CPB Netherlands Bureau for Economic Policy Analysis

CPB Memo

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
Postbus 80510
2508 GM Den Haag

T +31 70 3383 380
I www.cpb.nl

Contact
Abby Ostriker

To: Sector 5 CPB: Climate and Regional Economics

Date: 07-08-2014
Subject: Stochastic growth models for city sizes

1 Introduction

The urbanization of recent decades has caused the fraction of people living in cities to be higher than ever before (World Bank). This process makes the welfare of a country highly dependent on the functioning of its cities. From an economic perspective, policy-makers are concerned with the factors that determine the growth and decline of cities and their functioning in the economy as a whole. In particular, policy-makers are currently interested in studying how cities interact with each other economically. How does the growth of one city affect the growth of its neighbor, or of another city on the other side of the country? What about cities in other countries?

These questions have been partially addressed by a revival of interest in the distribution of city sizes, which can provide valuable information about the growth processes of the cities themselves. Seminal papers in this revival include Rosen and Resnick (1980), Alperovitch (1984, 1993), and Gabaix (1999[a], 1999[b]).

Recent discussion in the literature has centered on a striking phenomenon known as Zipf's Law, which is essentially the observation that the largest cities in a country follow a specific hierarchy in sizes. The law derives its name from Zipf (1949), who found a similar pattern in language, although Auerbach (1913) is generally credited with the seminal observation of the law as it applies to cities. Specifically, Zipf's Law states that a city's population is inversely proportional to its rank, so that the second-largest city is half as large as the largest, the third-largest is a third as large, and so on[1]. Extensive empirical work has verified that Zipf's Law is roughly (though not strictly) followed for all countries (Nitsch 2004; Mulder and de Groot, forthcoming).

[1] Throughout this paper, we will assume that the size of a city equals its population, as is common in this field of research. As is also common, we will use the word "city" to mean an urban agglomeration, whether defined by administrative or economic boundaries. See for a discussion Berry and Okulicz-Kozaryn (2011) and Nitsch (2004).

Any proposed growth process of cities must therefore at least generate a city size distribution conforming to Zipf's Law. An important example of such a growth process is the so-called Gibrat's Law of Proportional Growth, which states that the growth of a city is random and independent of its size. While it seems instinctive that the growth of a city should deterministically depend on economic factors, it is possible that the multitude of factors governing growth could cause the process to be effectively random. Therefore, while economists may feel that it ought to be possible to determine the optimal size of a city by balancing the effects of certain economic forces, such an optimal size may not exist. This is one possible explanation for the pervasively hierarchical nature of city sizes as observed by Zipf's Law.

The existence of Zipf's Law and the degree to which it is related to Gibrat's Law of Proportional Growth has serious implications for policies at all levels of government, from local to international. For example, it might give some countries second thoughts about advocating full integration of the European Union, if such integration could result in the shrinking of their largest city. On the other hand, it could also guide policies of internal regional development in a way that maximizes the growth of a country's cities. Clearly, it is important to study the growth process of cities, and in particular the ways in which that growth process might result in Zipf's Law.

After a brief elaboration on Zipf's and Gibrat's Laws, the contribution and outline of this paper will be described.

1.1 Zipf's Law

Mathematically, Zipf's Law can be expressed as

$$\boldsymbol{R = A\, P^{-\alpha}} \qquad \textbf{Equation 1.1}$$

where R is a city's rank, P is its population, and A and α are constants where A represents the population in the largest city. For Zipf's Law to hold strictly, α should be equal to 1. However, there is variation in how well countries fit this description. Empirical observations find that α usually lies between 0.8 and 1.2 (Nitsch 2004, Mulder and de Groot, forthcoming).

To see how closely a given country follows this rule, researchers in this field use one of two techniques to find the empirical value of α. The most common strategy is linear regression, due to its simplicity and flexibility. This approach is based on the observation that the logarithm of Equation 1.1 produces a straight line with a slope equal to the negative of the variable α:

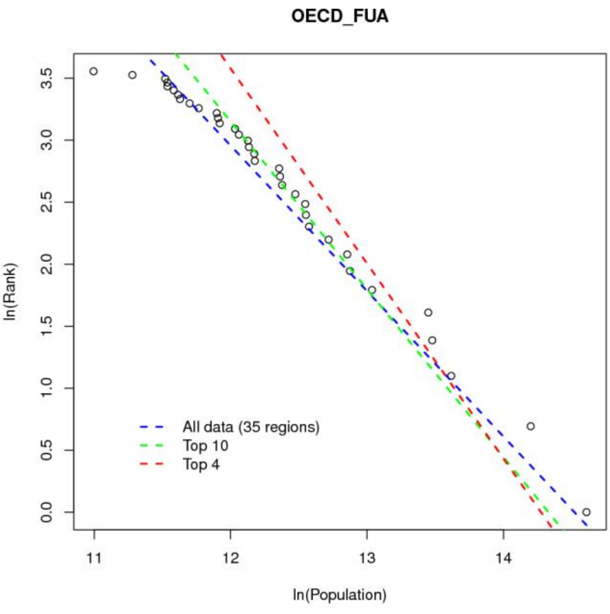
$$\ln(\boldsymbol{R}) = \ln(\boldsymbol{A}) - \boldsymbol{\alpha} \ln(\boldsymbol{P}).$$ **Equation 1.2**

Researchers will fit a line to the scatter plot of log rank versus log size for (some subset of) a country's cities and then estimate α as the slope of that line. Because it functions as a coefficient in this linear relationship, α is frequently known as the Zipf coefficient of a collection of cities. In order to get an unbiased estimate of the Zipf coefficient, a half should be subtracted from the rank in the equation above (Gabaix and Ibragimov, 2011). While this strategy is mathematically very robust, its estimates are highly reliant on both the number of cities considered[2] and the choice of agglomeration used to represent a city[3] . Furthermore, although Zipf's Law usually only applies to some subset of the largest cities in a country, there is no consensus on the appropriate amount of the upper tail to include in the measurement.

The other commonly used technique is known as Maximum Likelihood Estimation, in which researchers assume that the city size distribution follows a power law and find the most probable exponent given the data. This can be more accurate than the linear regression technique, but only if the underlying distribution is truly a power law (Gonzalez-Val 2012). Otherwise, it will be biased.

As an example we estimated the Zipf coefficient with a linear regression using the Dutch Functional Urban Areas (OECD, 2012). The results are shown in figure 1.1.

Figure 1.1 Estimations of the Zipf Coefficient using a linear regression with the rank minus 1/2 adjustment on Dutch Functional Urban Areas.



We find a Zipf coefficient of 1.17 for all data, 1.35 for the ten largest cities and 1.57 for the four largest cities. Although the data appear linear, the slope of the estimated line changes when using a different number of cities in the estimation.

---

2 It is a well-observed fact that in most cases, including more cities causes the Zipf coefficient to decrease (Gonzalez-Val 2012).

3 The two basic types of agglomeration units are administratively-defined cities and larger, economically-defined areas. It is typically easier to obtain data for administratively-defined cities, but it is more natural to use economically-defined areas. Estimates of the Zipf coefficient tend to be smaller and closer to 1 when agglomerations are larger and follow economic instead of administrative boundaries (Nitsch 2004; see for example Rozenfeld et al 2010).

Mathematically, Zipf's Law can be interpreted as a survival function (one minus the cumulative distribution function, or CDF) because the left-hand side of Equation 1 -- the rank of a city -- represents the number of cities that are at least as large as the city under consideration (Adamic, 2002). The expression of Zipf's Law in terms of a survival function is known as Pareto's Law, and it states that the probability that any city size X is bigger than a given size x is

$$\boldsymbol{P}(\boldsymbol{X} > \boldsymbol{x}) \sim \boldsymbol{x}^{-\boldsymbol{\alpha}} \, . \qquad \textbf{Equation 1.3}$$

Because of this formulation, α is also known as a Pareto exponent. From this equation, it is easy to see that the probability that a given city size X is a certain size *x* is given by the equation:

$$\boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{\alpha x}^{-\boldsymbol{\alpha}-\boldsymbol{1}}. \qquad \textbf{Equation 1.4}$$

This Pareto distribution is also known as a power law distribution, because *x* is raised to a power.

When X is Pareto-distributed with a minimum size x_m, the equations change slightly. The probability that any city size X is bigger than a given size *x* is

$$\boldsymbol{P}(\boldsymbol{X} > \boldsymbol{x}) \sim \begin{cases} \left(\frac{\boldsymbol{x_m}}{\boldsymbol{x}}\right)^{\boldsymbol{\alpha}} & \text{for } X \geq x_m \\ 1 & \text{for } X < x_m, \end{cases} \qquad \textbf{Equation 1.5}$$

and the probability that a given city size X is a certain size *x* is

$$\boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = \begin{cases} \boldsymbol{\alpha}\left(\frac{\boldsymbol{x_m}}{\boldsymbol{x}}\right)^{\boldsymbol{\alpha}+\boldsymbol{1}} & \text{for } X \geq x_m \\ 0 & \text{for } X < x_m. \end{cases} \qquad \textbf{Equation 1.6}$$

We will use this final equation later in the paper to "predict" a city size distribution from a previous distribution.

1.2 Gibrat's Law

In order to form a theoretical basis to help explain Zipf's Law, many researchers rely on an assumption first introduced by Gibrat in his 1931 paper on the topic. In order to make the problem more tractable, Gibrat assumed that city growth is independent of size. While this feature was just a condition required for his proof to be valid, it came to be known as Gibrat's Law of Proportional Growth and is included -- either in its original form or in some modified version -- in virtually every model of growth associated with Zipf's Law. Although applying Gibrat's Law with no additional constraints does not result in the power-law (perfect Zipf) distribution, the addition of one or two conditions is typically enough to generate one (de Wit, 2005). Furthermore, Gibrat's Law on its own produces a lognormal distribution, whose upper tail is almost impossible to distinguish from that of a Zipf distribution. Finally, while the distributions of economically-defined urban areas and the upper tails of the distributions of administratively-defined cities are usually fit well by a power law, the entire distribution of administratively-defined cities typically appears to be lognormal (Eeckhout, 2004). Therefore, Gibrat's lognormal distribution is a valuable model in its own right.

