RANK−1/2: A SIMPLE WAY TO IMPROVE THE OLS ESTIMATION OF TAIL EXPONENTS*

Xavier Gabaix

Stern School of Business, New York University, and NBER

Rustam Ibragimov

Department of Economics, Harvard University

May 26, 2009

Address for manuscript correspondence:

Xavier Gabaix; Department of Finance, Stern School of Business, New York University, 44 West Fourth St., Suite 9-190, New York, NY 10012-1126; *Email*: xgabaix@stern.nyu.edu; Phone: (212) 998-0257; Fax: (212) 995-4233

Rustam Ibragimov; Department of Economics, Harvard University, Littauer Center, 1805 Cambridge St., Cambridge, MA 02138; *Email*: ribragim@fas.harvard.edu; Phone: (617) 496-4795; Fax: (617) 495-7730

*We thank the Editor, Serena Ng, an anonymous Associate Editor, two anonymous referees, Gary Chamberlain, Victor Chernozhukov, Graham Elliott, Wolfgang Härdle, Yoko Konishi, Hu McCulloch, Marcelo Moreira, Yoshihiko Nishiyama, Peter Phillips, Artem Prokhorov, Sidney Resnick, Gennady Samorodnitsky, Jim Stock, Jonathan Thong, Charles van Marrewijk, Yanping Yi and the participants at seminars at the Departments of Economics at Concordia University, Harvard University, Institute of Mathematics of Uzbek Academy of Sciences, Massachusetts Institute of Technology, Tashkent State University and Tashkent State University of Economics for helpful comments and suggestions. Ibragimov gratefully acknowledges partial research support by the NSF grant SES-0820124, a Harvard Academy Junior Faculty Development grant and the Clark Fund (Department of Economics, Harvard University).

ABSTRACT

Despite the availability of more sophisticated methods, a popular way to estimate a Pareto exponent is still to run an OLS regression: $\log(\text{Rank}) = a - b\log(\text{Size})$, and take b as an estimate of the Pareto exponent. The reason for this popularity is arguably the simplicity and robustness of this method. Unfortunately, this procedure is strongly biased in small samples. We provide a simple practical remedy for this bias, and propose that, if one wants to use an OLS regression, one should use the Rank $-1/2$, and run $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$. The shift of $1/2$ is optimal, and reduces the bias to a leading order. The standard error on the Pareto exponent ζ is not the OLS standard error, but is asymptotically $(2/n)^{1/2}\zeta$. Numerical results demonstrate the advantage of the proposed approach over the standard OLS estimation procedures and indicate that it performs well under dependent heavy-tailed processes exhibiting deviations from power laws. The estimation procedures considered are illustrated using an empirical application to Zipf's law for the U.S. city size distribution.

KEYWORDS: power law, heavy-tailedness, OLS log-log rank-size regression, bias, standard errors, Zipf's law

JEL Classification: C13, C14, C16

1 Introduction

Last four decades have witnessed rapid expansion of the study of heavy-tailedness phenomena in economics and finance. Following the pioneering work by Mandelbrot (1960, 1963) (see also Fama, 1965, and the papers in Mandelbrot, 1997), numerous studies have documented that time series encountered in many fields in economics and finance are typically thick-tailed and can be well approximated using distributions with tails exhibiting the power law decline

$$P\Big(Z > s\Big) \sim Cs^{-\zeta}, \;\; C, s > 0. \tag{1.1}$$

with a tail index $\zeta > 0$ (see the discussion in Čížek, Härdle and Weron, eds, 2005; Rachev, Menn and Fabozzi, 2005; Gabaix, Gopikrishnan, Plerou and Stanley, 2006, and references therein). Here $f(s) \sim g(s)$ means that $f(s) = g(s)(1 + o(1))$ as $s \to \infty$. Throughout the paper, C denotes a positive constant, not necessarily the same from one place to another. Let

