Center for
Quantitative
Economics

Gibrat, Zipf, Fisher and Tippett: City Size and Growth Distributions Reconsidered

Christian Schluter[†] and Mark Trede[‡]

27/2013

[†] Aix-Marseille Université, France and University of Southampton, UK
[‡] Department of Economics, University of Münster, Germany

wissen.leben
WWU Münster

Gibrat, Zipf, Fisher and Tippett: City Size and Growth Distributions Reconsidered

Christian Schluter*

Aix-Marseille Université (Aix Marseille School of Economics), CNRS & EHESS and University of Southampton

Mark Trede†

Universität Münster

September 12, 2013

Abstract

This paper is about the city size and growth rate distributions as seen from the perspectives of Zipf's and Gibrat's law. We demonstrate that the Gibrat and Zipf views are theoretically incompatible in view of the Fisher-Tippett theorem, and show that the conflicting hypotheses about the size distribution are testable in a coherent encompassing estimating framework based on a single index.

We then show that the two views can be reconciled in a slightly modified but internally consistent statistical model: we connect economic activity and agglomeration in a model of Gibrat-like random growth of sectors, whose random number is linked to Zipf-like city size. The resulting average growth rate is a random mean, and we derive its invariant distribution.

Our empirical analysis is based on a recent administrative panel of sizes for all cities in Germany. We find strong evidence for the prediction of the growth model, as well as for a weak version of Zipf's law characterising the right tail of the size distribution.

Keywords: Zipf's law, Gibrat's law, city size, urban growth

JEL Codes: R11, R12

*`christian.schluter@univ-amu.fr`, Corresponding author. GREQAM, Centre de la Vieille Charité, 13002 Marseille, France, and Department of Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.

†`mark.trede@uni-muenster.de`, Center for Quantitative Economics and Center for Nonlinear Science, Westfälische Wilhelms-Universität Münster, Am Stadtgraben 9, 48143 Münster, Germany, Tel.: +49-251-83 25006, Fax: +49-251-83 22012.

1 Introduction

The size distribution of cities continues to be the subject of much controversy and debate ever since the 1913 paper by Auerbach, and the applications of ideas expounded in Gibrat (1931) and Zipf (1949). At the heart of the debate are two conflicting views. The Gibrat view holds that city sizes grow proportionately and independently of size, which, by a central limit theorem argument applied to log size, implies that sizes are asymptotically lognormally distributed. By contrast, the Zipf view in its weak form considers only the largest cities and holds that the size distribution is heavy tailed, so that the right tail decays like a power function and not exponentially fast as in the lognormal case. Stronger flavors claim that the exponent of the power function be -1, or that the entire size distribution is Pareto-like. Recent examples are the opposing views of Eeckhout (2004, 2009), and Gabaix (1999b), Córdoba (2008) and Levy (2009), Rozenfeld et al. (2011) and Ioannides and Skouras (2013).[1]

We show that this debate about the size distribution can be addressed in a common statistical framework based on the classic Fisher-Tippett theorem, and the extreme value theory emanating from it. The Gibrat and the Zipf view are revealed to correspond to two (out of three possible) distinct limit distributions admitted by the theorem, and are therefore theoretically incompatible. This implies that the frequently encountered claim in the recent literature that 'Gibrat's law implies and leads to Zipf's law' is wrong. The hypothesis about the tail behavior of the size distribution can thus be equivalently formulated as

[1]To illustrate, Eeckhout (2004) claims that "cities grow proportionately ... and this gives rise to a lognormal distribution of cities", and "it is shown that the size distribution of the entire sample is lognormal and not Pareto", whereas Gabaix (1999b) states "whatever the particulars driving the growth of cities, ... as soon as they satisfy (at least over a certain range) Gibrat's law, their distribution will converge to Zipf." Córdoba (2008) states that "the city size distribution in many countries is remarkably well described by a Pareto distribution."Gabaix and Ioannides (2004) provide an extensive survey of the literature on the city size distribution.

a hypothesis about the domain of attraction of the limit distribution. In particular, power-law behavior for the largest cities is *not* a surprising empirical regularity[2] but must hold for *any* thick tailed size distribution, i.e. distributions in the domain of attraction of the Fréchet distribution. By contrast, the lognormal distribution is in the domain of attraction of the thin-tailed Gumbel distribution. Whether the size distribution is in the domain of attraction of the Fréchet or the Gumbel distribution is an empirical question, which will be analysed in a coherent encompassing estimating framework based on the generalized extreme value distribution. This gives rise to a simple empirical test based on a single index.

While we argue that the standard version Gibrat and Zipf views are theoretically incompatible, we show that the two can be reconciled in a slightly modified but internally consistent statistical model: connecting economic activity and agglomeration, we consider a model of Gibrat-like random growth of economic sectors, whose random number is linked to Zipf-like city size in a manner that is consistent with Christaller's central place theory, a theory that has recently been revisited in Mori et al. (2008) and Hsu (2012). The resulting average growth rate being a random mean, we derive its invariant distribution and verify its empirical validity. In particular, we show that, under the maintained assumptions, the (annual and ten-year) growth rate distribution is heavy-tailed, follows asymptotically a student t-distribution, and in the leading case has an infinite variance, all despite finite variance random sectoral growth. These findings undermine the popular i.i.d. random growth models for city sizes.

[2] E.g. Gabaix (1999a) notes that "A striking pattern of agglomerations is Zipf's law for cities, which may well be the most accurate regularity in economics. It appears to hold in virtually al countries..", while Krugman (1996, p.4) observes that "we are unused to seeing regularities this exact in economics - it is so exact that I find it spooky.", and Córdoba (2008) observes that "at this point we have no resolution to the explanation of the striking regularity in city size distribution."Ioannides and Skouras (2013) conclude that "the Pareto law of city sizes and its exponent remain spooky!"

Our empirical analysis is based on a recent administrative panel of sizes for all cities in Germany. The German case is of interest, since Germany is the most populous state in Europe, and the highly accurate data, based on the legal obligation of residents to register with the authorities, constitute a panel that allows us to study the annual growth process (unlike census-based data studied in e.g. Eeckhout (2004)). We find strong evidence in our data for the prediction of the growth model, as well as for a weak version of Zipf's law characterizing the right tail of the size distribution while the strong forms of Zipf's law are soundly rejected. Not only is the hypothesis of lognormality rejected by implication as a description of the tail behavior of the size distribution, a simple test based on normalizing transforms also soundly rejects lognormality as a description of the main body of the distribution.

