



**HAL**  
open science

## Size distributions reconsidered

Christian Schluter, Mark Trede

► **To cite this version:**

Christian Schluter, Mark Trede. Size distributions reconsidered. *Econometric Reviews*, 2019, 38 (6), pp.695-710. 10.1080/07474938.2017.1417732 . hal-01994353

**HAL Id: hal-01994353**

**<https://amu.hal.science/hal-01994353>**

Submitted on 12 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Size distributions reconsidered

Christian Schluter & Mark Trede

To cite this article: Christian Schluter & Mark Trede (2018): Size distributions reconsidered, *Econometric Reviews*, DOI: [10.1080/07474938.2017.1417732](https://doi.org/10.1080/07474938.2017.1417732)

To link to this article: <https://doi.org/10.1080/07474938.2017.1417732>

---

 View supplementary material [↗](#)

---

 Accepted author version posted online: 02 Jan 2018.

---

 Submit your article to this journal [↗](#)

---

 Article views: 8

---

 View related articles [↗](#)

---

 View Crossmark data [↗](#)

---

## Size distributions reconsidered

Christian Schluter<sup>1,2</sup> and Mark Trede<sup>3</sup>

<sup>1</sup>Aix-Marseille School of Economics, Marseille Cedex, France

<sup>2</sup>Department of Economics, University of Southampton, Highfield, Southampton, UK

<sup>3</sup>Institute of Econometrics, Department of Economics, Westfälische Wilhelms-Universität  
Münster, Münster, Germany

### Abstract

We consider tests of the hypothesis that the tail of size distributions decays faster than any power function. These are based on a single parameter that emerges from the Fisher–Tippett limit theorem, and discriminate between leading laws considered in the literature without requiring fully parametric models/specifications. We study the proposed tests taking into account the higher order regular variation of the size distribution that can lead to catastrophic distortions. The theoretical bias corrections re-align successfully nominal and empirical test behavior, and inform a sensitivity analysis for practical work. The methods are used in an examination of the size distribution of cities and firms.

Christian Schluter christian.schluter@univ-amu.fr Aix-Marseille School of Economics, 5 Boulevard Maurice Bourdet CS 50498, Marseille Cedex 01 13205, France; Department of Economics, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

**KEYWORDS:** Size distribution; tail behavior; power law; extreme value index; regular variation

**JEL Classification:** C13; C14; R12

Accepted Manuscript

## 1. INTRODUCTION

The distribution of object sizes (the “size distribution” for short) continues to be the subject of much controversy and debate ever since the classic papers of Gibrat (1931) and Zipf (1949). This debate, which extends across many fields in economics and the natural sciences (as discussed below), is important since observable sizes (of e.g. cities/firms/disasters/returns or of e.g. proteins/solar flares) are used to make inferences about the (in)validity of an underlying unobserved generative growth process. At the opposite ends of the spectrum are two classic views. The classic Gibrat “law” holds that sizes grow without constraints proportionately and independently of size, which, by a central limit theorem argument, implies that sizes are asymptotically lognormally distributed.<sup>1</sup> By contrast, one can discern three nested versions of Zipf’s “law” in the literature: The Zipf law in its weakest form considers only the largest sizes and holds that the tail of the size distribution decays like a power function, while stronger forms of the law hypothesize that the exponent be unity or that the entire size distribution be exactly Pareto with unity exponent. Still other models (e.g. modified Simon models) suggest that the tail of the size distribution decays exponentially fast. However, the existing literature offers sharply conflicting views and alternative interpretations of the empirical evidence. We reconsider this debate by isolating its common denominator, namely the question as to whether the tail of the size distribution decays faster than any power function. We show that answering this question is challenging because of distortions induced by the higher order regular

<sup>1</sup>As defined in e.g. Sutton (1997), who, in reviewing Gibrat’s legacy, also discusses how “‘Gibrat’s Law’ [has been combined] with a range of ancillary assumptions.” The limit distribution changes if additional assumptions are imposed on the i.i.d. growth process. More recently, different researchers use the label “Gibrat’s law” differently. There “Gibrat’s law” refers simply to the i.i.d.-ness of the increments. Weaker versions of the “law” only impose constant mean and variances for the increments.

variation properties of the size distribution. Yet, these are invariably ignored by applied researchers, and confound the size distribution debate. This enables us to make three principal contributions.

Our first methodological contribution consists in recasting the key questions about the tail behavior of the size distribution within a unifying statistical framework using concepts from extreme value theory (see e.g. Embrechts et al., 1997, for a textbook treatment). The application of this general framework to the study of the tail of the size distribution is new. In particular, a single estimable parameter  $\gamma$ , the so-called extreme value index, enables us via the classic Fisher–Tippett limit theorem to juxtapose and clearly discriminate between power function-like behavior ( $\gamma > 0$ ) and faster tail decay ( $\gamma = 0$ ). Using the allied concepts of maximum domain of attraction (MDA), the situation  $\gamma > 0$  corresponds to the size distribution  $F$  being in the maximum domain of attraction of the Fréchet distribution (denoted by  $F \in MDA(G_{\gamma>0})$ ), whereas  $\gamma = 0$  corresponds to  $F$  being in the MDA of the Gumbel distribution ( $F \in MDA(G_0)$ ). We interpret thus the question as to whether the tail of the size distribution decays faster than any power function as requiring a test of the null hypothesis  $F \in MDA(G_0)$  versus the alternative  $F \in MDA(G_{\gamma>0})$ . To the extent that extreme value theory has been used before in the applied literature, researchers in urban and industrial economics usually *impose*  $\gamma > 0$ , and ignore the empirical challenges caused by subexponentiality and higher order regular variation discussed below.

We then study three complementary tests that have been proposed in the statistical literature that do not impose any parametrizations other than the assumption that the Fisher–Tippett theorem holds. Our test framework subsumes not only the classic “laws” discussed above but also situations in

which the tail decays exponentially fast, or in which power function behavior is generated by functions other than exact Pareto distributions. We do not test one fully parametric distribution against another (e.g. exact lognormal v. exact Pareto), the usual approach taken in the size distribution literature. It also turns out that the conditions for such tests are not met by our data. As there is widespread skepticism among researchers about the merit of testing whether one fully specified distribution fits exactly the data (Gabaix and Ioannides, 2004), we believe that our hypotheses impose very informative restrictions on the data of object sizes.

Our second methodological contribution focuses on diagnosing and overcoming the complications for inference that arise when the nuisance part of the size distribution function  $F$  decays only slowly. The empirical literature on the size distribution has to date – wrongly as it will turn out – ignored such higher order regular variation. We reveal that the three tests under consideration can exhibit catastrophic test size distortions. These theoretical biases are then studied under an assumption about second order regular variation. The subexponential lognormal case is considered in detail (while other distributions are considered in the Online Appendix). Knowledge of the theoretical biases is shown to re-align successfully the theoretical and empirical test sizes. The biases are also ranked for several distributions pertaining to our null hypothesis. In situations in which the theoretical bias cannot be estimated, we discuss the merits of a sensitivity analysis for the test statistics. We show that the parameter  $\sigma$  of the lognormal distribution can be interpreted as a sensitivity parameter. If the bias correction does not reverse the test conclusion for a “reasonable” range of sensitivity parameters, the test result can be considered as robust.

In our empirical applications we reconsider the size distribution for cities in the US and Germany, and for firm sizes in France. These examples constitute our third contribution as they demonstrate the relevance of our general hypotheses and the proposed test procedures: the empirical size distributions are neither exactly Pareto nor exactly lognormal, thus rendering some prominent tests from the literature (discussed below) inapplicable for our data. Our test results yield robust evidence in favor of Zipf's weak law for the upper tail of the city size distribution, i.e. power function-like tail decay while stronger forms of the law (a unity exponent, exact Pareto behavior, exact Pareto with unity exponent) are clearly rejected. These findings imply that some leading theoretical approaches to generative laws of invariant city size distributions are empirically flawed.

