelming that it achieved the status of an empirical law: Zipf's Law.

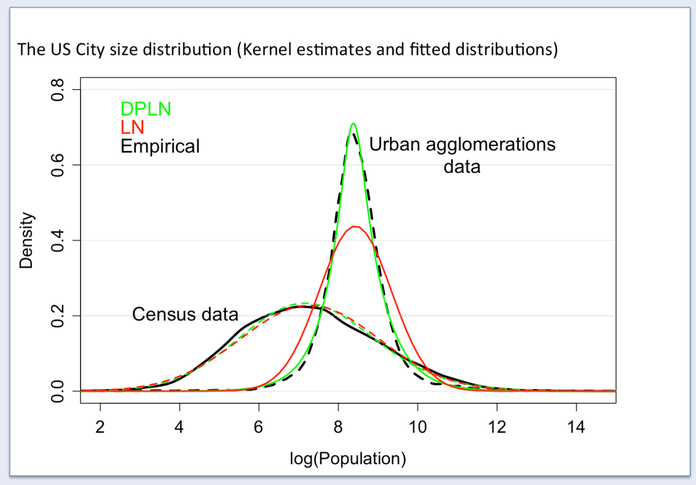
The Lognormal Distribution



Studies around Zipf's law are problematic since they focus only on large cities. In an influential article, Eeckhout (2004) has shown that the Pareto does not hold when taking into account all settlements of a country. He shows, that the overall, untruncated distribution is the Lognormal distribution. This initiated an intensive debate about city size distributions.

The straightforward question is, why so many previous studies provided evidence for a Zipfian power law among large cities? The reason according to Eeckhout (2004) is that the LN and the Pareto distribution have similar properties in the upper tail and can become virtually indistinguishable. In other words, Zipf's law is only an illusion because there is no Pareto distribution in the upper tail.

The Double Pareto Lognormal Distribution



This thesis argues that city sizes follow a more general form, the Double Pareto Lognormal distribution (DPLN).

The DPLN distribution was initially developed by the Canadian statistician and economist William J. Reed (2002). It has the following density for city sizes x:

$$DPLN(x) = \frac{\alpha \cdot \beta}{\alpha + \beta}\left[x^{-\alpha-1} e^{\left(\alpha\cdot\mu+\frac{\alpha^2\cdot\sigma^2}{2}\right)} \Phi\left(\frac{\log(x)-\mu-\alpha\cdot\sigma^2}{\sigma}\right) + x^{\beta-1} e^{\left(\beta\cdot\mu+\frac{\beta^2\cdot\sigma^2}{2}\right)} \Phi^c\left(\frac{\log(x)-\mu-\beta\cdot\sigma^2}{\sigma}\right)\right]$$

The parameters α and β are coeffcients to regulate the tails, whereas μ and σ determine the location and the spread of the distribution. Φ represents the normal cdf and $\Phi^c = 1 - \Phi$ represents the complementary cdf. A special feature of this distribution is that if *x* is large, then $f(x) \sim x^{-\alpha-1}$ and if x is small, then $f(x) \sim x^{\beta-1}$. Furthermore, it nests the LN as a limiting case when $\{\alpha,\beta\} \to \infty$. For other values the body of the distribution is also close to a lognormal shape.

→ Those features enable the DPLN distribution to be consistent with the mature literature on Zipf's law on one side and with Eeckhout's finding of a lognormal body on the other side. This thesis therefore provides a unifiying solution for Zipf's law and the lognormal distribution.

The DPLN should not be thought of as a rigid mixture of LN and two Paretos. It is rather a flexible parameterization that has several distributional features which the LN or the mixture model of LN and Pareto cannot capture.

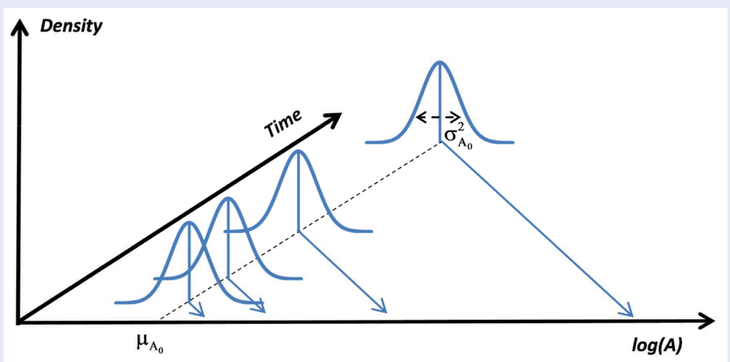
Why should we observe the DPLN in reality?

The DPLN is not an ad-hoc functional form, which is chosen because it has an impressive data fit; instead it has an explicit stochastic foundation. The DPLN results if cities grow according to a stochastic process, which is known as Gibrat's law, and if cities are created in different points of time so that age heterogeneity among cities prevails. This thesis provides empirical evidence for both.

The thesis furthermore provides an economic model of urban growth that naturally replicates this pattern. Within the model cities grow at the intensive and at the extensive margin and a growing population distributes endogenously over a rising number of cities.

While there are many economic models that try to explain residential location choice and city formation, none of those models is in accordance with empirical data on city size distributions. In detail, they cannot explain the body or the upper tail of the distribution. Therefore, the model of this thesis is the only existing model that is able to replicate empirical overall city size distributions.

The model is too complex to be described here in detail. The key features are positive and negative local population externalities along with city specific productivity enhancing amenities. Cities are heterogeneous in their age and grow according to Gibrat's law. Migration is induced by intercity differences in wages, rental prices and commuting costs. Cities are created over time by a central planner, who aims at maximizing national welfare. In the spatial equilibrium, city sizes follow the double Pareto lognormal distribution and the empirical lognormal-shaped overall city size distribution and the well known upper-tail Zipfian power law are features of this model.



Dr. rer. oec. Kristian Giesen, „summa cum laude"

Betreuer: Prof. Dr. Jens Südekum, Lehrstuhl für Mikroökonomik und Außenwirtschaft