$$Z_{(1)} \geq ... \geq Z_{(n)} \tag{1.2}$$

be decreasingly ordered observations from a population satisfying power law (1.1). Despite the availability of more sophisticated methods (see, among others, the reviews in Embrechts, Klüppelberg and Mikosch, 1997, and Beirlant, Goegebeur, Teugels and Segers, 2004), a popular way to estimate the Pareto exponent ζ is still to run the following OLS log-log rank-size regression with $\gamma = 0$:

$$\log\,(t - \gamma) = a - b \log\; Z_{(t)}, \tag{1.3}$$

or, in other words, calling t the rank of an observation, and $Z_{(t)}$ its size:

$$\log\,(\text{Rank} - \gamma) = a - b \log\,(\text{Size})$$

(here and throughout the paper, $\log(\cdot)$ stands for the natural logarithm).With N denoting the total number of observations, regression (1.3) with $\gamma = 0$ is motivated by the approximate linear relationships $\log\left(\frac{t}{N}\right) \approx \log(C) - \zeta \log\left(Z_{(t)}\right)$, $t = 1, ..., n$, implied by the empirical analogues of relations (1.1). The reason for the popularity of the OLS approach to tail index estimation is arguably the simplicity and robustness of this method. In various frameworks, the log-log rank-size regressions of form (1.3) in the case $\gamma = 0$ and closely related procedures were employed, among other works, in Rosen and Resnick (1980), Alperovich (1989), Krugman (1996), Eaton and Eckstein (1997), Brakman, Garretsen, van den Berg and van Marrewijk (1999), Dobkins and Ioannides (2000), Davis and Weinstein (2002), Levy (2003), Levy and Levy (2003), Helpman, Melitz and Yeaple (2004), Soo (2005) and Klass, Biham, Levy, Malcai and Solomon (2006). Further examples and the discussion of the OLS approach to the tail index estimation are provided in Persky (1992), Gabaix, Gopikrishnan, Plerou and Stanley (2003), Gabaix et al. (2006), Eeckhout (2004), Gabaix and Ioannides (2004) and Rossi-Hansberg and Wright (2007).

Let $\hat{b}_n$ denote the usual OLS estimator of the tail index ζ using regression (1.3) with $\gamma = 0$ and let $\hat{b}_n^\gamma$ denote the OLS estimator of ζ in general regression (1.3).

It is known that the OLS estimator $\hat{b}_n$ in the usual regression (1.3) with $\gamma = 0$ is consistent for ζ. However, the standard OLS procedure has an important bias. This paper shows that the bias is optimally reduced (up to leading order terms) by using $\gamma = 1/2$. Therefore, we recommend that, when using a log-log regression, one should always use $\log(\text{Rank} - 1/2)$ rather than $\log(\text{Rank})$.

We further show that the standard error of the OLS estimator $\hat{b}_n^\gamma$ of the tail index ζ in general regression (1.3) is asymptotically $(2/n)^{1/2}\zeta$. The OLS standard errors in log-log rank-size regressions (1.3) considerably underestimate the true standard deviation of the OLS tail index estimator. Consequently, taking the OLS estimates of the standard errors at the face value will lead one to reject the true numerical value of the tail index too often.

The 1/2 shift actually comes from more systematic results, in Theorems 1 and 2, which show that it is optimal and further demonstrate that the following asymptotic expansions hold for the general OLS estimator $\hat{b}_n^\gamma$:

$$E\Big(\hat{b}_n^\gamma/\zeta - 1\Big) = \frac{(2\gamma - 1)\log^2 n}{4n} + o\Big(\frac{\log^2 n}{n}\Big),$$

$$\hat{b}_n^\gamma/\zeta = 1 + \sqrt{\frac{2}{n}}\mathcal{N}(0,1) + O_P\Big(\frac{\log^2 n}{n}\Big)$$

(here and throughout the paper, for $\mu \in \mathbf{R}$ and $\sigma > 0$, $\mathcal{N}(\mu, \sigma^2)$ stands for a normal random variable (r.v.)

with mean μ and variance σ^2). We conclude that, when estimating the tail index ζ with an OLS regression, one should always use the regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$, with the standard error of the OLS estimator $\hat{b}_n$ of the slope given by $\sqrt{\frac{2}{n}}\hat{b}_n$.