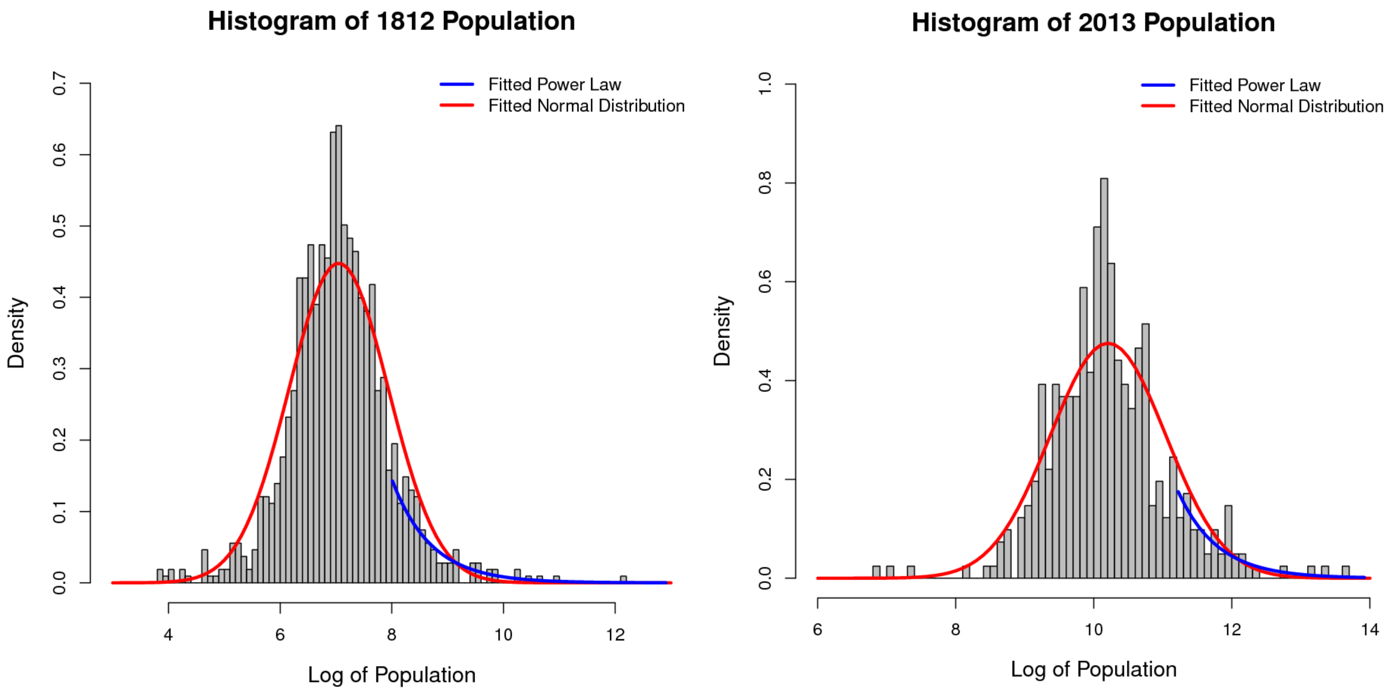
Although Gibrat's Law produces a theoretical distribution which is consistent with empirical findings, the validity of the law itself is contested empirically. The scientific consensus seems to be that Gibrat's Law is generally but not strictly true (Eeckhout 2004, Ioannides and Overman 2003, Black and Henderson 2003). Furthermore, it is not theoretically obvious why Gibrat's Law should hold true, as intuitively one expects the growth rates of smaller cities to have a higher standard deviation than those of more stable larger cities. Several economic models have been proposed to satisfy Gibrat's Law, most of them focusing on balancing positive and negative externalities, but none of them have gained much traction yet[4].

A variety of stochastic and economic growth models have been proposed which attempt to generate a city size distribution which satisfies Gibrat's Law. In this paper, we will only focus on the two most widely-cited stochastic growth models in the literature: one by Gibrat (1931), which produces a Gibrat (lognormal) distribution, and one which produces a Pareto distribution and whose continuous and discrete versions were produced independently by Gabaix (1999) and teams led by Solomon (Levy and Solomon 1996; Malcai, Biham and Solomon 1999; Blank and Solomon 2000). Gibrat's model arises directly from the Gibrat's Law assumption, while the Gabaix-Solomon model results from the combination of Gibrat's Law and a lower boundary on city size[5]. Note that there is sometimes confusion between Gibrat's Law and Gibrat's distribution. Gibrat's Law is the phenomenon of growth independent from size, while Gibrat's Distribution arises from Gibrat's Law when no other restrictions are added.

For a visual comparison of the Gibrat (lognormal[6]) and Gabaix-Solomon (power law) distributions fitted to a histogram of the *natural log* of Dutch city sizes, see Figure 1.2. Note that the lognormal distribution (red; becomes lognormal when fitted to a log plot) provides a good fit for the entire distribution, while the power law distribution (blue) which represents a perfect Zipf's Law fit, only fits the upper tail.

[4] Stochastic models include Gibrat, 1931; Kalecki, 1945; Levy and Solomon, 1996; Gabaix, 1999. Economic models include Duranton 2006, 2007; Eeckhout 2004; Rossi-Hansberg and Wright 2007; Cordoba 2008a,b.
[5] Other derivations of Zipf's Law from Gibrat's Law do exist. For example, one can use a Kesten process (Gabaix, 2009), or include the emergence and disappearance of cities (Blank and Solomon 2000; Saitchev, Malevergne and Sornette 2009).
[6] A distribution whose logarithm is normally distributed is known as the lognormal distribution (see Appendix A).

Figure 1.2 Comparison of fits of the lognormal distribution and the Pareto distribution based on the models of Gibrat and Gabaix-Solomon, respectively, for Dutch municipalities in 1812 and 2013.



1.3 Contribution of This Paper

Although the Gibrat and Gabaix-Solomon models described here are also introduced in the literature, authors are often not aware of errors and hidden assumptions in their seminal derivations. The ensuing scientific discussion then diffuses the corrections across the literature, making it useful to collect and present the final versions of each model in one place.

To this end, this paper will provide a clear and thorough explanation of the mathematical basis of the discrete versions of two fundamental stochastic growth models explaining the city size distribution[7]. In addition, we will identify the key assumptions of each of the models and discuss the economic and empiric justification behind them. Finally, this paper will evaluate how useful these stochastic growth models are for predictions and explanations of city sizes.

[7] Note that we will consider only the discrete versions of these two models, both because it is more natural for cities than a continuous model due to the finite number of cities, and because it avoids the introduction of additional assumptions required for a continuous process.

The structure of this paper is as follows: We will introduce the Gibrat distribution and assess its assumptions in the context of city sizes. We will then introduce the Gabaix-Solomon model and assess its assumptions in the same context. Finally, we will use computer simulations of these two models to conduct two tests. First, we will test whether we have correctly identified the assumptions in each of the models. Then, we will test whether it is possible to generate the distribution of municipalities in the Netherlands in 2000 on the basis of population data from Dutch municipalities from the years 1980-1990.

2 Discrete Gibrat Distribution

There is a large body of work that aims to describe city size distribution by pure randomness. A key starting point of many of these models is the Gibrat Distribution, which was one of the first city size distribution models to be proposed. This distribution is the result of imposing Gibrat's Law on the process of random proportionate growth. Gibrat first derived the result, a lognormal distribution with ever-increasing variance, in 1931, but a clear and detailed explanation of the proof is provided by Kalecki (1945), and we follow this derivation here. In order to better understand the various theoretical requirements for the derivation, we will outline the necessary assumptions at the beginning of each section. At the end of the derivation, we will discuss the empirical validity of each of these assumptions.

We begin by imposing one basic assumption that greatly simplifies the proof:

Assumption 1: Cities do not appear or disappear.

We also impose the assumption that a city's growth is random and independent of time:

Assumption 2: The size of a city grows or shrinks proportionately by a series of random and time-independent shocks.

Call these shocks m_t. When we are only describing the distribution of shocks for one city, it is sufficient that these shocks be independent of time. When we consider a collection of cities, it will also be important that each city's shocks are independent of all other cities' shocks. In this situation, it will also be necessary to combine the individual distributions of each city into one distribution over all the cities. Therefore, at every time step each city's distribution of shocks should have the same mean and variance[8]. This leads to the next assumption, which imposes Gibrat's Law on the process:

Assumption 3: At every time t, all cities draw their shocks randomly and independently from one distribution with finite mean μ_t and variance σ^2_t.

Because all cities draw their shocks randomly and independently from the same distribution, the shocks (and therefore the growth rates) are independent of size. To conclude our description of the shocks, we assume that they are small relative to one:

Assumption 4: The absolute value of every random variable m_t is small relative to 1.

Now, let X_t represent the size of a given city at time t. At every time t, the city size will be updated by the random process

$X_{t+1} = X_t*(1+m_t)$. **Equation 2.1**

Therefore, after T time steps (typically years), the value of X_T will be

$X_T = X_0*(1+m_1)*(1+m_2)*\ldots*(1+m_T)$ **Equation 2.2**

where X_0 is the initial value of X.

It can be shown that in the case of a single city, the distribution of $\log(X_T)$will approach the normal distribution as T approaches infinity. The intuitive explanation behind this result is that taking the logarithm of X breaks it into a sum of small values. The size of each new value becomes small relative to the standard deviation of the sum, so by a special version of the Central Limit Theorem[9] the distribution of the values of $\log(X_T)$ will approach a normal distribution.

As an extension of this result, we can consider a collection of cities that satisfy the above properties, and impose the additional assumptions that their initial sizes are expected to be the same and that the variation in their initial sizes will also become small relative to the total variation in size. Then we can show that the distribution of

[8] Note that this shared mean and variance can vary with time, as long as this variation is independent of time -- for example, the mean or variance could be drawn randomly at each time period from a distribution which does not vary with time.

[9] The version of the Central Limit Theorem used for this proof is known as the Laplace-Lyapunov Central Limit Theorem. All limit theorems describe the conditions under which a sum of *n* independent (or weakly dependent) random variables has a probability distribution approaching a normal distribution. In the Laplace-Lyapunov version, the random variables must be independent, with a finite mean and variance, but not necessarily identically distributed. Moreover, the third moment of the sum must go to 0 as *n* approaches infinity. This condition is equivalent to the condition that the normalized random variables become asymptotically negligible, and that the maximum value of any term in the sequence becomes vanishingly small compared to the whole sum. That these conditions hold under the assumptions of the Gibrat distribution is demonstrated in this proof. (Sources: http://en.wikipedia.org/wiki/Central_limit_theorem, http://www.encyclopediaofmath.org/index.php/Central_limit_theorem)

city sizes over the whole collection of cities also)] approaches the lognormal distribution as T approaches infinity. This will be our final result.

2.1 Single City Case

In this section we will prove that the city size distribution of a single city will approach a lognormal distribution as the number of time periods becomes sufficiently large. To formalize this mathematically, we first take the logarithm of equation 1:

$\ln(X_T) = \ln(X_0) + \ln(1+m_1) + \ln(1+m_2) + \ldots + \ln(1+m_T)$ **Equation 2.3**

In order to use the Central Limit Theorem, we must prove that the individual terms of this expression become small relative to the total variance of this expression, interpreted as the variance of the distribution of log of the city's size after T time steps[10]. We therefore denote the differences between the observed natural log of city sizes and growth rates and their respective means as:

$Y_T = \log(X_T) - E[\log(X_T)]$
$Y_0 = \log(X_0) - E(\log(X_0)]$
$y_t = \ln(1+m_t) - E[\ln(1+m_t)]$

We can now describe the total deviation of the log of a city's size from the log of its expected size:

$Y_T = Y_0 + y_1 + y_2 + \ldots + y_T$ **Equation 2.4**

Note that for a single city, Y_0 is equal to 0 (the expectation of the initial size is always exactly equal to the initial size when the expectation is generated after the fact). However, Y_0 will be non-zero in the case of a collection of cities, because not every city will have an initial size equal to the expectation.

Also note that as a corollary of Assumption 4 that every value m_t is small relative to 1, we can assume that $\log(1+m_t)$ and therefore also y_t are both small relative to 1 for every value m_t. Therefore, for the individual additional terms to become small relative to the variance, we need the variance to eventually exceed 1.

[10] It may seem odd to think about a probability distribution for an event which can only happen once, such as the size of a city in a given year. However, the idea is that if one were able to follow the city through time over and over, its size at time T could take on a range of values which would be described by the city size distribution we are deriving. It would be like performing many trials of an experiment in which you add together the values from T rolls of a die. Since the rolls are independent, by the Central Limit Theorem you will eventually see that the sum is normally distributed.

However, this is easy, because due to Assumption 2 that growth rates are independent, we know that the variance of the sum of Equation 2.4 is equal to the sum of the variances of each value at every time step:

$$Var\left(\sum_{t=1}^{T} y_t\right) = \sum_{t=1}^{T} \sigma^2(y_t)$$

and that therefore the standard deviation grows continually with each time step. To ensure that the variance does in fact eventually exceed 1, we finally assume[11]:

Assumption 5: $\sigma^2(y_T)$ does not approach 0 as T approaches infinity.

Then, by the Laplace-Lyapunov Central Limit Theorem, the distribution of deviations from the mean $Y_T = 0 + y_1 + y_2 + \ldots + y_T$ will be approximately normal for sufficiently large T. Recalling that the expectation of y_T is zero for all cities, and that at time t the city has the variance in size $\sigma^2(y_t)$, we see that Y_T for a single city will approach the normal distribution $N(0, \sum_{t=1}^{T} \sigma^2(y_t))$. The natural log of X_T will therefore also have a normal distribution with the same variance. The mean of this distribution will be determined by the expected initial size and expected growth rate at every time t: $E[\ln(X_0)]+\sum_{t=1}^{T} E[\ln(1+m_t)]$.