The size distribution debate extends across many fields in economics, and beyond into the natural sciences. This is not surprising since size data are informative about theoretical generative growth processes. Gabaix (2009) provides a survey of these, see also Schluter and Trede (2016). For instance, Kesten (1973) has shown that a stationary solution to a stochastic recurrence equation satisfies a regular variation condition (see also Goldie, 1991). In particular, adding a lower reflecting barrier to Gibrat's classic model of i.i.d. proportional growth yields a Pareto size distribution with unity exponent (Gabaix, 1999a, so that "Gibrat plus frictions" yields Zipf strongest law, which is sometimes abbreviated as "Gibrat's law causes Zipf's law"). By contrast, Reed (2002) shows that a subordinated geometric Brownian motion, specifically assuming that random lifetimes are exponentially distributed, yields a size distribution whose lower and upper tails are power functions. Reed and Hughes (2002) argue that such "killed" processes are widely applicable in the social and natural sciences. Cabral and Mata (2003) in turn argue that firm size distributions are less shaped by

survival/selection, and more by constraints on firm sizes that diminish as firms age. They conclude that the firm size distribution evolves over time towards a lognormal distribution. Sutton (1997) presents a modified Simon model that yields a size distribution with exponentially fast tail decay.

Recent examples in economics are the size distributions of macroeconomic disasters (Barro and Jin, 2011), of firms (e.g. Sutton, 1997, Cabral and Mata, 2003 or Luttmer, 2007) or of sales (Eaton et al., 2011), income distributions (see e.g. Schluter and Trede, 2002), the returns distribution in finance (see e.g. Schluter and Trede, 2008). In the literature on exchange rates, finance, insurance and risk management (where often  $1/\gamma \in (2, 4)$ ), Ibragimov et al. (2013, 2015) emphasize that heavy tails may lead to sub-optimal diversification in the value-at-risk framework, non-robustness of several economic and financial models, while finiteness of variances is crucial for the applicability of classical econometric approaches. Gabaix et al. (2006) propose a model that generates power tails for financial returns if the size distribution of market participants follows Zipf's law. The place of the size distribution debate in other sciences is surveyed in e.g. Mitzenmacher (2003). An illustration of the current debate taking place in urban economics are the opposing views of Eeckhout (2004) on one side, and Gabaix (1999b), Córdoba (2008), and Levy (2009) on the other side. Similar debates concern the firm size distribution.<sup>2</sup>

<sup>2</sup>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 Córdoba (2008) states that “the city size distribution in many countries is remarkably well described by a Pareto distribution.” Gabaix and Ioannides (2004) survey the city size distribution literature. Cabral and Mata (2003) summarize the firm size distribution (FSD) literature thus: “Conventional wisdom received from these studies has held that expected firm growth rates are independent of size (Gibrat's Law), and that the FSD is stable and approximately lognormal.”

The outline of the paper is as follows. In Section 2 we review various definitions of “tail behavior laws” that appear in the literature. We embed all approaches into a unifying framework based on extreme value theory. Section 3 presents the formal statistical tests of the Gumbel-Gibrat hypothesis, derives the biases under second order regular variation and suggests an adjustment to remedy them. A simulation study demonstrates that the adjustment works well in practical applications. In Section 4 we apply the tests to data on city and firm sizes. All proofs are collected in the Online Appendix, which also provides further supplementary material.

## 2. TAIL BEHAVIOR OF THE SIZE DISTRIBUTION

Does the tail of a size distribution decay faster than any power function? The literature on size distributions often refers to some prominent “laws”, although different researchers sometimes offer implicitly or explicitly different flavors or definitions. We first provide a nuanced discussion and some extended definitions of our own before embedding these in a unifying statistical framework based on maximum domain of attractions.

Zipf’s (1949) classic exposition of the rank size rule pertains to the *largest* sizes. Thus this 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 decays like a power function: the distribution function  $F$  of sizes is of the form

$$F(x) = 1 - L_F(x)x^{-\frac{1}{\nu}}, \quad \text{for large sizes } x \tag{1}$$

with  $\gamma > 0$  and  $L_F$  being a slowly varying function. Recall that a function  $g$  is called regularly varying at  $x_0$  with index  $\theta$  if  $\lim_{x \rightarrow x_0} g(tx)/g(x) = t^\theta$  with  $t > 0$  (if not stated otherwise,  $x_0$  is infinity). If  $\theta = 0$ , the function is said to be slowly varying. Hence, in (1) the right tail of the size distribution is *eventually* of the Pareto form. For instance, we will consider below distributions that are members of the Hall (1982) class given by

$$F(x) = 1 - ax^{-1/\gamma}[1 + bx^\beta + o(x^\beta)], \quad \text{for large sizes } x \quad (2)$$

with  $\gamma, a > 0$ ,  $b \in \mathbb{R}$ ,  $\beta < 0$ . As  $\beta \uparrow 0$ , the nuisance part in the tail decays more slowly, which, we show below, will induce power distortions in tests.

Stronger flavors of Zipf's law are the hypothesis that  $\gamma$  be unity, or that this power function behavior not only applies to large sizes but extends over the entire domain so that  $L_F$  is a constant and the *entire* size distribution exactly Pareto with exponent 1,

$$F(x) = 1 - cx^{-1/\gamma}, \quad \text{for all } x \geq c \quad (3)$$

with  $\gamma = 1$ . We refer to the latter two as Zipf's stronger and strongest law. These nuanced definitions, summarized in Table 1, differ from the language used by some researchers, who, when invoking "Zipf's law", refer to versions (b) or even (c) (e.g. Gabaix, 1999a). The nesting of the hypotheses is obvious. Note that the  $p^{\text{th}}$  moment of the size distribution is finite only if  $p < 1/\gamma$ , so the mean is infinite if the stronger or strongest Zipf's laws hold.

Table 1: “Laws” and Labels.

Law	$F$	Hypothesis
(a) Zipf’s weak law	equation (1)	$F \in MDA(G_{\gamma>0})$
(b) Zipf’s stronger law	equation (1) with $\gamma = 1$	$F \in MDA(G_1)$
(c) Zipf’s strongest law	equation (3) with $\gamma = 1$	Exact Pareto with exponent 1
(d) classic Gibrat’s law	equation (4) with $\delta = 2$	Exact lognormal
(e) Gumbel-Gibrat		$F \in MDA(G_0)$

By contrast, a lognormal size distribution is induced by Gibrat’s law in its classic form (i.e. unconstrained proportional i.i.d. growth of one mature cohort). For large  $x$ , the tail of this distribution can be written more generally as  $1 - F(x) \sim cx^\alpha e^{-\lambda \log^2 x}$  with  $\alpha \in \mathbb{R}$  and  $\lambda > 0$ . (Two functions, say  $b(t)$  and  $c(t)$ , are defined to be of the same order,  $b(t) \sim c(t)$ , if  $b(t)/c(t) \rightarrow 1$  as  $t \rightarrow \infty$ .) More generally still, lognormal-like tails can be written as

$$1 - F(x) \sim cx^\alpha e^{-\lambda \log^\delta x} \quad (4)$$

with  $\delta > 1$ . We denote this class  $LN(\delta)$ . Although the speed of tail decay of the lognormal distribution is faster than that of class (1), it is sufficiently slow to generate a tail that is commonly considered as “heavy”, i.e. for both distributional classes we have  $e^{\beta x}(1 - F(x)) = \infty$  for all  $\beta > 0$  as  $x \rightarrow \infty$ . Such tail decay is labelled *subexponential*. In the lognormal case, the speed of decay is faster than any power function, but also slower than exponential. This slow speed is partly at the origin of the confusing situation in the applied literature where the same data are given diametrically opposite interpretations. For our statistical theory, this will necessitate the use of asymptotic refinements.