2 The City Size Distribution, its Tail Behavior, and Zipf's Laws

Zipf's (1949) classic exposition of the rank size rule pertains to the *largest* sizes. Thus it is a statement about the tail behavior of the size distribution, and the weakest form of a Zipf law can be formulated as the hypothesis that the size distribution has a heavy, regularly varying, right tail which thus decays like a power function (rather than exponentially fast as would be the case for lognormally distributed sizes): for large sizes x, the CDF F of sizes is of the form

$$F_X(x) = 1 - L_1(x)x^{-\alpha} \tag{1}$$

with $\alpha > 0$ and L_1 being a slowly varying function.[3] Hence the right tail of the size distribution is eventually of the Pareto form. Stronger flavors of the law are the hypothesis that α be unity, or that this power function behavior not only applies to large sizes but extends over the entire domain (e.g. Gabaix (1999b) or Córdoba (2008)). We consider the statistical underpinnings of this hypothesis, in order to conclude that the weak form of the Zipf law naturally arises when the appropriate statistical theory is considered.

2.1 Tail Behavior: Extreme Value Theory and the Fisher-Tippett Theorem

The behavior of extreme quantiles and the associated distributional tail is the subject of extreme value theory, which arises from the classic Fisher-Tippett theorem (Fisher and Tippett (1928)) about the limit distribution of the maximum: If, for suitably chosen sequences of norming constants c_n and d_n, $c_n^{-1}(\max(X_1, \ldots, X_n) - d_n)$ converges in distribution to a non-degenerate CDF H, then H belongs to one of only three CDFs, namely the Fréchet, Weibull, or Gumbel. The Gumbel distribution has an upper tail that decays exponentially fast, whereas the upper tail of the Fréchet distribution follows a power law.[4]

Corollaries of the Fisher-Tippett theorem consider maximum domains of attraction (MDA).[5] In particular, the lognormal distribution belongs to the

[3]Equivalently, the quantile function Q_X is then $Q_X(1 - 1/x) = L_2(x)x^{1/\alpha}$ with L_2 slowly varying. Recall that a function g is called regularly varying at x_0 with index θ if $\lim_{x \to x_0} g(tx)/g(x) = t^\theta$ with $t > 0$. If $\theta = 0$, the function is said to be slowly varying.

[4]The Fréchet distribution is, for $x > 0$, given by CDF $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, with α as in equation (1). As $x \to \infty$, we have $1 - \Phi_\alpha(x) = 1 - \exp(-x^{-\alpha}) \approx x^{-\alpha}$. The Gumbel distribution is given by $\Lambda(x) = \exp(\exp(-x))$, and, as $x \to \infty$, $1 - \Lambda(x) \approx \exp(-x)$. The Weibull distribution has a finite upper limit, at which it exhibits a power tail.

[5]Recall that for CDF F, $F \in \mathrm{MDA}(H)$ if there exist norming constants such that the Fisher-Tippett theorem holds for extreme value distribution H. The MDAs are characterized in Embrechts et al. (1997), Theorems 3.3.7 and 3.3.26.

MDA of the Gumbel distribution. Whereas if the tail of the city size distribution is regularly varying with index $-\alpha$, then it lies in the MDA of the Fréchet distribution, and for large x we have $1-F(x) = L(x)x^{-\alpha}$ as stated by equation (1). The weak form of Zipf's law thus arises naturally from the appropriate statistical theory:

Proposition 1 *The weak form of Zipf's law is not a surprising regularity but necessary for* ***any*** *thick tailed size distribution.*

Hence the often observed empirical regularity for the largest sizes across a wide range of subjects (see e.g. Mitzenmacher (2003)), including sizes of cities, firms, and incomes, is not surprising nor "spooky" (Krugman (1996, p.40), Ioannides and Skouras (2013)), as often claimed in the literature, but rather expected. We now have a common statistical framework:

Proposition 2 *(The weak form of) Zipf's law corresponds to the hypothesis that the city size distribution is in the MDA of the Fréchet distribution, whereas the (standard) Gibrat view is that it is in the MDA of the Gumbel distribution.*

The standard Gibrat and Zipf view are thus incompatible, as they correspond to one or the other limit distribution. Hence the claim that 'Gibrat's law implies and leads to Zipf's law', frequently encountered in the recent literature, is wrong.

The analysis of the tail properties of the limit distribution can be conducted in a common estimating framework, since all three limit laws of the Fisher-Tippett theorem are embedded in the Generalized Extreme Value distribution, given by

$$G_\alpha(x) = \exp\left(-\left[1 + \frac{1}{\alpha}\left(\frac{x-\mu}{\sigma}\right)_+^{-\alpha}\right]\right). \tag{2}$$

$\alpha > 0$ is the Fréchet case, $\alpha \to 0$ is the Gumbel case, and $\alpha < 0$ is the negative Weibull case. Hence we do not need to impose a particular distributional model (such as lognormality, or a power tail[6]), unlike many contributors to the city size literature. An estimator of this index α of the Generalized Extreme Value distribution is proposed in Dekkers, Einmahl and de Haan (1989), and is discussed in greater detail below in Section 4.1. Hence, the question as to whether the size distribution has heavy tails or is lognormal can be tested directly and simply by testing the sign of the estimator of α:

Proposition 3 *A test of the city size distribution hypotheses of Proposition 2 is a test of the sign of* α.

We defer the implementation of this test on our data to the empirical Section 4.1 below.

3 The City Growth Rate Distribution

Despite the incompatibility of the standard Gibrat and Zipf views, we now show how the two can be reconciled in a slightly modified but internally consistent statistical model. Rather than considering a model of random growth for cities, consider cities composed of sectors that exhibit random growth. We connect economic activity, measured by the number of sectors, to city size and agglomeration, a link that has been amply documented in the literature. The randomness of the numbers of sectors leads to cities of different sizes, and the average growth rate is a random mean whose invariant distribution we are able to determine.