We therefore find that the distribution of the natural log of X_T is normal, with an expected mean value of $E[\ln(X_0)]+\sum_{t=1}^{T} E[\ln(1+m_t)]$ and an expected variance of $\sum_{t=1}^{T} \sigma^2(y_t))$ at time T. Then, by definition, we see that X_T is by definition *lognormally* distributed. Thus we have proved our result for the case of the single city.

2.2 Multiple City Case

We can now consider a collection of such cities, which all share the same expected value of shock μ_t and variance in shocks σ^2_t (Assumption 3), and which moreover all share the same expected initial size X_0:

Assumption 6: Each city's expected initial size X_0 is the same.

Now that we are considering multiple cities, let us use the superscript *i* to denote the value of a variable for an individual city. Note that in this case, because there is some variation in the initial city size X_0^i, the difference between the observed and expected log of city size for each city, Y_0^i, is not necessarily equal to 0. However, we assume that for each city, this variation is small:

[11] This assumption might be too strong, but theoretically it should not be very important. Kalecki says the variance needs to not "fall below a certain level", but the main priority is that it eventually needs to exceed 1.

Assumption 7: The standard deviation of Y_0^i is small relative to the sum of the standard deviations of the values $y_1^i + y_2^i + \ldots + y_T^i$.

Then, for each city, we can apply the Laplace-Lyapunov Central Limit Theorem just as we did in the single-city case, and observe that each city will approach a normal distribution independent of the distribution of Y_0^i.

By Assumptions 3 and 6, we know that every city has the same expected initial size and expected mean and variance in growth rates at every time t. Therefore, as T approaches infinity, the distributions can be combined.

2.2.1 Result

If a collection of cities with no appearance or disappearance grows by a series of random and independent shocks which become small relative to the sum of their variances as time goes on, and if at each time step the distribution of shocks has the same mean and variance for all cities, and furthermore if the initial size of each city is in expectation very similar and has a small variance, then the distribution of $\ln(X_T)$ approaches the normal distribution

$$\mathbf{N}(\mathbf{E}[\ln(\mathbf{X_0})] + \sum_{t=1}^{T} \boldsymbol{E}[\ln(\mathbf{1} + \boldsymbol{m_t})] \; , \; \sum_{t=1}^{T} \boldsymbol{\sigma^2}(\boldsymbol{y_t})\,). \qquad \text{Equation 2.5}$$

2.3 Discussion of Assumptions

2.3.1 Validity of Assumptions 3 and 4

It is encouraging that several of the assumptions on which this result relies are plausible. Assumption 4 that growth rates are small relative to 1 is validated empirically, as growth rates are usually centered close to 0 with a very small variance (Ioannides and Overman, 2003). Assumption 3 that the mean and variance of the growth rates are the same for all cities is still being debated, but it is supported by some empirical evidence. For example, Eeckhout (2004) finds that the mean and variation of growth rates is the same across all but the lowest decile of American “legal” cities. Ioannides and Overman (2003) also find that Gibrat’s Law holds broadly, though not exactly, for the growth processes of administratively-defined cities in the United States. However, for economically-defined regions in the US and/or Great Britain, Black and Henderson (2003) and Rozenfeld et al (2008) find that Gibrat’s Law does not hold. Specifically, Rozenfeld et al find that both the mean and standard deviation of the growth rate decrease with size.

2.3.2 Validity of Assumptions 1, 6, and 7

There are also some concerns about the validity of some of the other assumptions. The most restrictive assumption is the first, which states that cities neither appear nor disappear. In fact, cities certainly do both over time. For example, Almere, the newest city in the Netherlands, only formed in 1984 after a large amount of land near Amsterdam was drained. The related assumptions (6 and 7) that all cities have the same expectation and variation in starting size are also unrealistic, although they were imposed for technical reasons which may not be necessary should Assumption 1 be relaxed. Since cities are typically born at small sizes, including city births would increase the number of cities in the lower tail, making the distribution closer to a power law. On the other hand, since cities also exit from the lower tail, the effect of births would be somewhat mitigated. The introduction of births and deaths is discussed in Blank and Solomon (2000).

2.3.3 Validity of Assumption 2

The second very restrictive assumption is that the shocks are time-independent. Because of the division of population data into time scales of a year, growth rates that are close in time might be heavily auto-correlated -- if, for example, a city is experiencing a recession. The existence of auto-correlation reduces the predictive efficacy of the model, because a city's growth might be more accurately predicted by its growth in the previous year than by the mean growth of other cities.

2.3.4 Assumption 5 and Result

Finally, there are some concerns about the result, which is not a steady-state distribution. Note that by Assumption 5, the variance of the distribution of differences from the mean is continually increasing. Although this distribution of differences (and consequently the distribution of the log of city sizes) may be normal, due to its ever-increasing variance the city size will become indeterminate as time approaches infinity. Furthermore, as is discussed in Appendix A, if the variance in the log of city size increases with time, most of the cities will eventually become very small relative to a few extremely large cities. This is an unattractive feature of the model.

2.4 Conclusion

Gibrat's Law that a city's growth is independent of its size allows us to build a framework for showing that the distribution of city sizes is lognormal. However, this framework has a few structural concerns -- most importantly, the problematic assumptions that no cities ever appear or disappear and that the growth is independent of time.

3 Gabaix-Solomon Approach

In order to find a city size distribution that does not degenerate with time, researchers have imposed various additional conditions to the basic stochastic model. In general, these conditions either modify Gibrat's Law, so that growth rate decreases with increasing size; remove the condition that cities cannot enter or exit; or impose a lower size barrier that cities cannot cross. Imposing this last condition produces a power law distribution, which some have argued describes the upper tail[12] of observed city size distributions well (Levy 2009; Malevergne et al 2009, 2011).

This steady-state power-law distribution has been derived using two approaches. Gabaix (1999) generated a power-law distribution by employing a continuous-time approach, while Malcai, Biham, and Solomon (1999) independently derived a similar result using a discrete-time approach. Because a continuous distribution of city sizes requires there to be an infinite number of cities and an infinitely large maximum city size (Gabaix 2009, p. 268), which is a problematic assumption, we will focus on the derivation provided by Malcai et al. (1999). In the following section, we will outline the assumptions and mathematical procedures required to get their result of a steady-state power-law distribution.

3.1 Derivation of Steady-State Power Law Distribution

Malcai et al. derive a steady-state city size distribution by normalizing the sizes to account for a non-stationary growth rate which would cause the distribution to grow or shrink in absolute size, and then by imposing a lower bound[13]. By normalizing the city sizes, they are able to show that the probability distribution follows a power law in the steady state, and they determine the value of the power law exponent based on the number of cities and the lower limit. The outline of the proof, and the assumptions that are required for it, are explained below. A more detailed derivation of the result is provided in Appendix B.

The proof begins with the now-familiar stochastic system of random growth that is independent of size. For now, we will maintain the assumption that there is no appearance or disappearance of cities.

Assumption 1: Cities do not appear or disappear.

Our second assumption is also unchanged from the Gibrat distribution case:

---

[12] Most scholars of the subject agree that for administratively-defined municipalities (as opposed to economically-defined urban areas), the lower tail of observed distributions is well modeled by a lognormal distribution.
[13] The Gibrat distribution cannot be modified to produce a steady-state distribution by simply normalizing the city sizes, because its variance still increases monotonically with time.

Assumption 2: The size of a city grows or shrinks proportionately by a series of random and time-independent shocks.

The third assumption is similar, but not identical, to the third assumption in the first section. While before, we assumed that the mean and variance of the growth shocks could be time-dependent, for ease of calculation we will here assume they are not.

Assumption 3: At every time t, cities[14] draw shocks randomly and independently from one distribution with finite mean μ and variance σ^2.

However, the derived results are also valid for the more general case of time specific shocks (Blank and Solomon, 2000). Next, we impose the new assumption that there is a lower bound on city size.

Assumption 4: City sizes do not fall below a given (time-independent) fraction of the average city size of the previous period.

We can formulate this mathematically using the notation of Malcai et al:

Let $w_i(t)$be the size of a city i = 1 ... N at time t, and let $\lambda(t)$ be the multiplication factor experienced by all cities. By Assumptions 2 and 3, at each time t, all of the N cities undergo growth by the random process

$$\boldsymbol{w_i(t+1) \;=\; \lambda(t)\, w_i(t)\,.} \qquad \textbf{Equation 3.1}$$

The distribution of the growth shocks λ is denoted as $\Pi(\lambda)$ and is independent of both i and t (Assumption 2)[15]. Note that the multiplication factor λ is equivalent to the growth rate 1+m in the Gibrat approach discussed in Section 2. Also note that since city size cannot be negative, λ is positive:

Assumption 5: λ is always greater than 0.

We denote by $\hat{w}$ the average value of city size at time t:

$$\hat{w}(\mathbf{t}) \;=\; \boldsymbol{\frac{1}{N}\sum_{i=1}^{N} w_i(t)} \qquad \textbf{Equation 3.2}$$

[14] Malcai et al. (1999) actually use an asynchronous updating method, but since the growth rate distribution is independent of time and city size it should not make any difference.

[15] Blank and Solomon observe that for a realistic distribution, the mean growth rate should be slightly higher than 1, to allow growth, but with a standard deviation that allows randomly selected growth rates to drop below 1 - so that the population is not monotonically increasing. Their cited values of a mean of 1.02 and standard deviation of 0.05 agree with the corresponding values of 1.013 and 0.03 as calculated from a large dataset of the population of Dutch municipalities.

City sizes are then constrained to be greater than or equal to the lower bound $c{\cdot}\hat{w}(t)$ by the mechanism:

$$\boldsymbol{w_i(t+1)} = \mathbf{max}\{\boldsymbol{w_i(t+1)}, \boldsymbol{c} \cdot \hat{\mathrm{w}}(\boldsymbol{t})\}$$ **Equation 3.3**

where ŵ(t), evaluated just before the application of Equation 1, is used.

No new assumptions are required to reach the final result, which is that there exists a steady-state distribution of city sizes that follows a power law with an exponent denoted by α. If the minimum (normalized) city size is *c* and the maximum (normalized) city size is N[16], then the power law exponent α is implicitly defined by the expression

$$N = \frac{\alpha - 1}{\alpha}\left[\frac{\left(\frac{c}{N}\right)^{\alpha} - 1}{\left(\frac{c}{N}\right)^{\alpha} - \frac{c}{N}}\right]$$

Equation 3.4

It can be shown that various conditions on N and *c* generate values for α within the observed experimental range. Further details of the proof can be found in Appendix B.

Note that Gabaix's (1999) approach, with an infinitely large maximum (normalized) city size, generates a simplified version of the above expression which reduces to

$$\alpha = \frac{1}{1 - \frac{c}{N}}$$

Equation 3.5

As N approaches infinity, the value of the exponent approaches 1, thereby generating Zipf's Law. However, as previously discussed, this result is not valid for the discrete case.

3.2 Analysis of Assumptions

[16] As Gabaix (2009) points out, this upper limit is somewhat arbitrary, and a more logical choice might be the true maximum value, N - (N-1)*c. However, this does not affect the validity of the proof and only negligibly affects the value of the power law exponent α for typical values of N and c.

Fewer assumptions are required to derive this steady-state result than are required to derive the Gibrat distribution, but those that remain can still be problematic.

3.2.1 Assumptions 1 and 4

Assumption 1 that cities do not form or decay presents the same concerns that are detailed in Section 1. These concerns are related to the concerns that appear with Assumption 4 that cities do not fall below a certain minimum size, because a city that disappears clearly falls to size 0. Furthermore, the idea that city sizes might be maintained at some lower bound by an exogenous force is economically quite unreasonable -- governments do not "supplement" city populations in the same way that they supplement personal incomes below a certain level. However, Blank and Solomon (2000) show that by relaxing both of these assumptions simultaneously, a realistic model can be achieved.