We are not interested in testing for lognormality. More generally, we ask whether the decay speed is faster than any power function. For reasons that will become clear next, we will refer to this situation as Gumbel–Gibrat.

## 2.1. Maximum domains of attraction

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 of i.i.d. data: If, for suitably chosen norming constants  $c_n$  and  $d_n$ ,  $c_n^{-1}(\max(X_1, \dots, X_n) - d_n)$  converges in distribution to a non-degenerate distribution function, then this limit belongs to one of only three distribution functions, namely the Fréchet, Weibull, or Gumbel. Embedding these three possible outcomes in the Generalized Extreme Value distribution  $G_\gamma$ , the Fisher–Tippett theorem asserts the weak convergence of scaled maxima to

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

with  $\gamma > 0$  being the Fréchet case and  $\gamma \rightarrow 0$  the Gumbel case. As we consider  $x \rightarrow \infty$ , the negative Weibull case with  $\gamma < 0$  is irrelevant. Which limit case applies is thus given by the sign of  $\gamma$ , usually referred to as the extreme value index (and  $1/\gamma$  as the tail index).

Extensions of the Fisher–Tippett theorem consider maximum domains of attraction (MDA):  $F \in \text{MDA}(G_\gamma)$  if there exist norming constants such that the Fisher–Tippett theorem holds for extreme value distribution  $G_\gamma$  (the MDAs are characterized in e.g. Embrechts et al., 1997, Theorems 3.3.7 and 3.3.26); the hypothesis of their existence is usually labelled the extreme value condition. It

is well known that the subexponential lognormal distribution belongs to the MDA of the Gumbel distribution, as are distributions that decay exponentially fast (e.g. the Normal or the Gamma). By contrast distributions that are regularly varying, i.e. being of the form (1), are in the MDA of the Fréchet distribution. These include, e.g. the Hall model (2) or the Burr distribution. Zipf's weak law as formalized by (1) thus characterizes one *entire* MDA of the limit law.

We therefore label the hypothesis that the tail of the size distribution decays faster than power functions the Gumbel–Gibrat hypothesis. In particular, we seek to test

$$H_0 : F \in MDA(G_0) \quad v. \quad F \in MDA(G_{\gamma>0}) \quad (5)$$

We are not testing one parametric distribution (e.g. lognormal) against another (e.g. Pareto). If a researcher is willing to impose fully parametric assumptions, tailored tests are available in the literature (e.g. Malevergne et al., 2011). Other approaches seek to fit fully specified parametric models to the entire size data, notably subordinated i.i.d. growth processes yielding so-called double Pareto lognormal distributions (Reed, 2002), or switching models in which Pareto tails are pasted to lognormal bodies (Ioannides and Skouras, 2013). However, as regards our empirical applications, their framework cannot be used here: Neither the entire size distributions are exactly Pareto (i.e. they are not described by (3) for all sizes); nor, in terms of equation (1), is the nuisance function  $L_F$  a constant, nor is the lognormal quantile plot linear for large sizes. The generality of our hypotheses to be tested is attractive since it avoids misspecification errors induced by fully parametric approaches when the maintained hypotheses are rejected by the data. The test procedures we consider below do not impose any parametrizations other than the assumption that

the Fisher–Tippett theorem holds, so are semi-parametric in this sense. In order to address the challenges that arise from the slow decay in the nuisance part of the distribution function, we invoke a second order regular variation assumption.

### 3. TESTS OF THE GUMBEL–GIBRAT HYPOTHESIS

We study two complimentary approaches that have been proposed in the statistical literature for testing the Gumbel hypothesis: First, a direct test of the Gumbel–Gibrat hypothesis  $F \in MDA(G_0)$ , and then tests based on the estimated  $\gamma$ . The complimentary approaches are considered since it will turn out that none dominates the others in all circumstances considered below. Nor is it our intention to provide a comprehensive survey of all feasible estimators that have been proposed in the statistical literature.

Define the upper order statistics  $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(k)} \geq \dots \geq X_{(n)}$  where  $X_1, \dots, X_n$  is a sample of i.i.d. data. We consider only excesses over the random threshold  $X_{(k)}$ ;  $k$  is an intermediate sequence  $k = k(n)$  that is assumed to satisfy standard asymptotic monotonicity conditions:  $k(n) \rightarrow \infty$  and  $k(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ . A classic criterion for choosing  $k$  in applications is the minimisation of the MSE (see e.g. Hall, 1982, and Dekkers and de Haan, 1993). Rather than picking one particular value of  $k$ , we consider below a sequence of  $k$ . For our empirical applications, we discuss the choice of  $k$  in Section 4 below.

For the direct test approach, define the Greenwood-type statistic  $G_n(k) = M_{k,n}^{(2)}/(M_{k,n}^{(1)})^2$  with  $M_{k,n}^{(i)} = (1/k) \sum_{j=1}^k (X_{(j)} - X_{(k+1)})^i$ . The Hasofer–Wang test statistic (Hasofer and Wang, 1992; Neves and Fraga Alves, 2007) is a transformation, given by

$$W_n(k) = \frac{1}{k} \left( 1 - \frac{G_n(k) - 2}{1 + (G_n(k) - 2)} \right) \quad (6)$$

and equals the reciprocal of the empirical squared coefficient of variation.  $G_n(k)$  is a sample analogue of the moment expression  $E((X - t)^2 | X > t) / (E(X - t | X > t))^2$ , which equals 2 as  $t \rightarrow \infty$  when  $\gamma = 0$ . Under the alternative hypothesis  $\gamma > 0$ , Lemma 2 below requires  $\gamma < 1/4$  for convergence in distribution, otherwise  $\sqrt{k/4}(kW_n(k) - 1)$  diverges. We therefore should expect the test to be powerful for sufficiently large values of  $\gamma$ .

We also consider the problem of estimating  $\gamma$  directly before proceeding to the test. The well-known moment estimator due to Dekkers et al. (1989) is given by

$$\hat{\gamma}_M(k) = 1 + H_{k,n}^{(1)} - \frac{1}{2} \left( 1 - \frac{(H_{k,n}^{(1)})^2}{H_{k,n}^{(2)}} \right)^{-1} \quad (7)$$

with  $H_{k,n}^{(i)} = (1/k) \sum_{j=1}^k (\log X_{(j)} - \log X_{(k+1)})^i$ . Similar to the Greenwood-type statistic  $G_n(k)$ ,  $H_{k,n}^{(2)}/(H_{k,n}^{(1)})^2$  is a sample analogue of  $E((\log X - \log t)^2 | X > t) / (E(\log X - \log t | X > t))^2 (= 2$  as  $t \rightarrow \infty$  and  $\gamma > 0)$ . The log transform takes random variables in the domain  $G_{\gamma>0}$  into the domain  $G_0$ , ensuring that their means do not diverge.