We further provide similar asymptotic expansions for the tail index estimator $\hat{d}_n^\gamma$ in the dual to (1.3) regression

$$\log(Z_{(t)}) = c - d\log(t - \gamma) \tag{1.4}$$

(that is, $\log(\text{Size}) = c - d\log(\text{Rank} - \gamma)$), with logarithms of ordered sizes regressed on logarithms of shifted ranks. As follows from Theorems 1 and 2, the approaches to the tail index inference using regressions (1.3) and (1.4) are equivalent in terms of the small sample biases and standard errors of the estimators. The paper also discusses asymptotic expansions in the analogues of regressions (1.3) and (1.4) with the logarithms of shifted ranks $\log(t - \gamma)$ replaced by harmonic numbers (Section 3).

Numerical results indicate that the proposed tail index estimation procedures perform well for heavy-tailed dependent processes exhibiting deviations from power law distributions (1.1) (see Section 4). They further demonstrate the advantage of the new approaches over the standard OLS log-log rank-size regressions (1.3) and (1.4) with $\gamma = 0$.

The tail index estimation methods proposed in the paper are illustrated using an empirical analysis of Zipf's power law for the U.S. city size distribution (Section 5).

In recent years, several studies have focused on the analysis of asymptotic normality of the OLS tail index estimators in regressions (1.4) with $\gamma = 0$ and logarithms of ordered observations $\log(Z_{(t)})$ regressed on logarithms of ranks (see, among other works, the review in Ch. 4 in Beirlant et al., 2004). Such an approach to estimation of the tail shape parameters was introduced by Kratz and Resnick (1996) who refer to it as QQ-estimator. Nishiyama, Osada and Sato (2008) discuss asymptotic normality of the OLS tail index estimator in the regression of $\log(Z_{(t)})$ on $\log\ t$. Schultze and Steinebach (1999) consider closely related problems of least-squares approaches to estimation for data with exponential tails (see also Aban and Meerschaert, 2004, who discuss efficient OLS estimation of parameters in shifted and scaled exponential models). Kratz and Resnick (1996) establish consistency and asymptotic normality of the QQ-estimator in the case of populations with regularly varying tails. Their results demonstrate that in the case of populations in the domain of attraction of power law (1.1), the standard error of the QQ-estimator of the inverse $1/\zeta$ of the tail index based on n largest observations is asymptotically $\sqrt{2}/(\zeta\sqrt{n})$. Csörgő and Viharos (1997) prove asymptotic normality of the OLS estimators of the tail index in the case $\gamma = 0$ (see also Viharos, 1999; Csörgő and Viharos, 2006). Beirlant, Dierckx, Goegebeur and Matthys (1999) and Aban and Meerschaert (2004) indicate the possibility of modification of the QQ-estimator in which logarithms of ordered observations $\log(Z_{(t)})$ are regressed on $\log(t-1/2)$. Aban and Meerschaert (2004) mention in a remark without providing a proof that regressing logarithms of observations from a heavy-tailed population on logarithms of their ranks shifted by 1/2 reduces the bias of the QQ-estimator. Their remark seems to be motivated by simulations, not by the systematic understanding that Theorems 1 and 2 provide; in particular, they do not indicate that a shift of 1/2 is the best shift.

To our knowledge, general regressions (1.3) and (1.4) with $\gamma \neq 0$ and asymptotic expansions for them are considered, for the first time, in the present work. The modifications of the OLS log-log rank-size regressions with the optimal shift $\gamma = 1/2$ and the correct standard errors provided in this paper were subsequently used in the works by Hinloopen and van Marrewijk (2006), Bosker, Brakman, Garretsen, de Jong and Schramm (2007) and Gabaix and Landier (2008).