3.2.2 Assumptions 2 and 3

Assumption 2 that the distribution of shocks is time-independent presents the same concerns as described in Section 1. Another concern with the model is raised by Skouras (2010), who points out that introducing a lower barrier (Assumption 4) violates the assumption that growth rates are independently and identically distributed (Assumption 3). Because growth rates are adjusted when a city size is about to fall below a certain limit, the rates become effectively dependent on size. As a result, the parameters of the resulting Pareto distribution become dependent on the parameters of the growth rate distribution (which was not the case in the Gabaix-Solomon derivation). It is then apparent that very specific conditions on the growth rate distribution are required to generate a strict Zipf's Law (with $\alpha=1$) in the upper tail.

3.2.3 Initial City Size

Finally, note that, as with the Gibrat distribution, the shape of the initial city size distribution is not relevant to the shape of the eventual distribution. However, it is still true that a sufficiently long time needs to elapse before the distribution has its final shape (lognormal in the Gibrat case; following a power law in the Gabaix-Solomon case.)

4 Stochastic Simulations

To verify that the assumptions we isolated in the models above are both necessary and sufficient to generate the distributions they claim, we performed several computer simulations of the Gibrat and Gabaix-Solomon stochastic processes.

4.1 Methods

In the computer simulations, it was possible to specify a variety of parameters, including:

- model (Gibrat or Gabaix-Solomon)
- lower limit on size (for Gabaix-Solomon model only)
- number of cities
- initial distribution of city sizes (a distribution's shape, mean, and variance, or empirical distributions from a Dutch municipality dataset)
- distribution of growth rates (shape, mean, and variance)
- length (number of time steps) of each trial

For the Gabaix-Solomon model, a value had to be chosen for c, which determines the lower bound on size. We used Equation 4 and values of N and α from the Dutch Municipality data to find a value of c in the region of 0.4. Interestingly, this value was actually higher than the largest ratio for a given year of the smallest existing city size to the mean city size, which was 0.09. For N=500, reducing c from 0.40 to 0.09 reduces the predicted exponent from 1.66 to 0.93. The most likely explanation for this discrepancy is that the observed distributions are not strictly Pareto (see Figure 1.2), and they have not converged to a steady state. Therefore, for consistency, we chose to use a higher value of c in the simulations (the exact value was not important to the results, so we typically used c=0.34, which is the value derived from the number and Zipf coefficient of the top 20% of Dutch municipalities in 1812).

Following these specifications, the program then simulated the development of city sizes according to the assumptions of the given model. At each time step, growth rates were randomly and independently generated for each city, with the constraint that no multiplication factor could be less than or equal to 0 (therefore following Assumptions 1, 2, and 3). In the Gabaix-Solomon model, we added to the Gibrat model the constraint that each city be assigned the maximum of their stochastically generated size and the lower limit, calculated as described above (Assumption 4).

For some simulations, we were interested in determining the Zipf coefficient of a distribution. Since Zipf's Law usually applies only to the upper tail of city size distributions, in most cases we calculated the coefficient by performing a linear regression of log rank versus log population for the top 10% or 20% of cities[17]. However, in some cases (such as when we ran the Gabaix-Solomon model to convergence) we believed the power law would hold throughout the distribution, so we calculated the coefficient using the Hill (1975) maximum likelihood estimator, following Gabaix (2009):

$$\alpha = \frac{n-2}{\sum_{i=n}^{n-1} \ln(w_i) - \ln(w_{min})}$$

[17] In fact, we followed Gabaix and Ibragimov (2011) in performing the regression against log (rank - 1/2), in order to reduce bias.

Equation 4.1

4.2 Results

After simulating the development of city sizes, we were able to confirm that Gibrat's model leads to a lognormal distribution, which appears normal in a histogram of log sizes (Figure 4.1), while the Gabaix-Solomon model leads to a power law distribution (Figure 4.2).

4.2.1 Shapes of Final Distribution Match Predictions

For both the Gibrat and the Gabaix-Solomon models, we compared the average final distribution with the predicted final distribution, and confirmed that the observed parameters of mean and variance, in the first case, and power law exponent, in the second, both matched the predictions made by the model. This was verified graphically by plotting the expected distributions on the same graph as the histogram of the average distribution after the simulation was run for an appropriately long time period (see Figs. 4.1, 4.2, and 4.3).

The expected lognormal distribution for the Gibrat model was calculated based on the formula in Equation 2.5. Note that just as predicted, the variance in the initial size distribution, which was on the order of 3e06, became small relative to the total variance and therefore did not need to be included in the prediction of the variance of the final size distribution.

The expected Pareto distribution for the Gabaix-Solomon model had to be derived for use on the distribution of the natural log of the population. According to Equation 1.5, the CDF for the Pareto Distribution with a minimum value x_m is

$$P(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^{-\alpha} & \text{for } X \geq x_m \\ 1 & \text{for } X < x_m, \end{cases}$$

Equation 4.2

We introduce a variable Y = ln(X) and substitute it into the equation:

$$P(Y > y) = P(\ln(X) > y) = P(X > e^y) = \left(\frac{x_m}{e^y}\right)^{\alpha}$$

Equation 4.3

Now, we can take the derivative to find the PDF:

$$P(\ln(X) = y) = \alpha\left(\frac{x_m}{e^y}\right)^{\alpha}$$

Equation 4.4

To verify that the Zipf coefficient of the Gabaix-Solomon model converges to the predicted value, we also plotted its average over multiple runs of the simulation and observed that it does indeed converge to the expected value.

For Figures 4.1 and 4.2, simulations were initialized with municipality data from 1812, and had a normal shock distribution with mean of 1.01 and variance of 8e-04. There were 1077 cities and the lowest allowed fraction of the average size was c=0.34. For these values of N and c, the Gabaix-Solomon model predicts a Zipf coefficient of 1.5. For Figure 4.3, the simulations were initialized with a normal distribution with a mean of 7650 and a variance of 4.2e06, and used a normal shock distribution with a mean of 1.01 and a variance of 8e-04. There were 500 cities, and the lowest allowed fraction of the average size was c=0.34, so the predicted Zipf coefficient was 1.49. The Zipf coefficients were averaged over 100 trials.

In Figure 4.2, the exponent used to draw the Pareto Distribution was 1.53, as calculated by the Hill Maximum Likelihood Estimator. In Figure 4.3, the Zipf Coefficient was also calculated by the Hill Maximum Likelihood Estimator. Note that the speed of convergence in Figure 4.3 is typical for simulations initialized with artificial city size distributions, such as lognormal, normal, or uniform. However, some empirical city size distributions converge much faster, within on the order of 100 time steps.

Figure 4.1 Distribution of a sample run of the Gibrat simulation after 10,000 time steps.

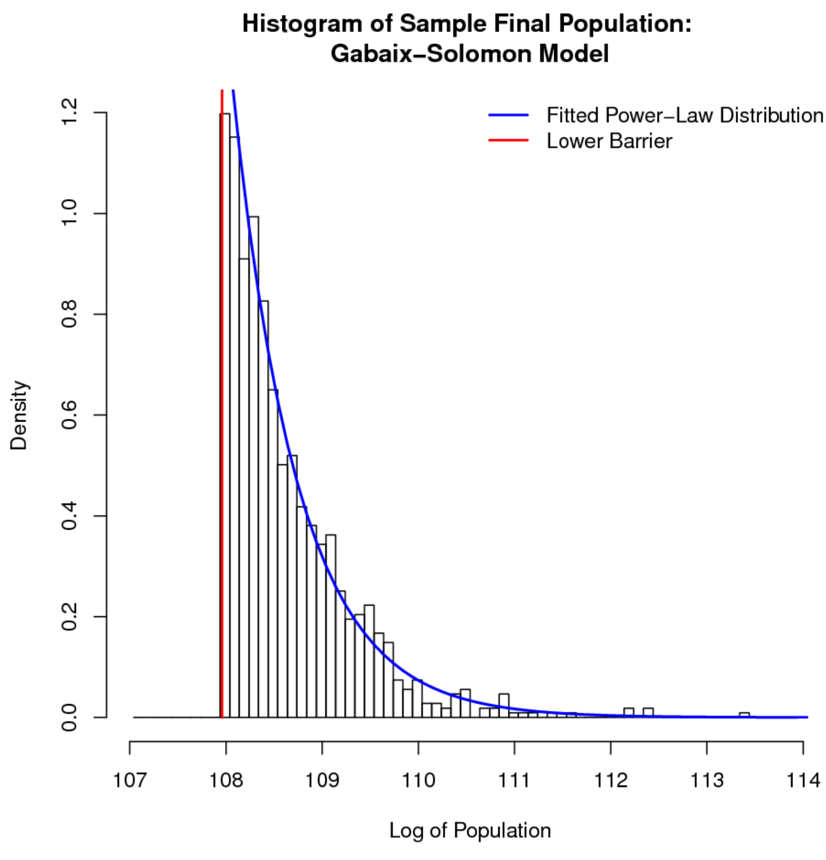
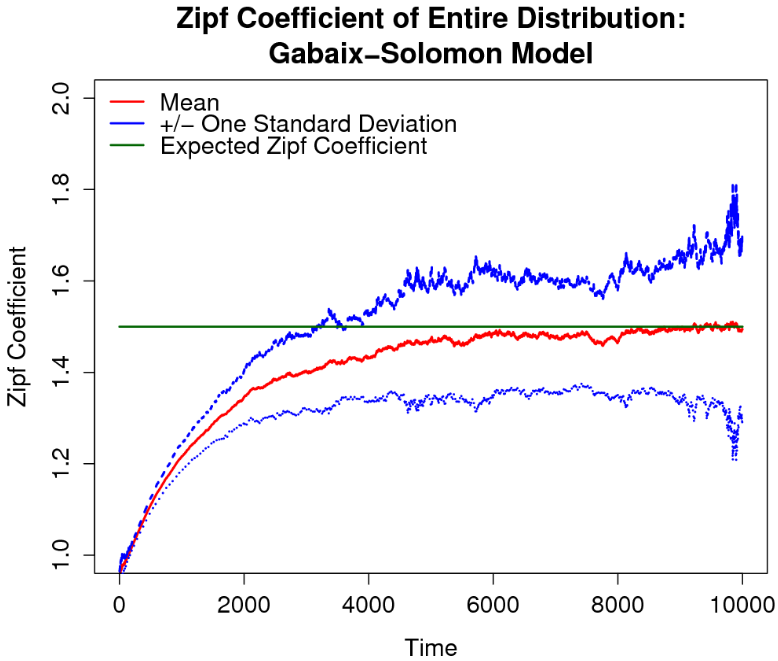


Figure 4.2 **me steps.**

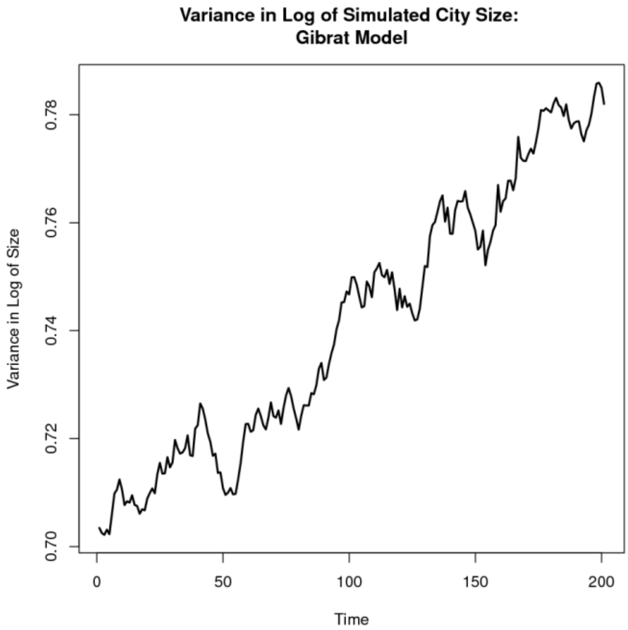
Figure 4.3 Zipf coefficient of Gabaix-Solomon simulation over 10,000 time steps, averaged over 100 trials.