Since the moment estimator is not invariant to shifts in the data, we consider also the estimator proposed in Matthys and Beirlant (2003). Considering the scaled log-ratio of spacings  $Y_j = j \log[(X_{(j)} - X_{(k)}) / (X_{(j-1)} - X_{(k)})]$ , and exploiting Rényi's representation of standard exponential order statistics, they propose

$$\hat{\gamma}_{MLE}(k) = \arg \max \sum_{i=1}^{k-1} (Y_j \lambda_{\gamma,i,k} - \log(\lambda_{\gamma,i,k})) \quad (8)$$

with  $\lambda_{\gamma,i,k} = 1 - (i/(k+1))^\gamma$ . They label this estimator a “maximum likelihood estimator” although it has this interpretation only asymptotically as no distributional assumptions are required apart from the distribution's belonging to an *MDA*.

The estimators  $\hat{\gamma}_M$  and  $\hat{\gamma}_{MLE}$  can be utilized to test the Gumbel–Gibrat hypothesis (5) in a straightforward way. The test statistic is  $T = \hat{\gamma} / SE(\hat{\gamma})$  where  $\hat{\gamma}$  is either the moment estimator (7) or the maximum likelihood estimator (8).

### 3.1. Distribution theory under second order regular variation

The distributional theory for  $W_n(k)$ ,  $\hat{\gamma}_M(k)$ , and  $\hat{\gamma}_{MLE}(k)$  is stated next, accounting for the potentially slow convergence speed in the Fisher–Tippett theorem (equivalently the slow speed of tail decay) that turns out to induce important biases. These will be quantified below in Figure 1. In the following, we provide second order refinements based on a second order regular variation condition. While the asymptotic theory for the estimators is known, we recast it here in a unifying way.

Define the tail quantile function  $U(t) = F^{-1}(1 - 1/t)$ . The content of the Fisher–Tippett theorem can be expressed equivalently in terms of first order regular variation and the tail quantile function (e.g. Dekkers et al., 1989): for  $\gamma \geq 0$  and if  $\lim_{t \rightarrow \infty} U(t) > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{\log U(tx) - \log U(t)}{a(t)/U(t)} = \log x \quad (9)$$

where  $a$  is a positive norming function with the property  $a(t)/U(t) \rightarrow \gamma$ . To account for the potentially slow convergence speed in the Fisher–Tippett theorem, we follow the extreme value literature and assume that the tail quantile function obeys the second order extended regular variation condition

$$\lim_{t \rightarrow \infty} \frac{\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \log x}{A(t)} = H_{\gamma, \rho}(x) \quad (10)$$

where  $H_{\gamma, \rho}(x) = \frac{1}{\rho} \left( \frac{x^{\rho + \gamma_-} - 1}{\rho + \gamma_-} - \frac{x^{\gamma_-} - 1}{\gamma_-} \right)$  with  $\gamma_- = \min\{0, \gamma\}$ , and  $H_{0,0}(x) = 0.5 \log^2 x = H_{\gamma > 0, 0}(x)$ . The parameter  $\rho \leq 0$  is the so-called second order parameter of regular variation, and  $A(t)$  is a rate function that is regularly varying with index  $\rho$ , with  $A(t) \rightarrow 0$  as  $t \rightarrow \infty$ . If  $\rho \rightarrow 0$ , the speed of decay in the nuisance part of the tail quantile function  $U$  is slow. The exposition below is simplified by using the equivalent representation (Dekkers et al., 1989, p. 1840),

$$\lim_{t \rightarrow \infty} \frac{\log U(tx) - \log U(t) - b_2(t) \log x}{b(t)} = H_{\gamma, \rho}(x) \quad (11)$$

where  $b_2(t) \sim a(t)/U(t)$ , so the new rate function satisfies  $b(t) \sim -A(t)a(t)/U(t)$ .

**Theorem 1.** Let  $\gamma = 0 = \rho$ , and let  $k = k(n) \rightarrow \infty$  and  $k(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ . Assume that (11) holds, and that  $\sqrt{k} \text{bias}(n, k) \rightarrow \lambda \neq 0$  as  $n \rightarrow \infty$ , where

$$\text{bias}(n, k) = \begin{cases} b_2(n/k) - \frac{b(n/k)}{b_2(n/k)} & \text{when considering } \hat{\gamma}_M, \hat{\gamma}_{MLE} \\ \hat{A}(n/k) & \text{when considering } W_n(k) \end{cases}$$

with

$$\hat{A}(t) = \begin{cases} A(t) = b(t)/b_2(t) & \text{if } b_2(t)^2/b(t) \rightarrow 0 \\ a(t)/U(t) = b_2(t) & \text{if } b_2(t)^2/b(t) \rightarrow \pm\infty \\ A(t)(1+c) & \text{if } b_2(t)^2/b(t) \rightarrow c \neq 0, \pm\infty. \end{cases}$$

Then

$$\sqrt{k}(\hat{\gamma}_M) \rightarrow^d N(\lambda, 1)$$

$$\sqrt{k}(\hat{\gamma}_{MLE}) \rightarrow^d N(\lambda, 1)$$

$$\sqrt{k/4}(kW_n(k) - 1) \rightarrow^d N(\lambda, 1).$$

### 3.1.1. Biases and the speed of tail decay

Next, we turn to bias functions. The following lemma states the bias functions for the key lognormal case. These results are new.

**Lemma 1.** Consider the lognormal distribution  $LN(2)$ . We have  $\gamma = 0 = \rho$ . As  $t \rightarrow \infty$ , to second order, the tail quantile function and rate functions are

$$\begin{aligned}
 U(t) &= \exp\left(\sigma \left[ (2 \log t)^{1/2} - \frac{\log \log t + \log 4\pi}{2(2 \log t)^{1/2}} \right]\right) \\
 a(t)/U(t) &= \sigma (2 \log t)^{-1/2} \left[ 1 + \frac{1}{2} (\log(4\pi) + \log(\log t) - 2) (2 \log t)^{-1} \right] \\
 A(t) &= -(2 \log t)^{-1} \left[ 1 + (2 \log t)^{-1} (\log(4\pi) + \log(\log(t)) - 3) \right]
 \end{aligned}$$

The bias functions are  $\hat{A}(t) = b_2(t) = a(t)/U(t)$  and  $b_2(t) - b(t)/b_2(t) = a(t)/U(t) + A(t)$ .

The distributional theory holds for all distributions that are in the MDA of the Gumbel distribution and which satisfy condition (10). In Online Appendix B we consider other distributions with  $\gamma = 0 = \rho$ , and obtain the following ranking of the biases:

**Corollary 1.** The biases in the lognormal case exceed those induced by the Gumbel, Gamma and Normal distributions. The biases for the normal distribution exceed those of the Gumbel and Gamma distribution. The magnitude of the bias function  $|b_2 - b/b_2|$  in the Weibull case exceeds those of the Gamma and the Gumbel distribution, and that of the normal distribution if  $\delta < 1/3$ ; in the Weibull case  $|\hat{A}(t)|$  is larger than for the Gumbel and Gamma cases if  $\delta < 1/2$  and larger than for the normal case if  $\delta < 1/3$ .

It is in this sense that the lognormal case constitutes a “worst case” scenario for the tests of  $F \in MDA(G_0)$  under  $H_0$ . This corollary suggests a sensitivity analysis, as considered in Section 3.2.1

below. Moreover, in the lognormal case, the slow speed of decay of the bias is compounded by large values of  $\sigma$  (as encountered in our empirical applications).

**Corollary 2.** *The biases in the lognormal case LN(2) and hence the test size distortions of the uncorrected test statistics are increasing in  $\sigma$ .*

### 3.1.2. Power considerations

We briefly consider the power of the tests and in particular, the role played by  $\rho$  driving the higher order biases which lead to power distortions.