2 Formal statement of the results

Throughout the paper, for variables $a_1, ..., a_n$, $\overline{a}_n$ stands for the sample mean $\overline{a}_n = \frac{1}{n}\sum_{t=1}^{n} a_t$. As usual, for a sequence of r.v.'s X_n and a sequence of positive constants a_n, we write $X_n = O_P(a_n)$ ($X_n = O_{a.s.}(a_n)$) if the sequence X_n/a_n is bounded in probability (resp., bounded a.s.) and write $X_n = o_P(a_n)$ ($X_n = o_{a.s.}(a_n)$) if $X_n/a_n \to_P 0$ (resp. $X_n/a_n \to_{a.s.} 0$).

Let $Z_{(1)} \geq Z_{(2)} \geq ... \geq Z_{(n)}$ be the order statistics for a sample from the population with the distribution satisfying the power law

$$P(Z > s) = \frac{1}{s^{\zeta}},\ s \geq 1, \zeta > 0. \tag{2.5}$$

Denote $y_t = \log(t - \gamma)$ and $x_t = \log(Z_{(t)})$. Let us consider the OLS estimator $\hat{b}_n^{\gamma}$ of the slope parameter b in log-log rank-size regression (1.3) with $\gamma < 1$ and logarithms of ordered observations regressed on logarithms of shifted ranks:

$$\hat{b}_n^{\gamma} = -\frac{\sum_{t=1}^{n}(x_t - \overline{x}_n)(y_t - \overline{y}_n)}{\sum_{t=1}^{n}(x_t - \overline{x}_n)^2} = -\frac{A_n^{\gamma}}{B_n}. \tag{2.6}$$

We will also consider the OLS estimator $\hat{d}_n^{\gamma}$ of slope in dual to (1.3) regression (1.4) with logarithms of ordered sizes regressed on logarithms of shifted ranks:

$$\hat{d}_n^{\gamma} = -\frac{\sum_{t=1}^{n}(x_t - \overline{x}_n)(y_t - \overline{y}_n)}{\sum_{t=1}^{n}(y_t - \overline{y}_n)^2} = -\frac{A_n^{\gamma}}{D_n}. \tag{2.7}$$

The following theorems provide the main results of the paper.

Theorem 1 *For any $\gamma < 1$, the following asymptotic expansions hold for the bias of the estimators $\hat{b}_n^{\gamma}$ and $\hat{d}_n^{\gamma}$:*

$$E\Big(\hat{b}_n^{\gamma}/\zeta - 1\Big) = \frac{(2\gamma - 1)\log^2 n}{4n} + o\Big(\frac{\log^2 n}{n}\Big), \tag{2.8}$$

$$E\Big(\zeta\hat{d}_n^{\gamma} - 1\Big) = \frac{(1 - 2\gamma)\log^2 n}{4n} + o\Big(\frac{\log^2 n}{n}\Big), \tag{2.9}$$

Theorem 2 *For any $\gamma < 1$, the following asymptotic expansions hold for the estimators $\hat{b}_n^{\gamma}$ and $\hat{d}_n^{\gamma}$:*

$$\hat{b}_n^{\gamma}/\zeta = 1 + \sqrt{\frac{2}{n}}\mathcal{N}(0,1) + O_P\Big(\frac{\log^2 n}{n}\Big), \tag{2.10}$$

$$\zeta\hat{d}_n^{\gamma} = 1 + \sqrt{\frac{2}{n}}\mathcal{N}(0,1) + O_P\Big(\frac{\log^2 n}{n}\Big). \tag{2.11}$$

The arguments for Theorems 1 and 2 are presented in the appendix.

Remark 1 *As follows from asymptotic expansions (2.8) and (2.9), the small sample biases of the OLS estimators $\hat{b}_n^{\gamma}$ and $\hat{d}_n^{\gamma}$ in regressions (1.3) and (1.4) involving logarithms of shifted ranks are both minimized under the choice of the shift $\gamma = 1/2$.*

The proof of Theorems 1 and 2 is based on the following results and methods. First, it exploits the Rényi representation theorem to relate the order statistics for observations following power law (1.1) to the partial sums of scaled i.i.d. exponential r.v.'s (see the beginning of the proof of Lemma 6). Then, we use martingale approximations to the bilinear forms that appear in the numerators of the statistics $\hat{b}_n^\gamma/\zeta - 1 = -(A_n^\gamma + \zeta B_n)/(\zeta B_n)$ and $\zeta\hat{d}_n^\gamma - 1 = -(\zeta A_n^\gamma + D_n)/D_n$ (relation (7.55) in the proof of Lemma 6 and relation (7.74) in the proof of Lemma 8). Third, the arguments use strong approximations to partial sums of independent r.v.'s provided by Lemma 1.