Consider then cities composed of economic sectors. The numbers of sectors S_i in city i are random and are related to the size of the city X_i according to

$$S_i = C + \lambda \ln X_i + \varepsilon_i, \tag{3}$$

[6] Note also that the well-known Hill estimator is not consistent in this general framework.

where C is a constant, and ε is a mean-zero error term. Empirical evidence in support of (3) is reported in e.g. Mori et al. (2008, Figures 6 and 7), who also argue that such a relation is in line with Christaller's (1966) central place theory and (a weak form of) the hierarchy principle, which asserts that sectors (/ industries / goods) found in cities of a given size will also, on average, be found in cities of larger sizes. For a recent formalization of central place theory that provides microfoundations see Hsu (2012).

To simplify the exposition, we assume that the error term ε_i has a density f_ε which is symmetric and has bounded support on $[-\bar{\varepsilon}, \bar{\varepsilon}]$. We introduce both Gibrat and Zipf features, by assuming that each sector exhibits random growth (detailed below) and that the city size distribution is heavy-tailed (empirically verified below), so its distribution function is of the form given by (1): $F_X(x) = 1 - L_1(x)x^{-\alpha}$. Hence the right tail of the size distribution is eventually of the Pareto form, so $\ln(X_i)$ is exponential eventually. We have the following lemma:

Lemma 1 *Under the maintained hypotheses, the distribution of economic sectors is exponential with parameter $p \equiv \alpha/\lambda$ for large s:*

$$\Pr\{S_i > s\} = L_3(s)\exp(-ps) \tag{4}$$

where L_3 is slowly varying.

The average growth rate R_i of city i is

$$R_i = S_i^{-1} \sum_{j=1}^{S_i} r_j \tag{5}$$

where each sector j grows at the random rate r_j, with $E(r_j) = \mu$ and $Var(r_j) = \sigma^2 < \infty$. It follows that the average growth rate is the mean of a random number of summands since S_i is random.

What is the limit distribution of the average growth rate of cities, if it exists, as the numbers of sectors become large (noting that $E(S_i) \to \infty$ as $p \to 0$) ? The answer is not immediate, since neither classic central limit theory does apply as we consider a random sum, nor do the classic limit theorems for random sums apply (e.g. Gnedenko and Korolev (1996)) because the maintained assumptions differ. However, our next proposition gives the remarkable answer to our question.

Theorem 1 *Under the maintained hypotheses,*

$$\sqrt{\frac{1}{p}}\left(\frac{R_i - \mu}{\sigma}\right) \to T \sim t_2 \quad \text{as } p \to 0, \tag{6}$$

irrespective of the sampling distribution of the individual growth rates r_j.

Thus the normalized city size growth rate distribution follows asymptotically the student t_2-distribution. R_i is then distributed for small p approximately as a scaled t_2 variate with scale parameter $\sigma p^{1/2}$. This scale parameter is estimable, so, although p is latent, σ and p are jointly identifiable.

We can accommodate small deviations from the statistical model given by (3) which has implied the exponential distribution of sectors, by assuming directly that sectors follow a gamma distribution with shape parameter q approximately equal to 1. In the exact case, we have of course the exponential distribution; in the neighborhood of 1 we have the same tail behavior, but allow for small deviations from the exponential density for medium-range sector numbers. This is attractive, since, like every model, (3) is likely to hold at best approximately. We can then generalize the above statistical theory as follows:

Theorem 2 *Assume that the distribution of sectors follows eventually a gamma* *distribution with shape parameter q. Then*

$$\sqrt{\frac{q}{p}}\left(\frac{R_i-\mu}{\sigma}\right)\rightarrow T\sim t_{2q}\quad \text{as } p\rightarrow 0, \tag{7}$$

irrespective of the sampling distribution of the individual growth rates r_j.

The shape parameter q is identified by the degrees of freedom parameter of the t-distribution, which is estimable. In particular, the EM-algorithm of Scheffler (2008) enables us not only to estimate the normalization factors μ and $p^{1/2}\sigma$, but also the degrees of freedom parameter df. In the light of (3) we expect $df = 2q$ to be close to 2 in our data.

As the limit distribution is a t_{2q}-distribution, its properties immediately yield two important results. First, although random sector growth is conventional and not further specified than having a finite mean and variance, the average growth rate exhibits power-law behavior, i.e. heavy tails of the limit distribution obtain even if the sampling distribution has finite variance (which is in contrast to some limit theorems in extreme value theory which rely on a Fréchet domain-of-attraction assumption). Second, in the leading case whith $q = 1$, its tail is so heavy that its variance is infinite. We collect these results in:

Corollary 1 *The growth rate distribution is heavy-tailed, if $q = 1$ its variance is infinite.*

This infinite variance invalidates one key hypothesis underlying the standard version of Gibrat's Law, i.e. the standard i.i.d. proportional growth model.

We next show that the growth rates will also exhibit medium range temporal dependence. Hence we expect that the assumptions underlying Gibrat's law are likely to be invalid in our empirical application below, leading to a failure of Gibrat's law and the implied lognormality of the size distribution.

3.1 Longer Run Growth Rate Distributions

The limit law of Theorem 2 is the basis for the limit law of the growth rate distribution over longer horizons, such as ten years. Add a time index to the city size and growth variables and let

$$R_{i,t+1} = \log \frac{X_{i,t+1}}{X_{i,t}}.$$

Iterating the equation we have

$$\log(X_{i,t+1}) = \log X_{i,0} + \sum_{\tau=0}^{t} R_{i,\tau+1}. \tag{8}$$

Consider the ten-year growth rate. As the number of sectors S_i does not change much from year to year, it constitutes a common component in the vector of the ten annual growth rates $(R_{i,1}, .., R_{i,10})$. This vector, as a result, approximately follows a multivariate t-distribution. The sum of ten multivariate t-distributed random variables is also t-distributed with the same degrees of freedom (Kotz and Nadarajah (2004)). Thus:

Corollary 2 *The ten-year growth rate approximately follows a t_{2q}-distribution.*

In the very long run, however, the number of sectors can no longer be regarded as approximately constant since it grows or shrinks along with the city size. Hence, for very long horizons, this proposition no longer applies.

The empirical validity of Theorem 2 and Corollary 2 are examined below in Section 4.2 for our data for German cities. Before turning to this empirical evidence, we reconsider first the empirical validity of the maintained hypothesis given by (1) about the city size distribution.