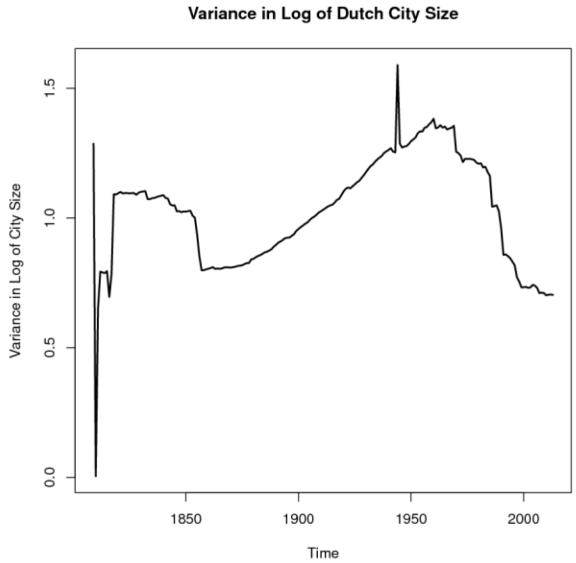


4.2.2 Behavior of Variance Matches Predictions

To check the result of the Gibrat model that the variance of the log of the city size increases steadily with time, we plotted the variance of both the city size and the log of the city size for sample runs of both the Gibrat and Gabaix-Solomon models, as well as the observed values for Dutch municipalities from 1811-2013 (Figure 4.4). All simulations were initialized with municipality data from 1812, and had a normal shock distribution with mean of 1.01 and variance of 8e-04. For the Gabaix-Solomon simulation, the lowest allowed fraction of the average size was c=0.34. Indeed, the Gibrat model produces a steadily increasing variance in log of size, while the Gabaix-Solomon model shows no such pattern. The pattern of variance in the empirical data does not match the behavior of either model.

Figure 4.4 Variance in log of city sizes, simulated and observed.



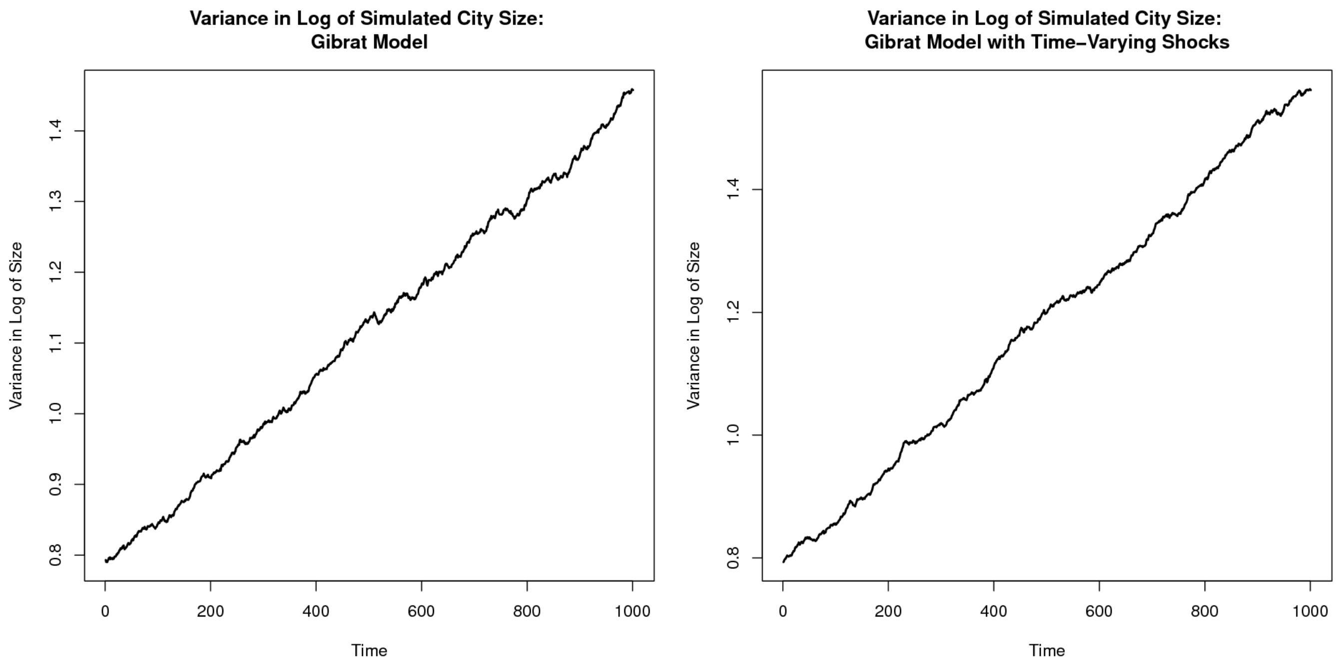


4.3 Checking Assumptions

4.3.1 Effect of Time-Varying Shock Distribution

To check the assumption that the results of both models did not change when the mean and variance of the growth rate distribution varied with time, we compared the results of two different conditions on both the Gibrat and the Gabaix-Solomon models: one simulation with a time-independent mean and variance, and one in which the mean and variance at every time step were selected from time-independent distributions. In the time-independent case, the mean growth rate was 1.01 and the variance was 8e-04. In the time-dependent case, the mean was selected from a distribution with mean of 1.01 and variance of 1e-06, while the standard deviation was selected from a distribution with mean of √(8e-04) and a variance of 1e-06. Simulations were initialized with municipality data from 1812. For the Gabaix-Solomon model, the lower limit fraction was c=0.34. We found that, indeed, after a sufficiently long time the resulting city size distributions did have the same shape (lognormal in the Gibrat simulation and Pareto in the Gabaix-Solomon simulation) and, in the case of the Gibrat model, the same linear patterns of variance over time (see Fig. 4.5).

Figure 4.5 Comparison of time-independent and time-dependent shock distributions for the Gibrat model after 1000 time steps.



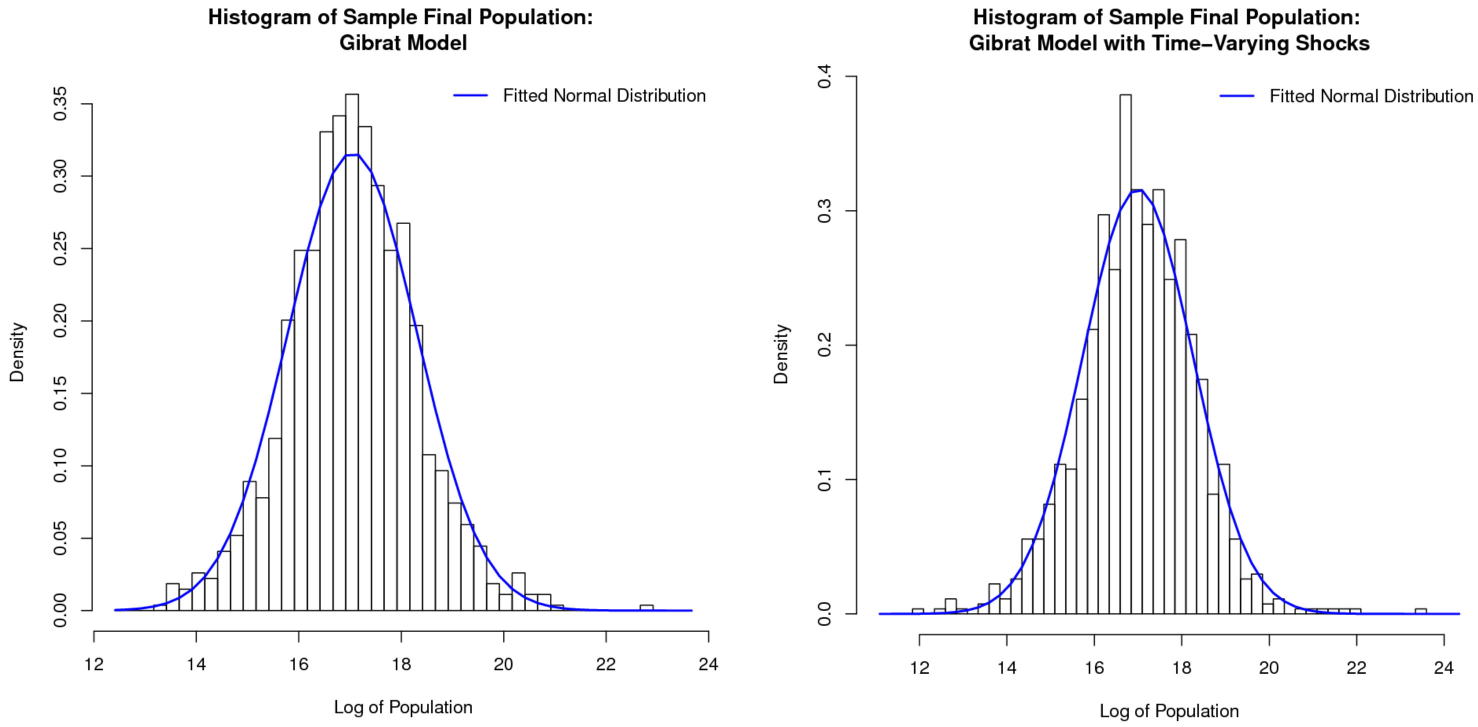
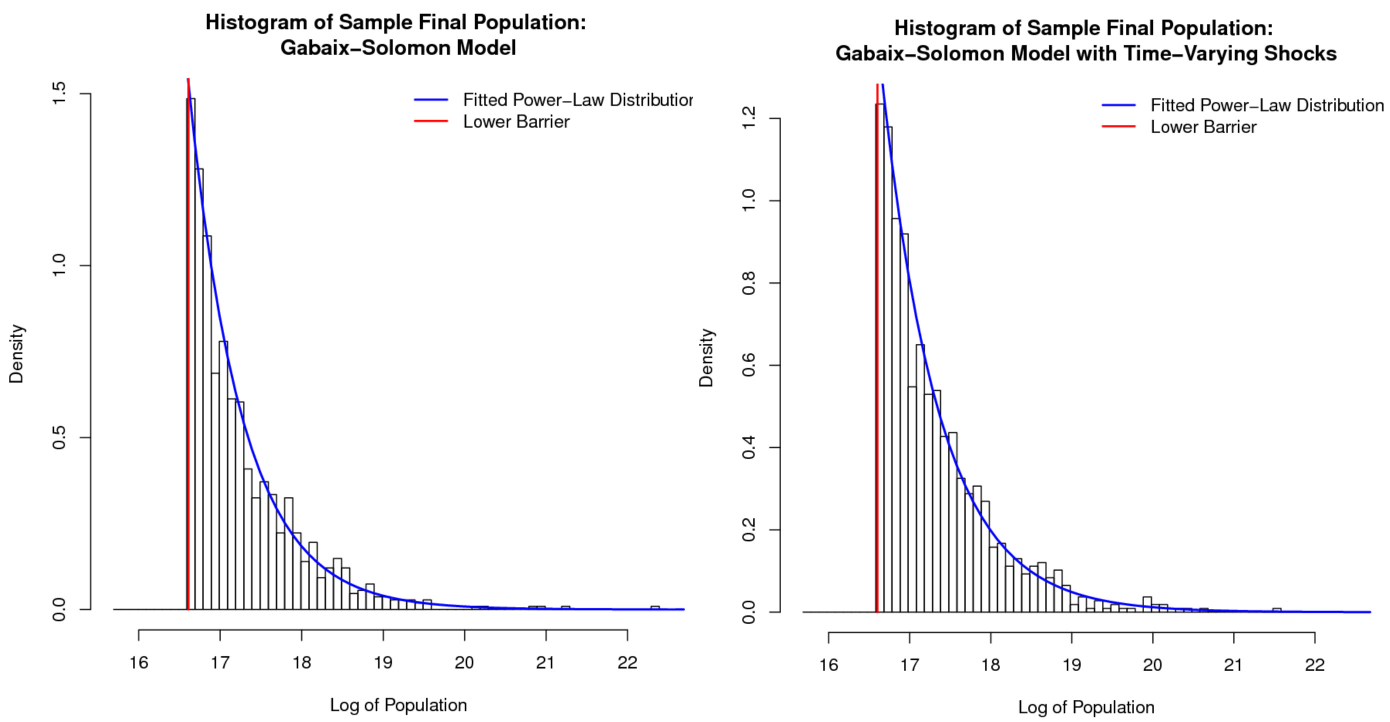


Figure 4.6 **Comparison of time-independent and time-dependent shock distributions for the Gabaix-Solomon model after 1000 time steps.**



4.3.2 Sensitivity to Distribution Shapes

To test how sensitive the models were to changes in the initial city size and growth rate distributions, we held one distribution constant and compared the average Zipf coefficients of the final population distributions when the other distribution was changed.