**Lemma 2.** *Let  $\gamma > 0$ . Let  $k = k(n) \rightarrow \infty$  and  $k(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ . Assume that (11) holds, and that  $\sqrt{k} \text{bias}(n, k) \rightarrow \lambda \neq 0$  as  $n \rightarrow \infty$ , where*

$$\text{bias}(n, k) = \begin{cases} \frac{\gamma + \rho(1 - \gamma)}{\rho(1 - \rho)^2} A(n/k) & \text{when considering } \hat{\gamma}_M, \hat{\gamma}_{MLE} \text{ with } \rho < 0 \\ \frac{1 - \gamma}{(1 - \gamma - \rho)(1 - 2\gamma\rho)} A(n/k) & \text{when considering } W_n(k) \end{cases}$$

Then

$$\sqrt{k}(\hat{\gamma}_M - \gamma) \xrightarrow{d} N(\lambda, 1 + \gamma^2)$$

$$\sqrt{k}(\hat{\gamma}_{MLE} - \gamma) \xrightarrow{d} N(\lambda, 1 + \gamma^2)$$

Also, for  $\gamma < 1/4$ ,

$$-\sqrt{\frac{k}{4}}(kW_n(k) - 2\gamma) \xrightarrow{d} N(\lambda, 1 + \gamma^2).$$

The asymptotic power function is given, for a one-sided test of size  $\alpha$ , by

$$\pi(k, \gamma, \lambda) = 1 - \Phi \left( \frac{\Phi^{-1}(1 - \alpha) - k^{1/2}\gamma}{(1 + \gamma^2)^{1/2}} - \lambda \right),$$

where  $\Phi$  denotes the normal cdf. As the bias  $\lambda$  increases, the power is distorted upwards.

In particular, in the Hall class (2),  $U(x) = c^\gamma(1 + dx^\rho + o(x^\rho))$  with  $\rho = \gamma\beta < 0$ , and  $A(t) = (\rho^2/\gamma)dt^\rho$ . Second order regular variation leads to power distortions. As  $\rho \uparrow 0$ , the nuisance part of  $F$  decays more slowly, and the biases tend to increase; the resulting overestimation leads to an overstatement of the power since  $\partial\pi/\partial\lambda > 0$ . Simulation evidence is presented in Online Appendix A.1.

### 3.2. Simulation evidence

Theorem 1 suggests that higher order regular variation properties of distributions can give rise to biases, which, if ignored, lie at the heart of the test size distortion. If the functions  $A$  and  $a$  were known, the theorem suggests a bias correction that should re-align nominal and empirical test size behavior. We examine these implications for the focal lognormal case (and in Online Appendix B for Weibull, Gumbel, Gamma, and Normal distributions).

The simulation design involves drawing samples of size  $N = 10,000$ , and repeating the experiments  $R = 1,000$  times. The performance of the three test procedures is assessed using the coverage error rate of their confidence intervals for the Gumbel–Gibrat null  $\gamma = 0$ , for a nominal 5% rate.

Figure 1 depicts the results. Consider first the test-statistics, depicted in Panels A–C. In each case considered, the uncorrected test statistic exhibits a considerable bias (solid lines), while its variability is stable (weak dashed line) and close to the theoretical value of unity. The analytical bias functions (bold dashed lines) capture the empirical biases very well. The differences between empirical bias and the value of the bias function of Lemma 1 decrease in  $k$ , and are negligible for  $k > 50$ .

In Panels D–F of Figure 1 we consider the implications of the distortions and the proposed corrections for the empirical test sizes for the nominal rate of 5%. First, we plot the empirical test sizes using the uncorrected test statistics (solid lines). As the biases have been shown to be very substantial, it is not surprising that the empirical test sizes using the uncorrected test statistics quickly diverge to 100%. By contrast, using the appropriate bias corrections (bold dashed lines) successfully re-aligns nominal and empirical test sizes.<sup>3</sup> Figure 1 illustrates thus a key result of this paper.

### 3.2.1. Sensitivity analysis

In practice,  $(a, A, \rho)$  are not known and difficult to estimate. We therefore suggest a sensitivity analysis that is informed by the preceding theory. The overall concern is that the distortions could be sufficiently large as to invalidate the conclusion of the test of the null hypothesis  $\gamma = 0$ , as

<sup>3</sup>As regards the “maximum likelihood estimator”  $\hat{\gamma}_{MLE}$  in Panel F, the empirical rate exceeds the nominal rate but is stable. Panel C reveals an elevated variability of the estimator (which exceeds the theoretical value of 1). Including the appropriate scale correction now lets actual and nominal coverage error rates coincide (dotted line in Panel F).

demonstrated by the preceding simulation study. Under the null hypothesis the biases are such that the uncorrected Hasofer–Wang test statistic constitutes a lower bound, and the  $\gamma$  estimates and their lower confidence limits an upper bound (illustrated in Figure 1 for the lognormal case). Correcting the distortions therefore increases the former and reduces the latter. Corollary 1 shows that the lognormal distribution constitutes the “worst case” situation, Lemma 1 defines the theoretical biases in this case, while Corollary 2 reveals that the distortions are increasing in  $\sigma$ . We therefore interpret  $\sigma$  more generally as a sensitivity parameter that controls a very conservative upper bound. The proposal for the sensitivity analysis is then to set an arbitrary yet “reasonable” value for  $\sigma$  to establish whether the rejection of the null hypothesis is overturned, or to find the critical value of  $\sigma$  above which rejection of the null hypothesis is overturned. If, for “reasonable” values of  $\sigma$ , the test conclusions cannot be overturned, it follows by Corollary 1 that the tail of the size distribution decays faster than lognormal as the resulting distortions are then smaller. It is, of course, conceivable that the correction is excessive and the test conclusion wrongly overturned. We would consider any situation in which the conclusion based on the uncorrected test statistic is overturned as yielding an inconclusive result. Fortunately, this situation does not arise in our empirical examples below, as both uncorrected test and our sensitivity analysis point to the same, and hence robust, conclusion.

#### 4. EMPIRICAL APPLICATIONS

We apply the discussed methods to investigate the size distributions of cities in Germany and the US, and the size distribution of firms in France. In all three cases, we test the Gumbel–Gibrat

null hypothesis for the upper tail of the distributions.<sup>4</sup> As discussed in the Introduction, applied researchers generally ignore the potentially catastrophic distortions that we have studied above. Here we provide *robust* evidence for power function behavior.

#### 4.1. City size distribution in Germany

Our first empirical application concerns the size distribution of cities. We use first an administrative dataset for Germany for the year 2000, provided by the German Federal Statistical Office. 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, and the year 2000 size distribution comprises 13,854 cities.

We present first some descriptive graphs. Figure 2 Panel A depicts the kernel density estimate of the entire German log city size distribution. As the density is clearly not exponential, it follows that the entire size distribution is not Pareto. Hence clearly rejected already are the exact Pareto law (3) for the entire size distribution and thus Zipf’s strongest law (as given by (3) with  $\gamma = 1$ ). In order to describe the upper tail behavior, we present in Panel B a Pareto quantile plot; i.e. we plot  $\log(X_{(j)})$  on  $(-\log(j/(n+1)))$  for  $j = 1, \dots, k$  with  $k = 300$ . If the distribution were exactly Pareto, the plot would be linear; if only the tail were (exactly) Pareto so that  $L_F(x) = c$  in (1), most

<sup>4</sup>In Online Appendix C.1 we consider in detail the lower tails, which are of interest since some data generating processes imply a Pareto-type tail at the lower end of the distribution (e.g. Reed, 2002; Malevergne et al., 2013). There we show that Pareto-type behavior at the lower tails of the city size distributions cannot be confirmed.

points on the right part of the graph would be on a line. Since the graph is not exactly linear, we refrain from assuming exact Pareto behavior in the tails of the size distribution. Finally we present in Panel C also the normal quantile plot for large log sizes, having rescaled  $\log(X_{(j)})_{j=1,\dots,300}$  using the maximum likelihood estimates of the mean and variance of the assumed lognormal distribution (with  $\hat{\mu} = 7.3$  and  $\hat{\sigma} = 1.5$ ). Again, the graph is not linear (nor is it on the diagonal), so we decide not to presuppose exact lognormality for the upper tail. In view of Panels B and C, our data suggest not to conduct tests such as proposed in Malevergne et al. (2011), who consider testing a Pareto null hypothesis against a lognormal alternative for large sizes.