4 Size and Growth Rates Distributions for all German Cities

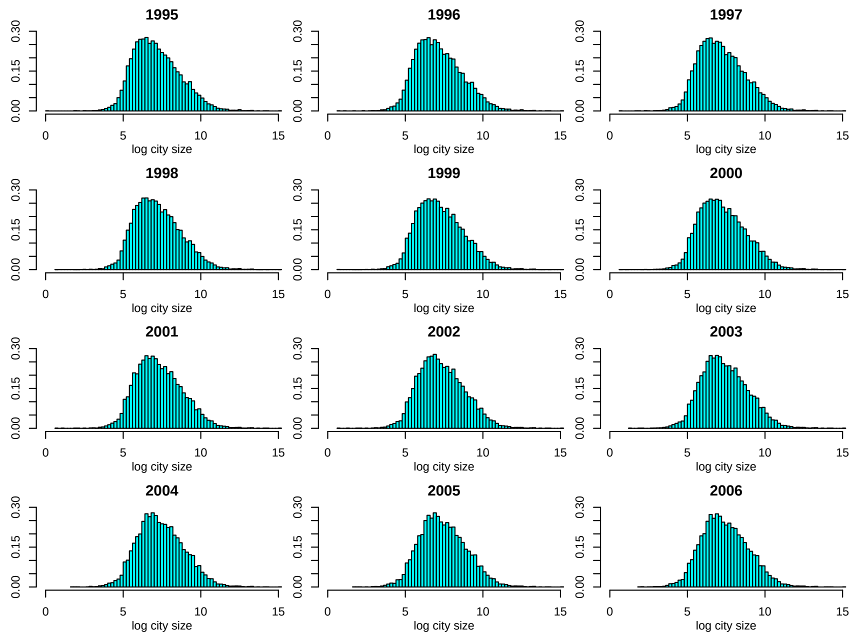
We conduct our statistical analysis using a 12 year panel of administrative data for *all* cities covering the years 1995-2006 provided by German Federal Statistical Office.[7] These administrative data are highly accurate due to the legal obligation of citizens to register with the authorities. The unit of analysis is the "city", or more precisely the municipality or settlement ("Gemeinden"). Population sizes are as of December 31st of each year, and we use a panel of about 14,000 cities.

We summarize some general features of the size distribution. Only three cities have more than 1m inhabitants (Berlin, Hamburg, Munich), 12 to 14 cities have more than 0.5m inhabitants, and about 80 cities have more than 0.1m inhabitants. The size evolution for the 15 largest cities is reported in the data appendix (while there are some changes in their rank order, this group remains unchanged). The mean number of persons in a city is roughly 6,000. Figure 1 depicts histograms of the log size distribution of all cities, which appear fairly stable over time, and look qualitatively similar to other size distributions for other countries reported in the literature (so there is nothing "peculiar" about this German data). One conclusion is already obvious: These distributions are clearly not exponential, so the size distribution cannot be Pareto over the entire support: the data clearly reject the strongest form of Zipf's law.

The histograms also exhibit a distinct skewness, and given the prominence in the literature of the lognormal hypothesis, it is of interest to examine di-

[7] Bosker et al. (2006) consider the case of Germany in a different setting: for the largest 62 German cities they consider a long time series (1925-1999), and examine the impact of the population shock of WWII on subsequent growth rates using time series methods. Since they examine the largest cities, our Proposition 1 is relevant.

Figure 1: Histograms of German city sizes

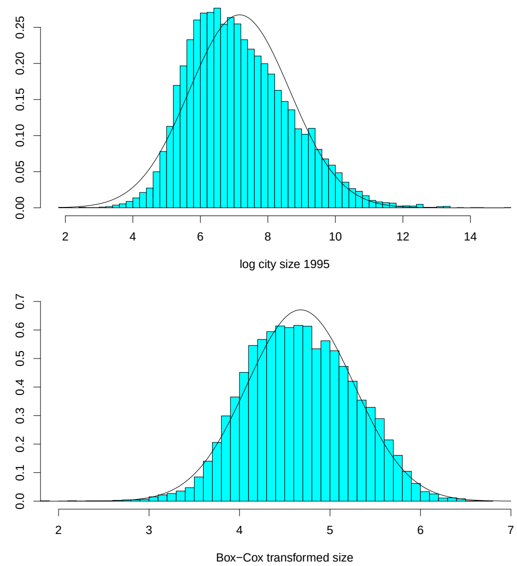


rectly whether this skewness is compatible with lognormality. To this end we consider a class of normalizing transformations that nests the lognormal case. In particular, for city size X, a normalizing transform g seeks to annihilate the skewness of X, so that the distribution of $g(X)$ is close to normal. The specific class of transformations we consider is the Box-Cox transformation given by

$$g_\beta\left(x\right) = \begin{cases} (x^\beta - 1)/\beta & \text{for } \beta \neq 0 \\ \log(x) & \text{for } \beta = 0 \end{cases} \tag{9}$$

The log-transformation $\log(X)$ is thus a special case ($\beta = 0$), as is the linear transformation ($\beta = 1$). For $\beta < 0$, $g_\beta(x)$ has an asymptote at $|\beta|^{-1}$, and the normal target density needs to be truncated. The transformation parameter β is estimable by maximum likelihood, and we can test for lognormality simply using a Wald test for the transformation parameter β.

Figure 2: Fitted Size Distributions



We consider first the representative year 1995. The estimate of the transformation parameter is $-.1262$, statistically different from 0, thus clearly rejecting lognormality. Figure 2 panel 1 depicts the histogram of log sizes and the fitted normal distribution and thus illustrates the excess skewness of the actual size distribution. In Panel 2 we illustrate the success of the normalizing Box Cox transformation which closely matches the normal target density. Turning to the remaining years, Table 1 reports the estimates for our data. It is clear that all estimates are significant and negative. All estimates are inconsistent with the hypothesis of lognormality ($\beta = 0$), which is formally confirmed by Wald tests.

We thus conclude that the skewness in the histograms of Figure 1 is too excessive to be compatible with lognormality. We now turn our attention from the main body of the size distribution to its tail behavior.