In the first test, we held the initial size distribution constant (a lognormal distribution, with mean of log size = 8.9 and variance of log size = 0.07). We varied the growth rate distributions, which were all calibrated to have a mean of 1.01 and variance of 8e-04. In the second test, we held the growth rate distribution constant (a normal distribution, with mean = 1.01 and variance = 8e-04.) We varied the initial size distributions, which were calibrated to have a mean of 7650 and a variance of 4.2e06.

For both tests, we averaged the results of 100 trials of 4000 time steps each. The Gabaix-Solomon model had a lower limit of 0.34 times the average size. For the Gibrat simulation, we calculated the Zipf coefficient of the distribution by using linear regression on the top 20% of cities. For the Gabaix-Solomon simulation, we used the Hill Maximum Likelihood Estimator on the entire population.

We found that, after a sufficiently long time, both models were quite robust to changes in the distribution of both their initial city size and their growth rates (see Tables 4.1 and 4.2).

Table 4.1 Average Zipf coefficients of final city size distribution, with a fixed lognormal initial size distribution and a varying growth rate distribution.

Growth Rate Distribution	Gibrat	Gabaix-Solomon
Uniform	1.13	1.43
Normal	1.16	1.45
Gamma	1.11	1.49
Logistic	1.16	1.46

Table 4.2 Average Zipf coefficients of final city size distribution, with a fixed normal growth rate distribution and a varying initial size distribution.

Initial Size Distribution	Gibrat	Gabaix-Solomon
Identical	1.18	1.43
Normal	1.16	1.46
Logistic	1.15	1.45
Lognormal	1.16	1.45

These findings verify the assumptions in both of the models that the shape of the final size distribution does not depend on either the shape of the growth rate distribution or the shape of the initial size distribution.

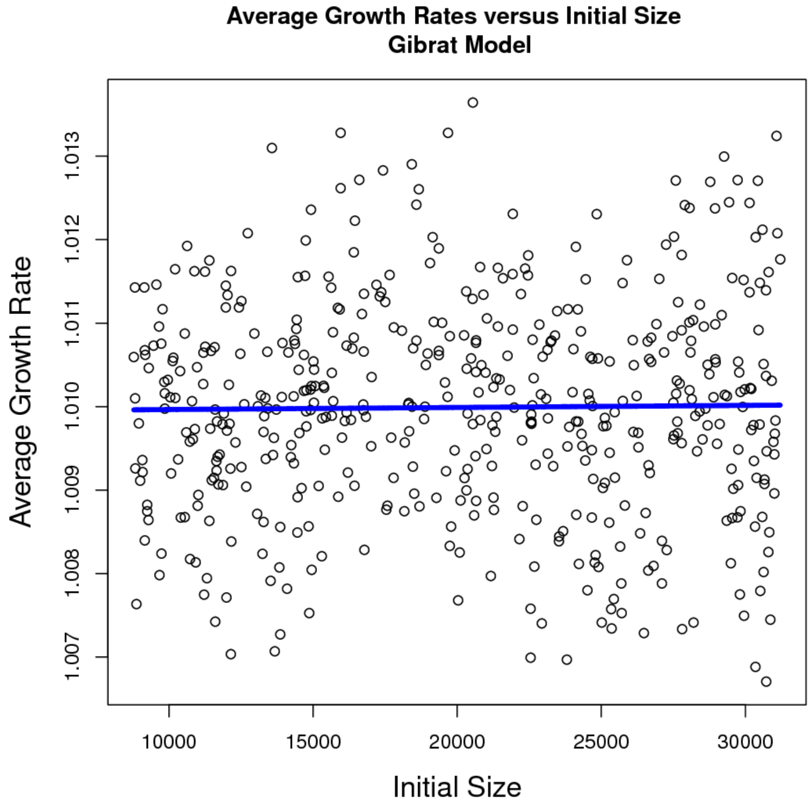
4.3.3 Gibrat's Law of Proportionate and Size-Independent Growth

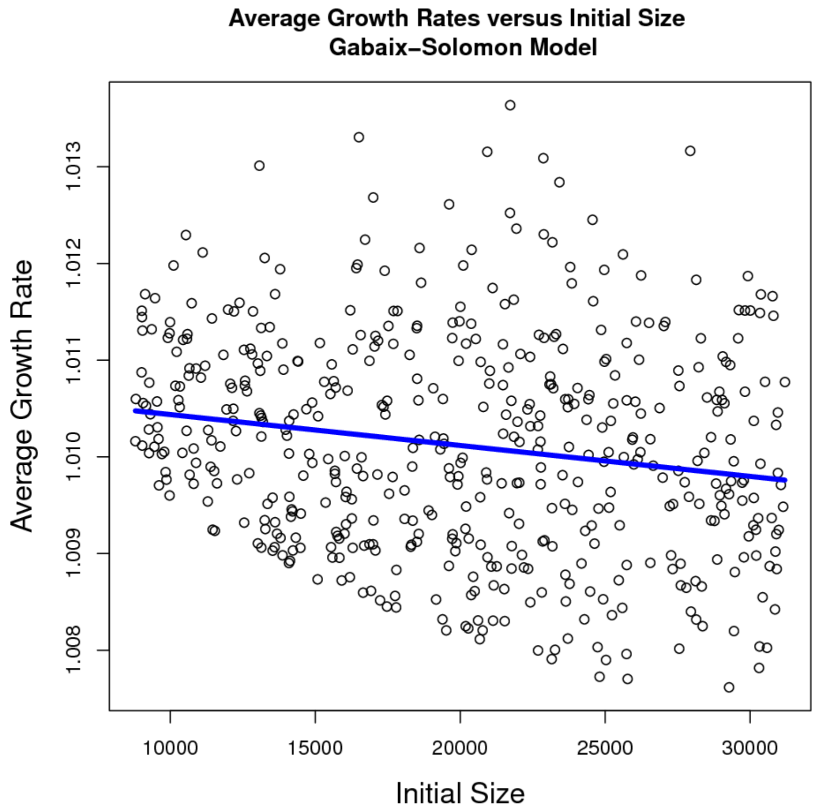
To test whether Gibrat's Law is maintained in each of the two models, we simulated the growth of a collection of cities and plotted the average growth rates of each city against their initial size. Both simulations were initialized with a uniform city size distribution with a minimum of 8755 and a maximum of 31,225 and experienced a normal shock distribution with a mean of 1.01 and a variance of 8e-04. There were 500 cities and the simulations were run for 500 time steps. In the Gibrat model, growth rates below 0 were forced to 0.01. In the Gabaix-Solomon model, growth rates resulting in a size below the lower limit of 0.34 times the average city size were adjusted so no size would fall below the lower limit.

In each simulation, we used linear regression to fit a line to the average growth rate of a city versus its initial size (see Figure 4.7). While we report results for only one trial, the findings are representative. We found that in the Gibrat model, the slope of the line was not significantly different from 0 ($p=0.236$, adjusted R-squared = 8.1e-04). In contrast, in the Gabaix-Solomon model we found that the slope of the line was significantly different from 0 ($p=3.82e-05$, adjusted R-squared = 0.03). We can conclude, therefore, that while the assumption that the distribution of growth rates is independent of size - that is, Gibrat's Law - is valid for the Gibrat model, it is violated in the Gabaix-Solomon model. This result verifies the finding of Skouras (2010), who found that imposing a lower limit violates Gibrat's Law in the Gabaix-Solomon model by artificially increasing the growth rate of smaller cities.

The implications of this finding for the validity of the model are unclear. Empirically, it has been found that the variation in growth rates decreases with size (Eeckhout 2004, Rozenfeld et al. 2008), but Figure 4.7 indicates that under the Gabaix-Solomon model, the variation in growth rates *increase* with size, due to the artificially imposed lower bound. On the other hand, the negative correlation between average growth rate and size is consistent with some empirical findings on deviations from Gibrat's Law (Black and Henderson 2003, Rozenfeld et al. 2008). Further investigation of this issue is warranted.

Figure 4.7 **Comparison of correlation between average growth rate and initial city size in sample runs of the Gibrat and Gabaix-Solomon simulations.**



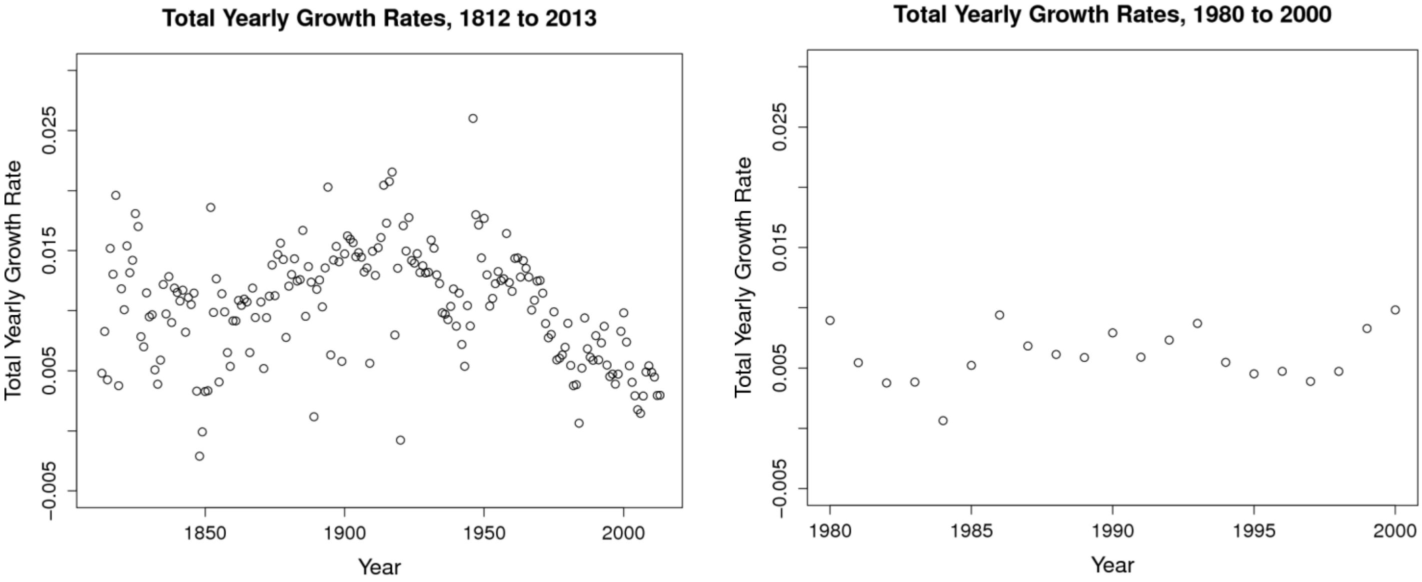


4.4 Comparing Models to Empirical Data

Finally, we compared predictions of a city size distribution by the Gibrat and Gabaix-Solomon models to real population distributions. To make the predictions, we gathered information about the starting distribution and the growth rates in the intervening period. Then, we used the theory of the models to predict a final distribution. Note that in contrast to previous sections, we were not performing simulations -- rather, we used empirical data and equations from the theoretical models to generate an expected distribution.