We proceed therefore to the application of our test procedure, and we consider first the uncorrected test statistics, which are depicted in Figure 3. Panel A depicts the uncorrected Hasofer–Wang test statistics (solid line), and suggests a rejection of the null hypothesis for values of  $k$  as low as 50. The test procedures based on the uncorrected estimates of  $\gamma$ , suggest a similarly clear conclusion.<sup>5</sup> In Panels B and C we depict  $\hat{\gamma}_M$  and  $\hat{\gamma}_{MLE}$  (solid lines). After some initial fluctuations for very low  $k$ , both estimators agree and stabilise around a value of .72 which also implies an infinite variance but finite mean of the size distribution). The uncorrected confidence band (not depicted) is sufficiently tight to suggest a clear rejection of the Gumbel–Gibrat null hypothesis  $\gamma = 0$ . Interestingly it also excludes the Zipf value of 1 for  $k > 50$ , which means that Zipf’s stronger law (i.e. (1) with  $\gamma = 1$ )

<sup>5</sup>If  $\gamma > 0$ , alternative estimators of  $\gamma$  become available. In Online Appendix C.2 we compare our estimates to the Hill and rank-1/2 (log-log rank-size regression) estimates. These suggest the same qualitative conclusions across our city and firm size data.

is also rejected. Similar qualitative and quantitative conclusions apply to the test based on  $\hat{\gamma}_{MLE}$ , depicted in Panel C (solid line).

Although the uncorrected test statistics suggest a clear rejection of the Gumbel–Gibrat null hypothesis  $\gamma = 0$ , we have seen that higher order regular variation properties can lead to catastrophic test size distortions. We therefore conduct a sensitivity analysis, as discussed in Section 3.2.1 above. Specifically, we consider what would happen in the “worst case” distribution (under the null hypothesis), i.e. we consider the bounds on our test statistic or our lower confidence limits that would arise from subexponential lognormal distributions. We treat  $\sigma$  as a sensitivity parameter, and set it arbitrarily but “reasonably” to 1.5 (equal to the maximum likelihood estimate reported above).

We consider first the Hasofer–Wang test statistic. The uncorrected test statistic constitutes a lower bound. The conservative upper bound thus obtained is depicted in Figure 3 Panel A (dashed line). The Gumbel–Gibrat hypothesis can then only be rejected at the 5% level for roughly  $k > 300$ . As regards the  $\hat{\gamma}$ -based procedures depicted in panels B and C, the uncorrected estimates constitute upper bounds, and the bias adjustments reduce the  $\gamma$  estimates (dashed lines) considerably. But the reduction is not sufficient to overturn the conclusion that the null hypothesis be rejected: The shifted lower 95%-confidence limits (dotted lines) still exclude the value 0 for  $k > 70$  in both cases.

Our sensitivity analysis has suggested that, relative to the bounds established by the uncorrected test statistics, the number of upper order statistics to be taken into account should modestly increase.

The null hypothesis for the Hasofer–Wang test is robustly rejected for  $k > 300$ , and for  $k > 70$  for the  $\gamma$ -estimator based methods.

Should a researcher wish to pick a particular  $k$ , specifically to pin down a point estimate for  $\gamma$ , established methods tend to suggest values of  $k$  for our data that are not in contradiction with the suggestions above, i.e. they yield sufficiently large values of  $k$ . For instance, one visual method of choosing  $k$  is to consider an extended area of the  $\hat{\gamma}(k)$  plot in which the estimate is rather stable. For the moment estimator depicted in Figure 3b, this happens around  $k \in [100, 250]$ . Given this extended stability, it is not surprising that alternative methods yield similar values. For instance, consider minimizing the mean squared error (mse) based on Lemma 2, which requires knowledge of the theoretical bias function. Assuming a Burr distribution with  $\rho = -.3$  and replacing  $\gamma$  by its estimate, yields a mse-minimizing value of  $k = 135$ . Increasing the magnitude of the second order parameter  $\rho$  increases  $k$ .

#### 4.2. City size distribution in the US

For our examination of the US city size distribution, we use the US census 2000 data of city sizes used and extensively described in Eeckhout (2004). The number of observations is  $n = 25,358$ . The density of the US size distribution is similar to the German, and therefore not displayed. The results are displayed depicted in Figure 4. The uncorrected point estimators of  $\gamma$  for the upper tail (solid lines in Panels B and C) quickly stabilize around a value of  $.68 > 0$  suggesting that the Gumbel–Gibrat hypothesis should be rejected for the upper tail (and also implies an infinite variance but finite mean of the size distribution). The same is true for the uncorrected Hasofer–Wang test.

Turning to our sensitivity analysis, fitting a lognormal density results in estimated  $\sigma$  of 1.75. The bias adjusted moment estimates (Panel B, dashed line) are considerably smaller than the uncorrected estimates. The lower 95% confidence bounds (dotted line) crosses the abscissa at  $k = 100$ . Hence, the Gumbel–Gibrat hypothesis is robustly rejected for  $k > 100$ . The same results holds for the MLE (Panel C). As to the bias adjusted Hasofer–Wang test (Panel A), the Gumbel–Gibrat hypothesis is rejected for  $k > 250$ . In any case, we also clearly reject the stronger forms of Zipf’s law (equation (1) with unity exponent).

#### **4.3. Firm size distribution in France**

As a second empirical application we consider the firm size distribution in France. As in the literature on the city size distribution, the literature on the firm size distribution contains conflicting views. Some studies postulate an exponential or lognormal distribution, see e.g. the review in Sutton (1997). Others assume a Pareto upper tail; this strand of the literature goes back at least to Simon and Bonini (1958). Empirical evidence in support of the Pareto law is presented e.g. in Axtell (2001). By contrast, using data for Portugal, Cabral and Mata (2003) argue that the lognormal distribution provides a good fit only for mature firms: As firms age, constraints on the firm size diminish, and the size distribution evolves over time to a lognormal distribution.

The data are extracted from the “Amadeus” dataset provided by Bureau van Dijk. Firm size is measured by the number of employees in 2015, and  $n = 17,757$  observations. The distribution is highly skewed with mean 125 employees but a median of only 14.