Table 1: Estimated Box Cox parameters 1995-2006

Year	$\hat{\beta}$	SE
1995	-0.1262	0.0063
1996	-0.1245	0.0064
1997	-0.1167	0.0064
1998	-0.1147	0.0065
1999	-0.1074	0.0066
2000	-0.1081	0.0066
2001	-0.0979	0.0067
2002	-0.0928	0.0068
2003	-0.0859	0.0070
2004	-0.0788	0.0070
2005	-0.0764	0.0070
2006	-0.0767	0.0071

4.1 The Tail Behavior of the Size Distribution

A consistent estimator of the index of the generalized extreme value distribution (2) is proposed in Dekkers, Einmahl and de Haan (1989), and is given by

$$\hat{\alpha} = \left(1 + H_{K,n}^{(1)} + \frac{1}{2} \frac{H_{K,n}^{(2)}}{(H_{K,n}^{(1)})^2 - H_{K,n}^{(2)}} \right)^{-1}, \tag{10}$$

where $H_{K,n}^{(i)}$ are functions of excesses over a threshold

$$H_{K,n}^{(i)} = \frac{1}{K} \sum_{j=1}^{K} \left(\log X_{(j)} - \log X_{(K+1)} \right)^i$$

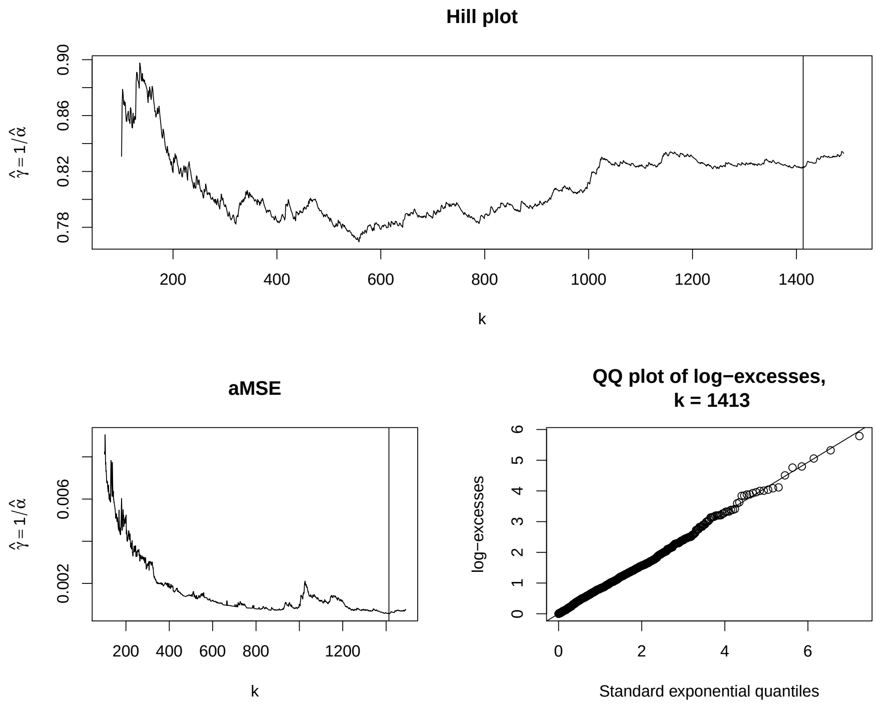
with $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(K+1)}$ denoting the upper order statistics. $H_{K,n}^{(1)}$ is the popular Hill (1975) estimator, which is inadmissible in our generalized setting since it is only consistent for distributions with regular varying tails, and thus requires pre-testing. The threshold $X_{(K)}$ is chosen optimally in a data-dependent way by minimizing the asymptotic mean-squared error (aMSE) criterion. Dekkers et al. (1989, theorem 3.1) show that $\hat{\alpha}$ follows asymptotically a normal distribution. For completeness, we also compute the consistent estimator proposed in Smith (1987) which is based on a likelihood approach.

Table 2: Estimates of the Tail Index $\hat{\alpha}$

	Dekkers et al.		Smith		Hill	
Year		SE		SE		SE
1995	1.321	(0.020)	1.423	(0.038)	1.313	(0.025)
1996	1.323	(0.020)	1.426	(0.038)	1.316	(0.025)
1997	1.325	(0.020)	1.435	(0.038)	1.320	(0.025)
1998	1.325	(0.020)	1.429	(0.038)	1.318	(0.025)
1999	1.330	(0.020)	1.431	(0.038)	1.323	(0.025)
2000	1.330	(0.019)	1.420	(0.038)	1.324	(0.025)
2001	1.329	(0.019)	1.418	(0.039)	1.324	(0.025)
2002	1.330	(0.020)	1.426	(0.038)	1.324	(0.025)
2003	1.313	(0.021)	1.390	(0.040)	1.303	(0.026)
2004	1.313	(0.021)	1.390	(0.040)	1.303	(0.026)
2005	1.313	(0.021)	1.390	(0.040)	1.303	(0.026)
2006	1.313	(0.021)	1.390	(0.040)	1.303	(0.026)

Table 2 reports the results. For all years, the Dekkers et al. estimator (10) is statistically different from 0, so the tail of the size distribution is always heavy. At the same time, the index estimate is statistically different from unity (the value in a strong version of Zipf's law), and very stable over time. The Smith estimator is comparable, as is the now admissible Hill estimator (given this pre-testing). All estimates are coherent, in the sense that all pairwise difference-of-means tests by columns do not reject the null hypothesis that the row estimates are the same (the size of the overall test is controlled by applying the Bonferroni correction to the sizes of the individual tests, and we ignore the positive correlation between the estimates, so the overall test is conservative). The mean of the tail index estimate across all estimates equals 1.35. As this average value is smaller than 2 this implies that the tails are very heavy: the second and higher moments of the size distribution do no exist.

Figure 3: Hill Plot Analysis



4.1.1 Robustness Considerations

In view of the stability of the tail index estimates, and the admissibility of the Hill estimator, we briefly consider the robustness of the estimate with respect to the threshold choice for the representative year 1995. Figure 3 depicts the Hill plot $(k, H_{k,n}^{(1)})$. The Hill estimate is very stable for threshold values 1000 to 1500, and $K_{opt} = 1413$ minimizes the asymptotic mean squared error (computed using the second order approach detailed in Beirlant et al. (2004)) depicted in panel 2. For this threshold Panel 3 of the figure depicts the QQ plot of the log excesses versus the standard exponential distribution, as well as a line with slope $H^{(1)}_{1413,n}$. [8] The model fits the data well.