Because average growth rates vary over time, we selected 1980-2000 as a time period close to the present during which growth rates were relatively constant (see Figure 4.8). We calculated the total growth rate from 1980 to 1990 and assumed that the total growth rate from 1990 to 2000 would be the same. Specifically, we calculated the average yearly growth of the natural logarithm of the mean city size from our "training" dataset of 1980-1990, and then we extrapolated that growth for the appropriate length of time of our prediction. We focused on the growth of the logarithm instead of the actual size because both models are more easily visualized on the log scale.

Figure 4.8 Yearly growth rates of total population of Dutch municipalities, 1812 to 2013.



4.4.1 Lognormal Distribution

First, we predicted a lognormal distribution for municipalities in the Netherlands in 2000 from the distribution in 1990 and the average growth rate from 1980 to 1990, using the Gibrat model. To make this prediction, we used the following variables and formulas:

P_{1980} = mean of the natural log of city population in 1980
P_{1990} = mean of the natural log of city population in 1990
V_{1990} = variance of natural log of city population in 1990
V_{GR} = variance in growth rates of all cities between 1980 and 1990
T_1 = 10 = number of years between1980 and 1990
T_2 = 10 = number of years between1990 and 2000

Average year-to-year growth of log sizes = G = $\frac{P_{1990}-P_{1980}}{T_1}$

Predicted mean of lognormal distribution in 2000 = $\ln(P_{1990}) + T_2*G$

Predicted standard deviation of lognormal distribution in 2000 = $\sqrt{V_{1990} + T_2 * V_{GR}}$

This prediction is graphed in Figure 4.9. Note that the predicted variance includes both the variance of the initial population and the variance predicted by the Gibrat model. It was necessary to account for the initial variance because for a time scale of only 10 time steps, it would not yet have become small relative to the variance in growth rates.

4.4.2 Pareto Distribution

Similarly, we predicted the upper tail of the 2000 distribution from the population distributions of 1980 and 1990 using the Gabaix-Solomon model. Because this model requires a Zipf coefficient, we assumed that the Zipf coefficient in 1990 was constant through time. We focused on modeling the top 10% of cities because the available data described municipalities, which are distributed lognormally (in contrast to more sophisticated definitions of urban areas, which have a power law distribution). Only the upper tail of the municipality data might be considered to be Pareto distributed, so we used the Zipf coefficient of the top 10% of cities. To make our prediction, we used the following variables and formulas:

t_{1980} = mean of the natural log of the top 10% of city populations in 1980
t_{1990} = mean of the natural log of the top 10% of city populations in 1990
T_1 = 10 = number of years between1980 and 1990
T_2 = 10 = number of years between1990 and 2000

Z = Zipf coefficient of top 10% of city populations in 1990
N = 10% of total number of cities in 1990
c therefore implicitly determined by Equation 3.4

Average growth rate of top 10% = $G_{10} = \frac{t_{1990} - t_{1980}}{T_1}$

Reflection barrier = B = c * exp(t_{1990} + T_2*G_{10})

Then, the PDF of the Pareto distribution for city size S, as given by Equation 1.6, is

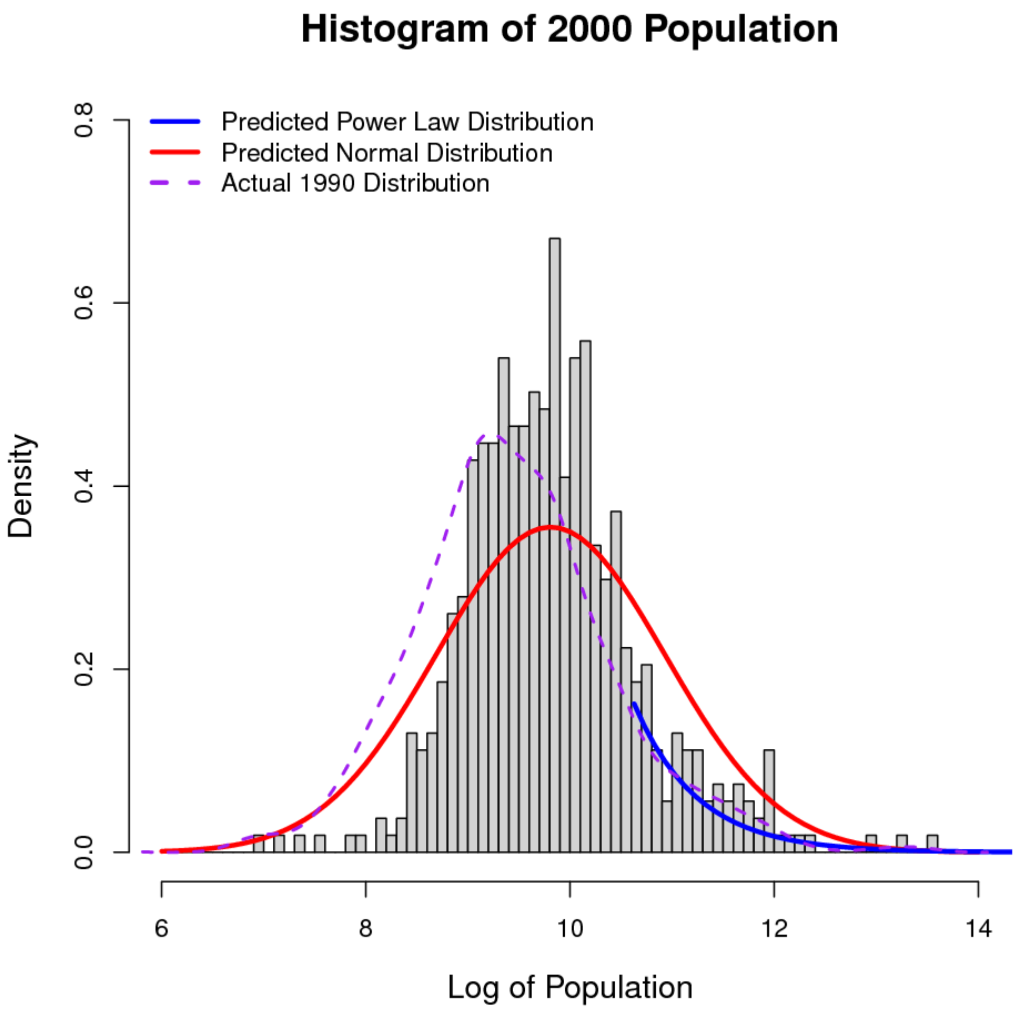
$$P(\ln(S) = s) = 0.10 * Z * \left(\frac{B}{\exp(s)}\right)^{Z}$$

where the factor of 0.10 accounts for the fact that the total probability of the truncated distribution should be 10%. The figure generated by this procedure is also depicted in Figure 4.9.

4.4.3 Evaluation

The Gabaix-Solomon prediction does not perform better than simply assuming the 1990 distribution will remain unchanged. The Gibrat prediction improves on this prediction in the lower tail, but overestimates on the upper tail. Also note that the Gibrat prediction already demonstrates signs of degeneration due to its increased variance. On the question of whether the graphs demonstrate that pure randomness can explain the observed changes in city size distributions, there is room for interpretation.

Figure 4.9 **A comparison of two predictions for the 2000 distribution of Dutch municipality sizes based on the population distributions of 1980 and 1990.**



5 Conclusion

By performing simulations of two stochastic models for city growth, we were able to verify that both models performed as expected. We confirmed that the Gibrat model would produce a lognormal distribution with the log of the city size normally distributed as N(E[log(X_0)]+$\sum_{t=1}^{T} E[log(1+m_t)]$, $\sum_{t=1}^{T} \sigma^2(y_t)$), and we verified that the Gabaix-Solomon model would produce a power-law distribution with exponent given by $N = \frac{\alpha-1}{\alpha}\left[\frac{\left(\frac{c}{N}\right)^{\alpha}-1}{\left(\frac{c}{N}\right)^{\alpha}-\frac{c}{N}}\right]$. Moreover, we confirmed the result that the Gibrat model will have steadily increasing variance. In addition, we verified that both models were not sensitive to changes in shape of initial city size and growth rate distributions, and that they were robust to time-varying growth rates. We also verified that Gibrat's Law is maintained in the Gibrat model but is violated in the Gabaix-Solomon model.

These results confirm that our understanding of the Gibrat and Gabaix-Solomon models is thorough and correct. On the basis of this understanding, we can now better assess whether deviations from the model might produce distributions similar to those that have been empirically observed. This will allow us to evaluate whether or to what extent city size distributions are shaped by effective randomness as opposed to specific economic forces, which is the basic question faced by every student of Zipf's and Gibrat's Laws.

The relevance of this question can be demonstrated by a thought experiment. Imagine that all borders in the European Union disappear and the region becomes fully economically integrated. Zipf's Law predicts that Paris and London, which currently both have populations of around 11 million, will diverge in size. Both of the stochastic models studied in this paper would produce this result. However, neither one can predict which city will be the "winner." For policy makers, both the identity of the winner and the factors leading to its success are of great importance, and unfortunately neither of these models provides insight in those areas.

In this context, a historical perspective on the dissolution of borders could be illuminating. Dittmar (2010) examines European cities from the eighth to the nineteenth centuries, and finds that Zipf's Law only emerged in Western Europe between 1500 and 1800. During this period, land constraints on the growth of cities which restricted the availability of food were drastically relaxed, due to a dramatic reduction in transport costs and increases in agricultural productivity (Dittmar, 2010). After the lifting of these constraints, city growth rates became independent of size and Gibrat's Law began to hold after 1500 (ibid).

Dittmar's finding supports the theory that random growth, as opposed to other economic factors, is the primary force behind Zipf's Law. Some economic models attempt to explain Zipf's Law as deriving from the random allocation of geographic advantages (Krugman 1996) or the industrial specialization of cities (for example Cordoba, 2004 or Rossi-Hansberg and Wright, 2007). However, because geographical attributes remained unchanged during 1500-1800, and the growth of factories only appeared afterwards, these theories cannot explain the emergence of Zipf's Law during this period.

While Dittmar's results support the pure-randomness hypothesis, they also identify transport costs as an important economic factor influencing growth rates. Another part of the paper associates limited labor mobility (in the form of serfdom) with decreased growth. It could be illuminating to conduct further investigations into how these factors relate to growth. Additionally, it might be useful to supplement Dittmar's work with an investigation of another period of economic integration, such as the consolidation of the United States in the 18th century. Another potential avenue of related research is agent-based modeling, which could mimic how both economic forces such as wages and price of living and social forces such as language barriers and distance from family might impact an individual's migration decision.

Finally, it is important to recognize that while it is clear that random growth can generate Zipf's Law, and that it is possible that cities do exhibit effectively random growth, it is not obvious that this random growth is necessarily the cause of the observed Zipf's Law. This is good news for policy-makers, who would struggle to find ways to encourage city growth if it were truly a random process. But until the growth process is proved to be random, we can continue to search for an economic way to predict and influence it.

6 References

Auerbach, F. (1913). "Das Gesetz der Bevölkerungskonzentration." *Petermanns Geographische Mitteilungen*, pp. 74–76.

Adamic, L.A., (2002). "Zipf, Power-Laws, and Pareto - A Ranking Tutorial."

Alperovich, G. (1984). "The Size Distribution of Cities: On the Empirical Validity of the Rank-Size Rule." *J. Urban Economics*, Vol. 16, pp. 232-239.

Alperovich, G. (1993). "An Explanatory Model of City-Size Distribution: Evidence from Cross-Country Data." *Urban Studies*, Vol. 30, No. 9, pp. 1591-1601.

Berry, B.J.L. and Okulicz-Kozaryn, A. (2012). "The City Size Distribution Debate: Resolution for US Urban Regions and Megalopolitan Areas." *Cities*, Vol. 29, Supplement 1, pp. S17-S23

Black, D. and Henderson, V. (2003). "Urban Evolution in the USA." *J. Economic Geography*, Vol. 3, No. 4, pp. 343-372.

Blank, A. and Solomon, S. (2000). "Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components)." *Physica A: Statistical Mechanics and its Applications*, Vol. 287, pp. 279-288.