Figure 5 Panel A shows the density plot of the logarithm of the number of employees. As in the case of city sizes, the strongest form of Zipf's law (i.e. exact Pareto behavior) does obviously not hold. Estimating the parameter  $\sigma$  of a lognormal distribution by maximum likelihood, we obtain an estimate of 1.2. Panels C and D of Figure 5 depict the moment estimator and the maximum likelihood estimators of  $\gamma$ . Both methods yield estimates exceeding 1 even for moderate values of  $k$ . The dashed lines show the bias adjusted estimators. The dotted lines are the corresponding lower 95% confidence limits. Even for small values of  $k$  (say,  $k = 30$ ) we can reject the Gumbel–Gibrat hypothesis as  $\hat{\gamma}_M$  and  $\hat{\gamma}_{MLE}$  are significantly larger than 0. For  $k$  between about 50 and 120 both confidence intervals enclose the value of 1 supporting Zipf's strong law (and an infinite mean of the size distribution); for larger  $k$ , this result continues to hold for the moment estimator but the estimates increase above 1 in the maximum likelihood case. The Hasofer–Wang test statistic, plotted as a solid line in Panel B, indicates that the Gumbel–Gibrat null of  $\gamma = 0$  can be rejected for  $k$  as small as about 30. Taking the bias adjustment into account, the dashed line shows that the null can be rejected for  $k > 80$ .<sup>6</sup>

#### 4.4. Asymptotic distortions when $\gamma > 0$

Our methods have provided robust empirical evidence for power tail behavior of city and firm size distributions. If the researcher seeks to obtain a particular point estimate of  $\gamma$ , Lemma 2 suggests that care needs to be exercised. Although  $A(t) \rightarrow 0$  as  $t \rightarrow \infty$ , this decay might be slow, inducing

<sup>6</sup>Cabral and Mata (2003) suggest that the size distribution depends on firms' age, in particular that the lognormal distribution tends to provide a better fit for older firms. Looking only at firms incorporated before 1999 (roughly half of the firms), the estimation results hardly change and are therefore not shown. Hence, the Gumbel–Gibrat hypothesis is also robustly rejected for older firms.

a higher order bias. An asymptotic bias arises if  $k$  increases too quickly relative to the tail decay so that  $\lambda \neq 0$ . In Online Appendix C.3 we investigate the potential size of the distortion in Hall-type models, for which the sign of the distortion is given by  $\text{sgn}(\gamma + \rho(1 - \gamma))$ . Hence the estimators suffer a positive distortion if  $\rho$  is sufficiently small in magnitude, or if  $\gamma > 1$ . Quantitatively, we show that the distortions are small for  $|\rho| > .5$ . For smaller magnitudes, the distortions are bigger, but this situation is less likely in our data. However, in view of the positive sign of the distortion our inference that  $\gamma < 1$  is strengthened. Finally, note that alternative estimators of  $\gamma$  are likely to suffer similar higher order biases, yet practitioners usually tend to ignore such potential pitfalls.

## 5. CONCLUSION

We have recast key issues of the size distribution debate in a unifying statistical framework based on extreme value theory, that has led us to a test of whether the tail of the size distribution decays faster than any power function. Traditionally, the size distribution literature has focused on the juxtaposition of two fully specified parametric models (such as exact lognormal and Pareto); yet, the conditions for the ensuing tests are not met by our data. By contrast, our tests consider restrictions on the speed of tail decay that are less restrictive, while nesting the leading “laws” of the literature. Our test framework is therefore not only of theoretical interest but has also empirical necessity. We have addressed the practical complications arising from the higher order regular variation properties of the size distribution, which, if ignored, could lead to catastrophic distortions. If the theoretical bias functions are known, nominal and theoretical test sizes can be re-aligned successfully. If the theoretical functions are not known, our results inform a sensitivity analysis.

In our substantive empirical illustrations, we consider the size distribution of cities (in the US and Germany) and firms (in France) and conclusively demonstrate that the upper tails decay like power functions. While a unity exponent of the power function cannot be rejected for firm sizes, the tests unambiguously reject this stronger forms of Zipf's law for city sizes. These are important findings since the rejection of the classic Gibrat's law implies that some leading approaches to generative laws of invariant size distributions are empirically compromised. At the same time, the class of generative laws to be considered should be extended, since power law behavior of the upper tail is in fact one of the outcomes of the Fisher–Tippett theorem for the largest sizes.

Finally, we recall that our formal analysis is based on the hypothesis that the size data are i.i.d. (as is standard in the empirical literature). Results in Hsing (1991) suggest that the current theory might be a reasonable guide if the dependence is sufficiently weak so that approximations to a normal law still hold (see also Resnick and Starica, 1998). In Online Appendix D this conjecture is verified for dependent data generating processes considered in Gabaix and Ibragimov (2011). For a recent theoretical analysis of tail index estimation for dependent data see e.g. Hill (2010). Extensions of the current framework to dependent data will be the subject of future work.

#### **ACKNOWLEDGEMENT**

We thank the editor and our referees for their constructive comments that have helped to improve the paper. Financial support from ANR-DFG (grant ANR-15-FRAL- 0007-01) is also gratefully acknowledged.

## REFERENCES

- Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, 293:1818–1820.
- Barro, R. and Jin, T. (2011). On the size distribution of macroeconomic disasters. *Econometrica*, 79(5):1567–1589.
- Cabral, L. and Mata, J. (2003). On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93(4):1075–1090.
- Córdoba, J. C. (2008). A generalized Gibrat's law. *International Economic Review*, 49(4):1463–1468.
- Dekkers, A. and de Haan, L. (1993). Optimal choice of sample fraction in extreme value estimation. *Journal of Multivariate Analysis*, 47:173–195.
- Dekkers, A., Einmahl, J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17:1833–1855.
- Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from French firms. *Econometrica*, 79(5):1453–1498.
- Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.
- Fisher, R. and Tippett, L. (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. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1:255–293.
- Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. (2006). Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121:461–504.
- Gabaix, X. and Ibragimov, R. (2011). Rank - 1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business and Economic Statistics*, 29(1):24–39.
- Gabaix, X. and Ioannides, Y. (2004). The evolution of city size distributions. In Henderson, J. and Thisse, J.-F., editors, *Handbook of Regional and Urban Economics, Volume 4: Cities and Geography*. Elsevier, Amsterdam.
- Gibrat, R. (1931). *Les inégalités économiques*. Librairie du Recueil Sirey, Paris.
- Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability*, 1(1):126–166.
- Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society B*, 44:37–42.
- Hasofer, A. M. and Wang, Z. (1992). A test for extreme value domain of attraction. *Journal of the American Statistical Association*, 87(417):171–177.
- Hill, J. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, 26:1398–1436.
- Hsing, T. (1991). On tail index estimation using dependent data. *Annals of Statistics*, 19:1547–1569.

- Ibragimov, M., Ibragimov, R., and Kattuman, P. (2013). Emerging markets and heavy tails. *Journal of Banking and Finance*, 7:2546–2559.
- Ibragimov, M., Ibragimov, R., and Walden, J. (2015). *Heavy-tailed Distributions and Robustness in Economics and Finance*. Springer.
- Ioannides, Y. and Skouras, S. (2013). US city size distributions: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, 73:18–29.
- Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248.
- Levy, M. (2009). Gibrat’s law for (all) cities: Comment. *American Economic Review*, 99(4):1672–1675.
- Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122:1103–1144.
- Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, 83:036111–1–036111–11.
- Malevergne, Y., Saichev, A., and Sornette, D. (2013). Zipf’s law and maximum sustainable growth. *Journal of Economic Dynamics and Control*, 37(6):1195–1212.
- Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 13:853–880.
- Mitzenmacher, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:226–251.

- Neves, C. and Fraga Alves, M. I. (2007). Semi-parametric approach to the Hasofer-Wang and Greenwood statistics in extremes. *Test*, 16(2):297–313.
- Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.
- Reed, W. J. and Hughes, B. D. (2002). From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, 66:067103–1–067103–4.
- Resnick, S. and Starica, C. (1998). Tail index estimation for dependent data. *Annals of Statistics*, 8:1156–1183.
- Schluter, C. and Trede, M. (2002). Tails of Lorenz curves. *Journal of Econometrics*, 109:151–166.
- Schluter, C. and Trede, M. (2008). Identifying multiple outliers in heavy-tailed distributions with an application to market crashes. *Journal of Empirical Finance*, 15:700–713.
- Schluter, C. and Trede, M. (2016). Weak convergence to the Student and Laplace distributions. *Journal of Applied Probability*, 53:121–129.
- Simon, H. A. and Bonini, C. P. (1958). The size distribution of business firms. *American Economic Review*, 48(4):607–617.
- Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35:40–59.
- Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Cambridge, MA.