Another robustness concern might be that the tail index estimate is driven by the largest city, Berlin, which, being the capital, might be structurally unrepresentative. More generally, we can investigate whether the sizes of the j largest cities are compatible with the sizes of the next few largest cities. We do this by conducting the outward testing procedure for heavy-tailed distributions proposed in Schluter and Trede (2008). We find that neither the size of Berlin, nor of any other of the largest 15 cities, is incompatible with the overall power tail behavior.[9]

4.2 The Growth Rate Distribution

We turn to an examination of the growth rate distribution, noting that the established power tail behavior of the size distribution verifies the empirical relevance of one of the maintained hypotheses of Theorem 1. As the theory only requires that the size distribution exhibits a power-function behavior eventually, we ensure that we are approximately in the Pareto tail of the city size distribution, by first examining the Pareto plot of $\log(1-\hat{F}(x))$ on $\log(x)$ and identifying visually the city size for which the plots starts to become approximately linear. In our case, this leads us to consider the largest 5300 cities.

The generalized limit Theorem 2 has two aspects. First, the limit distribution of growth rates is a student t-distribution. Second, its degrees of freedom parameter df equals $2q$, and the model (3) suggests q to be approximately one.

Consider then first the shape of the growth rate distribution. Figure 4 Panel A depicts the histogram of the annual growth rates for the representative years

[8] To see this, consider the quantile function $U(x) \equiv Q_F(1-1/x)$. We have $\log U(x) \to \alpha^{-1}\log x$. Let $p \in (0,1)$, and consider $p_n \simeq p$ such that $j=(n+1)(1-p_n)$ is an integer. Then the log quantile excess satisfy $\log U((n+1)/j) - \log U((n+1)/(k+1)) \to \alpha^{-1}(\log(k+1)-\log(j))$, and are approximated by the log excesses $\log X_{(j)} - \log X_{(k+1)}$.

[9] Full details are available from the authors on request.

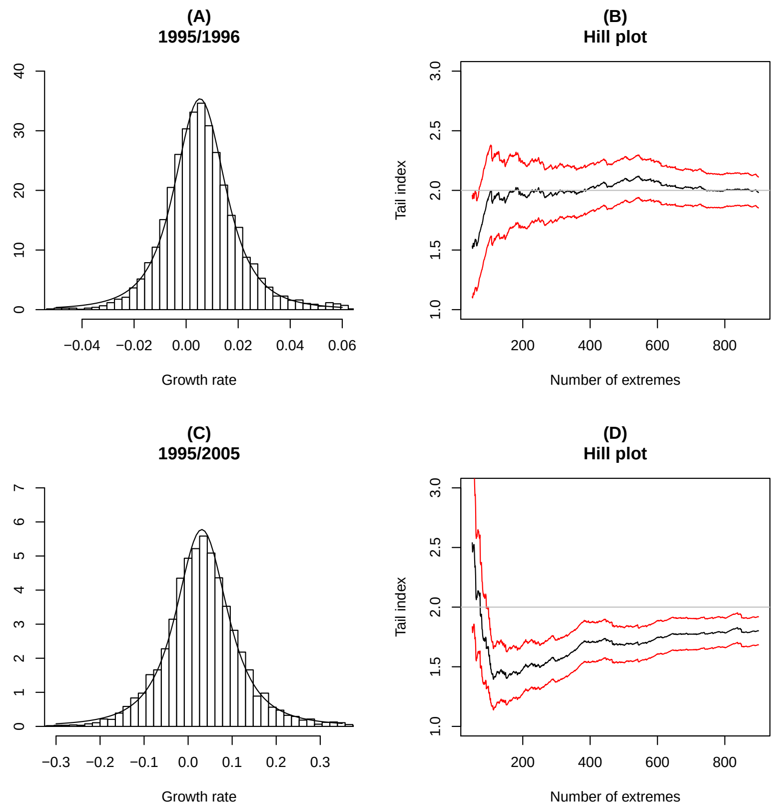
1995/6, as well as the fitted scaled $t_{\hat{df}}$ density, having used Scheffler's (2008) EM-algorithm to estimate $(\mu, p^{1/2}\sigma, df)$. In Panel C of this figure, we consider the ten year growth rate 1995/2005, the subject of Corollary 2. The t-densities fit the data well.

Next, we turn to the degrees of freedom parameter $df = 2q$. The point estimates are reported in Table 3 Panel A. Most estimates suggest a neighborhood of $q = 1$. Rather than taking a "global" approach to estimating q using all data, an alternative local approach is to consider the tail index of the growth rate distribution. The theoretical t_{2q} distribution is, of course, heavy-tailed and has a tail index of $2q$. Under this hypothesis, the Hill estimator becomes available, and the estimates are reported in Panel B of the table, while Panels B and D of Figure 4 depict the Hill plots (including a 95% pointwise confidence band) for the annual and ten-year growth rates. The tail index estimates equally suggest that q is in the neighborhood of 1. We conclude, overall, despite some statistical departures from the focal df and tail index value of 2, that the t_{qr} distribution with q approximately one fits the actual growth rate distribution well for all years.

5 Conclusion

We have argued that extreme value theory establishes that the Gibrat and the Zipf view about the tail behavior of the city size distribution are incompatible as they correspond to two different limit distributions of the Fisher-Tippett theorem. Which hypothesis is empirically relevant has been shown to be testable in an encompassing framework based on the estimate of the index of the generalized extreme value distribution. The empirical evidence in our data for power-tail behavior, i.e. the weakest form of Zipf's law, is overwhelming, as is the evidence against lognormality.

Figure 4: One year and ten year city size growth rates: histograms, the fitted scaled t-distributions, and tail index analysis.