Córdoba, J. C. (2008a). "On the distribution of city sizes." *J. Urban Economics*, Vol. 63, pp. 177–197.

Córdoba, J. C. (2008b). “A generalized Gibrat’s Law for cities.” *International Economics Review*, Vol. 49, pp. 1463–1468.

Dittmar, J. (2011). “Cities, Institutions, and Growth: The Emergence of Zipf’s Law.” Under review. Available at http://www.jeremiahdittmar.com/files/Zipf_Dittmar.pdf

Duraton, G. (2006). “Some foundation for Zipf’s Law: Product proliferation and local spillovers.” *Regional Science and Urban Economics*, Vol. 36, pp. 542–563.

Duraton, G. (2007). “Urban evolutions: The fast, the slow and the stall.” *American Economic Review*, Vol. 97, pp. 197–221.

Eeckhout, J. (2004). “Gibrat’s Law for (all) Cities.” *American Economic Review*, Vol. 94, No. 5, pp. 1429-1451.

Gabaix, X. (1999a). “Zipf’s Law and the Growth of Cities.” *American Economic Review*, Vol. 89, No. 2, pp. 129-132.

Gabaix, X. (1999b). “Zipf’s Law for Cities: An Explanation.” *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 739-767.

Gabaix, X. (2009). “Power Laws in Economics and Finance.” *Annual Review of Economics*, Vol. 1, pp. 255-293.

Gabaix, X. and Ibragimov, R. (2011). “Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents.” *J. Business & Economics Statistics*, Vol. 29, pp. 24-39.

Gibrat, R. (1931). “Les Inégalités économiques.” Paris, France: Librairie du Recueil Sirey.

González-Val, R. (2012). “Zipf’s Law: Issues in Empirical Work.” *Région et Développement*, Vol. 36, pp. 147-164.

Hill, B. M. (1975). “A simple approach to inference about the tail of distribution.” *Annals of Statistics*, Vol. 3, pp. 1163–1174.

Ioannides, Y. M., & Overman, H. G. (2003). “Zipf’s Law for cities: An empirical examination.” *Regional Science and Urban Economics*, Vol. 33, pp. 127–137.

Kalecki, M. (1945). “On the Gibrat Distribution.” *Econometrica*, Vol. 13, pp.161-170.

Krugman, P. (1996). “Confronting the mystery of urban hierarchy.” *Journal of the Japanese and International Economies*, Vol. 10, pp. 399–418.

Levy, M. (2009). “Gibrat’s Law for (all) cities: Comment.” *American Economic Review*, Vol. 99, pp. 1672–1675.

Levy, M., & Solomon, S. (1996). “Dynamical explanation for the emergence of power laws in a stick market model.” *International Journal of Modern Physics C*, Vol. 7, pp. 65–72.

Malcai, O., Biham, O., & Solomon, S. (1999). “Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements.” *Physical Review E*, Vol. 60, pp. 1299-1303.

Malevergne, Y., Pisarenko, V., & Sornette, D. (2009). “Gibrat’s Law for cities: Uniformly most powerful unbiased test of the Pareto against the lognormal.” Research paper series: 09-40. Swiss Finance Institute.

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*, Vol. 83, 036111.

Mulder, P. and H. de Groot (forthcoming), “The Economics of Zipf’s Law: A Meta-Analysis.”

Nitsch, V. (2004). “Zipf Zipped.” *J. Urban Economics*, Vol. 57, pp. 86-100.

OECD (2012), Redefining “Urban”: A New Way to Measure Metropolitan Areas, OECD Publishing. http://dx.doi.org/10.1787/9789264174108-en

Rosen, K.T. and Resnick, M. (1980). “The Size Distribution of Cities: An Examination of the Pareto Law and Primacy.” *J. Urban Economics*, Vol. 8, pp. 165-186.

Rossi-Hansberg, E., & Wright, M. L. J. (2007). “Urban structure and growth.” *Review of Economic Studies*, Vol. 74, pp. 597–624.

Rozenfeld, H.D., Rybski, D., Andrade, J.S. Jr., Batty, M., Stanley, H.E., & Makse, H.A. (2008). “Laws of population growth.” *Proceedings of the National Academy of Sciences*, Vol. 105, pp. 18702-18707.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). “The area and population of cities: New insights from a different perspective on cities.” *American Economic Review*, Vol. 101, pp. 2205–2225.

Saitchev, A., Malevergne, Y., & Sornette, D. (2009). “Theory of Zipf’s Law and Beyond.” Springer.

Skouras, Spyros. (2010). “Gibrat’s Law Does Not Imply Zipf’s Law.” Available at SSRN: http://ssrn.com/abstract=1717112 or http://dx.doi.org/10.2139/ssrn.1717112

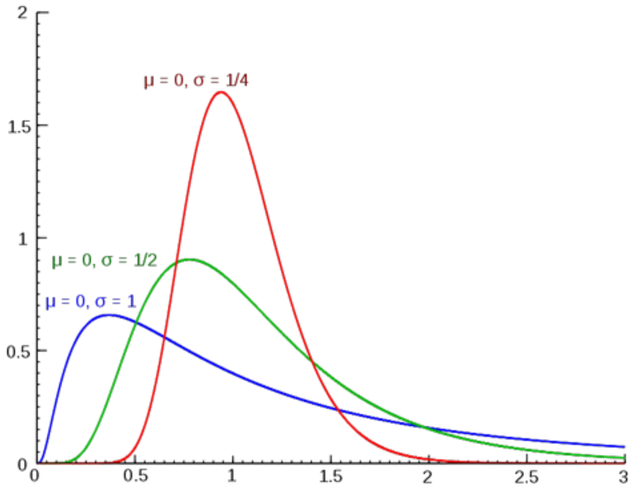
World Bank. (2014). “Urban population (% of total).” http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS/countries/1W?display=graph

Zipf, G. K. (1949). Human behavior and the principle of least effort. Cambridge, MA: Addison Wesley.

7 Appendix A - The Lognormal Distribution

The Gibrat Distribution, which was formalized in section 2 and whose mean and variance are given in section 2.2.1, is a special case of the lognormal distribution. The lognormal distribution is a probability distribution in which the logarithm of a variable X is normally distributed, with mean μ and variance σ^2.

Here are three examples of lognormal distributions:



(Source: http://www.quora.com/Finance/What-does-it-mean-if-a-future-stock-price-is-lognormally-distributed)

7.1.1 Mean and Variance of the Lognormal Distribution

The mean and variance of the lognormal distribution of city sizes at time T can be derived from the mean and variance of the normal distribution. Note that the mean of the lognormal distribution cannot be obtained directly from the mean of the normal distribution -- it must also take into account the variation. The equations for transformation of the normal mean and variance μ and σ^2 into the mean and variance m and v of the lognormal distribution are:

$$m = e^{\mu+\frac{\sigma^2}{2}}$$
$$v = e^{2\mu+\sigma^2}\,(e^{\sigma^2} - 1)$$

(Source: http://www.mathworks.nl/help/stats/lognstat.html)

7.1.2 Median of the Lognormal Distribution

Because Y = log(X) is symmetrically distributed (the normal distribution is symmetric), the mean μ is also the median of the Y values. Therefore, the median of the untransformed variable X is e^{μ}.

Examining the ratio of the mean to the median also lends some insight into the main problem with the result of the Gibrat Distribution. To make a small simplification without loss of generality, assume that all the variances of the city sizes are the same and increasing over time. Then the variance of log(size) across all cities is $\sigma^2 = t\,\sigma_y^2$. We can interpret this in terms of the untransformed distribution by looking at the ratio of the mean to the median:

$$\frac{mean}{median} = e^{\frac{t\sigma^2}{2}}$$

Clearly, as time progresses, the mean will grow larger and larger in relation to the median: the upper tail of the untransformed distribution will stretch out in the positive direction to a disproportionately large size. It is possible that this unrealistic result may be solved by applying another model to the growth of cities.

8 Appendix B - Details of Derivation in Malcai et al. (1999)

Continuing from Equation 3.3, we now describe the dynamic rule of Equation 3.1 in terms of a master equation for the probability distribution on city sizes, which we denote *p(w,t)*. With some manipulation, this master equation will resemble an equation that was previously shown by Levy and Solomon (1996) to result in a power law distribution. The master equation is:

$$p(w,t+1) = p(w,t) + \frac{1}{N}\left[\int_{\lambda} \mathrm{p}\left({}^{\mathrm{w}}\!/_{\lambda},\mathrm{t}\right)\,\Pi(\lambda)\mathrm{d}\lambda\; - \mathrm{p}(\mathrm{w},\mathrm{t})\right]$$

Equation 8.1

In words, this equation says that the probability that a city size is equal to w at time T = t+1 is equal to the probability that it is equal to w at the previous time T=t (and it doesn't move) plus the probability that it moves to size w from any other size, minus the probability that it moves away from size w to a different size[18].

In order to account for the possibility of boundless drift, we now normalize the city size variables[19]:

$w_i \to w_i / \hat{w}$

To avoid confusion, we will not use ŵ to mean the average of the normalized city size variables, because this average is by definition[20] equal to 1. Malcai et al. take the minimum value of w to be c and the maximum value to be N. Therefore, we have the normalization condition over $p(w,t)$:

$$\int_c^N \boldsymbol{w}\, \boldsymbol{p}(\boldsymbol{w}, \boldsymbol{t}) d\boldsymbol{w} = \mathbf{1} \qquad \textbf{Equation 8.2}$$

To put the master equation in a more convenient form, we now consider the logarithm of the sizes. We define the new variable

$W_i = \ln(w_i)$

and note that Equation 3.1 becomes

$$W_i(t+1) = W_i(t) + \ln(\lambda(t)). \qquad \textbf{Equation 8.3}$$

This transformation means that the probability density function p(w) must be transformed as well. We will state without proof that the new probability density function is

$P(W) = e^W p(e^W)$

but details of this derivation can be found at http://www.ebyte.it/library/docs/math04a/PdfChangeOfCoordinates04.html.

The master equation then becomes

---

[18] There is a factor of 1/N due to the fact that only one city undergoes a shock at any given time. It is unclear whether this is necessary to apply the later calculations that lead to the result, or whether this is an "optional" feature of the model.

[19] Since this only amounts to a system-wide multiplicative factor, this normalization does not impact the results in a significant way.

[20] To clarify: $\frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{\hat{w}} = \frac{1}{N\hat{w}}\sum_{i=1}^{N} w_i = \frac{1}{N\hat{w}}(N\,\hat{w}) = 1.$

$$P(W, t+1) = P(W, t) + \frac{1}{N}\left[\int_{\lambda} \mathrm{P}(\mathrm{W} - \ln(\lambda), \mathrm{t})\ \Pi(\lambda) \mathrm{d}\lambda\ - \mathrm{P}(\mathrm{W}, \mathrm{t})\right]$$

Equation 8.4

It was proved by Levy and Solomon (1996) that equations of this form have a steady-state solution

$$\boldsymbol{P}(\boldsymbol{W}) \sim \boldsymbol{e}^{-\boldsymbol{\alpha} \boldsymbol{W}}\ .$$ **Equation 8.5**

Translating back into the original (normalized) city size variable w, we have

$$\boldsymbol{p}(\boldsymbol{w}) = \boldsymbol{K}\ \boldsymbol{w}^{-\boldsymbol{\alpha}-\mathbf{1}}.$$ **Equation 8.6**

Dividing Equation 8.2 by the probability normalization condition

$$\int_{\boldsymbol{c}}^{\boldsymbol{N}} \boldsymbol{p}(\boldsymbol{w}, \boldsymbol{t}) \boldsymbol{d}\boldsymbol{w}\ = \mathbf{1}$$ **Equation 8.7**

we can derive an implicit expression for the value of the exponent α:

$$N = \frac{\alpha - 1}{\alpha}\left[\frac{\left(\frac{c}{N}\right)^{\alpha} - 1}{\left(\frac{c}{N}\right)^{\alpha} - \frac{c}{N}}\right]$$