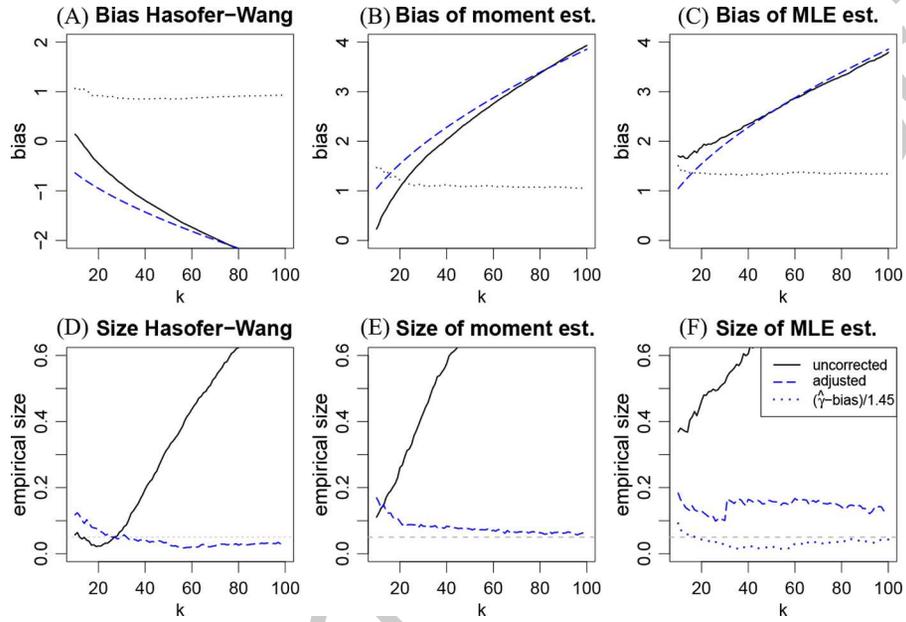


Figure 1: Bias and test size distortions in the lognormal case LN(2).

The top row depicts the mean (solid line) and standard deviation (dotted line) of the Hasofer-Wang test statistic (A), the moment estimator (B), and the MLE estimator (C); the long dashed lines show the theoretical biases. The bottom row shows the empirical sizes (solid lines) of the three approaches at significance level 5% along with the bias adjusted sizes (dashes).

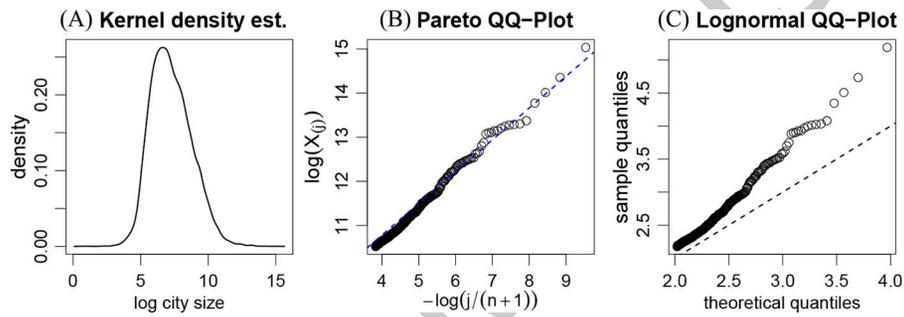


Figure 2: German city size distribution descriptives: Body and upper tail.

Panel A depicts the kernel density estimate of the log city size in Germany (with bandwidth 0.2).

Panel B shows the Pareto QQ plot, Panel C shows the lognormal QQ plot.

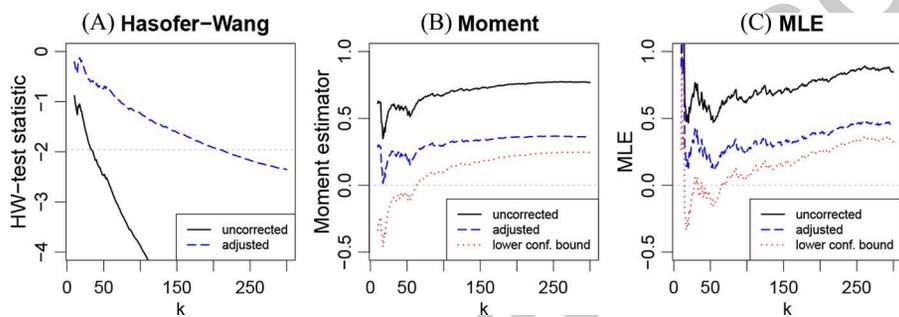


Figure 3: City size distribution in Germany: Tests of the Gumbel–Gibrat hypothesis.

Panel A shows the uncorrected (solid line) and the bias adjusted (dashes) Hasofer–Wang test statistics as a function of  $k$  along with the critical level (short dashes). Panels B and C show the uncorrected (solid lines) and bias adjusted (dashes) moment estimate (Panel B) and maximum likelihood estimate (Panel C) for varying  $k$ , the dotted line is the lower 95% confidence level.

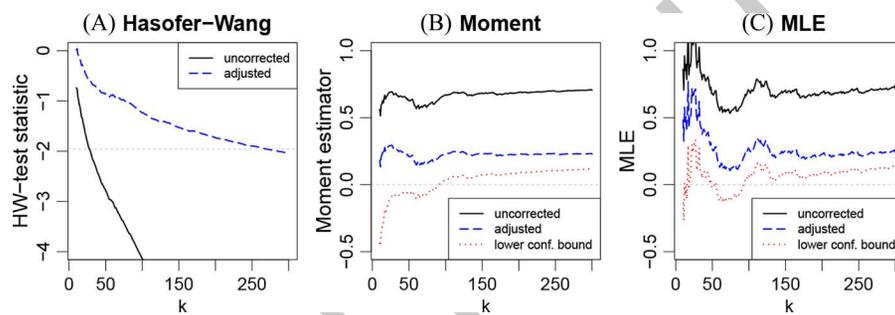


Figure 4: City size distribution in the U.S.: Tests of the Gumbel-Gibrat hypothesis.

Notes as per Figure 3.

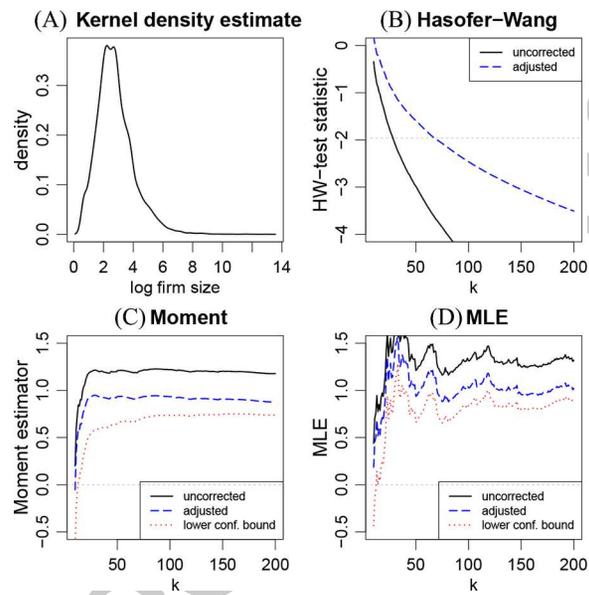


Figure 5: Firm size distribution in France.

Panel A depicts the kernel density estimate of the log firm size in France (with bandwidth 0.2).

Panels B-D: Notes as per Figure 3.