Apart from this tail analysis of the size distribution, we have also examined separately its main body, specifically since the distinction between tail and main body restrictions at times gets ignored in the debate centred around Zipf's law. Using normalizing transforms, lognormality is also rejected as an adequate description of the main body of the size distribution. Nor does Pareto behavior extend over the entire support. Both these empirical observations thus falsify, at least for this data, some current microeconomic theories yielding one or the other completely specified size distribution. By contrast, the observed empirical regularity, and the associated statistical limit theory,

Table 3: The growth rate distribution and parameter estimates for the student t distribution

	A						B		
	μ	SE	$p^{1/2}\sigma$	SE	df	SE	tail index analysis		
	$[\times 10^3]$	$[\times 10^3]$	$[\times 10^2]$	$[\times 10^2]$			Hill	SE	k
1995/96	5.300	0.177	1.031	0.017	2.741	0.112	1.987	0.066	896
1996/97	3.813	0.174	0.997	0.017	2.519	0.097	1.864	0.068	756
1997/98	3.258	0.173	1.005	0.017	2.745	0.112	1.985	0.073	732
1998/99	4.014	0.166	0.904	0.016	1.817	0.055	1.337	0.035	1500
1999/00	2.541	0.165	1.014	0.017	3.985	0.224	3.035	0.192	251
2000/01	3.168	0.156	0.925	0.015	3.156	0.138	2.125	0.073	840
2001/02	1.879	0.148	0.876	0.014	3.122	0.139	1.978	0.073	734
2002/03	-0.151	0.134	0.768	0.012	2.532	0.093	1.487	0.048	977
2003/04	-0.543	0.130	0.750	0.012	2.667	0.102	1.533	0.050	935
2004/05	-2.112	0.134	0.795	0.012	3.258	0.143	1.667	0.068	610
2005/06	-4.022	0.128	0.780	0.012	3.987	0.207	1.861	0.094	394
1995/05	31.073	1.089	6.231	0.110	2.370	0.088	1.715	0.050	1174

pertain only to the largest cities.

However, a slightly modified statistical model reconciles the Gibrat and Zipf view: we have proposed a model of Gibrat-like random growth of sectors, whose random number is linked to Zipf-like city size by Christaller's central place theory. The invariant growth rate distribution implied by our statistical model has been shown to fit our data for German cities well.

A Data Appendix: The 15 Largest German Cities

Table 4: The 15 largest cities and their sizes ($\times 10^3$), 1995-2006

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Berlin	3,471	3,459	3,426	3,399	3,387	3,382	3,388	3,392	3,388	3,388	3,395	3,404
Hamburg	1,708	1,708	1,705	1,700	1,705	1,715	1,726	1,729	1,734	1,735	1,744	1,754
Munich	1,236	1,226	1,206	1,189	1,195	1,210	1,228	1,235	1,248	1,249	1,260	1,295
Cologne	966	964	964	963	963	963	968	969	966	970	983	990
Frankfurt	650	647	643	644	644	647	641	644	643	647	652	653
Essen	615	612	609	603	600	595	592	585	589	588	585	583
Dortmund	599	597	595	592	590	589	589	591	590	589	588	588
Stuttgart	586	586	585	582	582	584	587	588	589	591	593	594
Düsseldorf	571	571	571	568	569	569	571	572	573	573	575	578
Bremen	549	549	547	543	540	539	541	543	545	546	547	548
Duisburg	535	533	529	523	520	515	512	509	506	504	502	499
Hannover	523	523	521	516	515	515	516	517	516	516	516	516
Nuremberg	492	493	490	487	487	488	491	493	494	495	499	501
Leipzig	471	457	446	437	490	493	493	495	498	498	503	507
Dresden	469	461	459	453	477	478	479	480	484	487	495	505

B Mathematical Appendix

Proof of Lemma 1. Under the stated assumptions we have

$$
\begin{aligned}
\Pr(S > s) &= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} P\left(X > \exp\left(\frac{1}{\lambda}(s - C - \varepsilon)\right)\right) f_\varepsilon(\varepsilon) d\varepsilon \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} L\left(\exp\left(\frac{1}{\lambda}(s - C - \varepsilon)\right)\right) \exp\left(-\frac{\alpha}{\lambda}(s - C - \varepsilon)\right) f_\varepsilon(\varepsilon) d\varepsilon \\
&= L_2(s) \exp\left(-\frac{\alpha}{\lambda} s\right) \int_0^{\bar{\varepsilon}} \exp\left(\frac{\alpha}{\lambda}\varepsilon\right) f_\varepsilon(\varepsilon) d\varepsilon \\
&= L_3(s) \exp\left(-\frac{\alpha}{\lambda} s\right)
\end{aligned}
$$

where L_2 and L_3 are slowly varying functions. ■

The proof of Theorems 1 and 2 invokes the following limit theorem:

Theorem 3 *Let $r_1, r_2, \ldots$ be a sequence of i.i.d. random variables with $E(r_k) = 0$ and finite variance $Var(r_k) = \sigma^2 < \infty$ for $k \in \mathbb{N}$. Let ν have a negative binomial distribution, independent of the r_i, with parameters $q > 0$ and $0 < p < 1$. Define the random mean of a random number ν of draws as $\bar{r}_p = \nu^{-1} \sum_{k=1}^{\nu} r_k$. Then as $p \to 0$,*

$$\sqrt{\frac{q}{p}} \cdot \bar{r}_p \to T \sim t_{2q}. \tag{11}$$

Proof of Theorem 1 and 2. The geometric distribution is the discrete counterpart of the exponential distribution of Lemma 1, and the geometric distribution is a special case of the negative binomial distribution, and follows with $q = 1$. Hence the claim of Theorem 1 follows from Theorem 3. Similarly, the negative binomial distribution is the discrete counterpart of the gamma distribution, which establishes Theorem 2. ■

For the proof of Theorem 3, we set, without loss of generality, $\sigma^2 = 1$, and establish first the following result:

Lemma 2 *As $p \to 0$, the random variates $p\nu$ and $\sqrt{\nu}\bar{r}_p$ are asymptotically independent.*

Proof. Consider the joint survival function

$$\begin{aligned} P\left(p\nu > t, \sqrt{\nu}\bar{r}_p > x\right) &= P\left(\nu > \frac{t}{p}, \sqrt{\nu}\bar{r}_p > x\right) \\ &= \sum_{n>t/p} P\left(\nu = n, \sqrt{n}\bar{r}_p > x\right). \end{aligned}$$

Since ν and $r_1, r_2, \ldots$ are independent, the joint probability can be factored as

$$\begin{aligned} \sum_{n>t/p} P\left(\nu = n, \sqrt{n}\bar{r}_p > x\right) &= \sum_{n>t/p} P\left(\nu = n\right) P\left(\sqrt{n}\bar{r}_p > x\right) \\ &= \sum_{n>t/p} P\left(\nu = n\right) \bar{\Phi}(x) \\ &\quad + \sum_{n>t/p} P\left(\nu = n\right) \left[P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x)\right]. \end{aligned}$$

where $\bar{\Phi}$ is the survival function of the standard normal distribution. We now show that the last sum converges to zero as $p \to 0$,

$$\begin{aligned} &\sum_{n>t/p} P\left(\nu = n\right) \left[P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x)\right] \\ \leq &\sum_{n>t/p} P\left(\nu = n\right) \sup_x \left|P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x)\right|. \end{aligned} \tag{12}$$

Since

$$P\left(\nu = n\right) < \frac{1}{(r-1)!}\left(\frac{np}{1-p}\right)^r (1-p)^n$$

it is immediate that

$$P\left(\nu = n\right) < \frac{p}{n}$$

for given p and sufficiently large n. Then

$$\begin{aligned} & \lim_{p\to 0} \sum_{n>t/p} P(\nu = n) \sup_x \left| P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x) \right| \\ \leq \; & \lim_{p\to 0} p \sum_{n>t/p} \frac{1}{n} \sup_x \left| P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x) \right|. \end{aligned} \tag{13}$$

According to theorem 7.8 in Gut (2005)

$$\sum_{n=1}^{\infty} \frac{1}{n} \sup_x \left| P\left(\sqrt{n}\bar{r}_p > x\right) - \bar{\Phi}(x) \right| < \infty,$$

so (13) converges to 0. It follows that

$$\begin{aligned} \sum_{n>t/p} P(\nu = n) P\left(\sqrt{n}\bar{r}_p > x\right) \;&\to\; \sum_{n>t/p} P(\nu = n) \cdot \bar{\Phi}(x) \\ &= \bar{\Phi}(x) \sum_{n>t/p} P(\nu = n) \\ &= \bar{\Phi}(x) \cdot P(p\nu > t). \end{aligned}$$

Hence, the joint probability can be factorized asymptotically, and $p\nu$ and $\sqrt{\nu}\bar{r}_p$ are asymptotically independent. ■

Proof of Theorem 3. Rewrite

$$\sqrt{\frac{q}{p}} \cdot \bar{r}_p = \sqrt{\frac{2q}{2p\nu}} \cdot \sqrt{\nu}\bar{r}_p.$$

Conditioning on ν it is evident that $\sqrt{\nu}\bar{r}_p$ is asymptotically $N(0,1)$ as $p \to 0$. In addition, for $p \to 0$, the random variable $2p\nu$ converges weakly to $V \sim \chi^2_{2q}$. According to the preceding lemma, $\sqrt{2q/(2p\nu)}$ and $\sqrt{\nu}\bar{r}_p$ are asymptotically independent. Hence, as $p \to 0$

$$\sqrt{\frac{q}{p}} \cdot \bar{r}_p \to \sqrt{\frac{2q}{V}} U$$

where $U \sim N(0,1)$ and $V \sim \chi^2_{2q}$ are independent. Since $\sqrt{\frac{2q}{V}}U \sim t_{2q}$ we conclude that the normalized mean converges weakly to a t-distribution with $2q$ degrees of freedom. ■

References

Auerbach, A. (1913). Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen 59*, 73–7.

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (Eds.) (2004). *Statistics of Extremes: Theory and Applications*, Chichester. Wiley and Sons.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2008). A century of shocks: The evolution of the German city size distribution 1925-1999. *Regional Science and Urban Economics*.

Christaller, W. (1966). *Central Places in Southern Germany*. Englewood Cliffs: Prentice Hall.

Córdoba, J. C. (2008). A generalized Gibrat's law. *International Economic Review 49*(4), 1463–1468.

Dekkers, A., J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics 17*, 1833–1855.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review 94*(5), 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. *American Economic Review 99*(4), 1676–1683.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. Berlin: Springer.

Fisher, R. and L. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society 24*(2), 180–190.

Gabaix, X. (1999a). Zipf's law and the growth of cities. *American Economic Review, Papers and Proceedings 89*(2), 129–132.

Gabaix, X. (1999b). Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114*(3), 739–767.

Gabaix, X. and Y. Ioannides (2004). The evolution of city size distributions. In J. Henderson and J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics, Volume 4: Cities and Geography*. Amsterdam: Elsevier.

Gibrat, R. (1931). *Les inegalites economiques; applications: aux inegalites des richesses, a la concentration des enterprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionel*. Paris: Librairie du Recueil Sirey.

Gnedenko, B. V. and V. Y. Korolev (1996). *Random Summation: Limit Theorems and Applications*. Boca Raton: CRC Press.

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

Hsu, W.-T. (2012). Central place theory and city size distribution. *Economic Journal 122*, 903–932.

Ioannides, Y. and S. Skouras (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics 73*, 18–29.

Kotz, S. and S. Nadarajah (2004). *Multivariate t Distributions and Their Applications*. Cambridge, MA: Cambridge University Press.

Krugman, P. R. (1996). *The Self-Organizing Economy*. Cambridge, MA: Blackwell.

Levy, M. (2009). Gibrat's law for (all) cities: Comment. *American Economic Review 99*(4), 1672–1675.

Mitzenmacher, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics 1*, 226–251.

Mori, T., K. Nishikimi, and T. E. Smith (2008). The number-average size rule: A new empirical relationship between industrial location and city size. *Journal of Regional Science 48*(1), 165–211.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review 101*(5), 2205–2225.

Scheffler, C. (2008). A derivation of the EM updates for finding the maximum likelihood parameter estimates of the student's t distribution. Technical note.

Schluter, C. and M. Trede (2008). Identifying multiple outliers in heavy-tailed distributions with an application to market crashes. *Journal of Empirical Finance 15*, 700–713.

Smith, R. (1987). Estimating tails of probability distributions. *Annals of Statistics 15*, 1174–1207.

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.