3 A related approach based on harmonic numbers

For $t \geq 1$, denote by $H(t)$ the $t-$th harmonic number: $H(t) = \sum_{i=1}^{t} \frac{1}{i}$. Further, let $H(0) = 0$. Consider the analogues of regressions (1.3) and (1.4) that involve logarithms of ordered sizes $y_t = \log(Z_{(t)})$ and the functions $\tilde{x}_t = H(t-1)$ of ranks of observations:

$$H(t-1) = a' - b' \log(Z_{(t)}). \tag{3.12}$$

$$\log(Z_{(t)}) = c' - d' H(t-1); \tag{3.13}$$

Similar to the proof of Theorem 2, one can show that the following asymptotic expansions hold for the tail index estimators $\hat{b}_n'$ and $\hat{d}_n'$ using regressions (3.12) and (3.13):

$$\hat{b}_n'/\zeta = 1 + \sqrt{\frac{2}{n}}\mathcal{N}(0,1) + O_P\Big(\frac{\log n}{n}\Big); \tag{3.14}$$

$$\zeta\hat{d}_n' = 1 + \sqrt{\frac{2}{n}}\mathcal{N}(0,1) + O_P\Big(\frac{\log n}{n}\Big). \tag{3.15}$$

Comparison of expansions (3.14) and (3.15) with (2.8)-(2.11) shows that, *ceteris paribus*, tail index estimation using regressions involving harmonic numbers is to be preferred, in terms of the small sample bias, to that based on the logarithms of shifted ranks $\log(t-\gamma)$ for any γ. On the other hand, regressions (1.3) and (1.4) are simpler to implement and more visual than estimation procedures based on (3.14) and (3.15). In particular, we are not aware of works that employed estimation approaches based on harmonic numbers similar to (3.14) and (3.15), while regressions (1.3) and (1.4) with $\gamma = 0$ are commonly used, as discussed in the introduction. Comparison of the asymptotic expansions for the tail index estimators using regressions (3.12) and (3.13) with the OLS tail parameter estimators in log-log rank-size regressions (1.3) and (1.4) also sheds light on the main driving force behind the small sample bias improvements using logarithms of shifted ranks $\log(Rank - 1/2)$. This driving force is, essentially, the fact that $\log(n - 1/2)$ provides better approximation to the harmonic numbers $H(n-1)$ than does $\log(n)$ and, more generally, than $\log(n-\gamma)$, $\gamma < 1$. This is because, as follows from the inequalities for $H(n) - \ln(n + 1/2)$ in Havil (2003), Section 9.3 on pp. 75-79, for all $\gamma < 1$,

$$H(n-1) = C + \ln(n-\gamma) + (\gamma - 1/2)n^{-1} + O(n^{-2}) \tag{3.16}$$

as $n \to \infty$, where $C = \lim_{n\to\infty}(H(n) - \ln n)$ is Euler's constant, so the optimal choice of the shift γ in the sense of the best asymptotical approximation is 1/2 (note that the last inequality on p. 76 in Havil, 2003, should read, in the notations of this section, $1/(24(n+1)^2) < H(n) - \ln(n + 1/2) - C < 1/(24n^2)$).

4 Simulation results

In this section, we present simulation results on the performance of the traditional regression (1.3) with $\gamma = 0$ and the modified regression (1.3) with the optimal shift $\gamma = 1/2$ and the correct standard errors given by Theorem 2. We present the numerical results for the OLS Pareto exponent estimation procedures under dependence and under deviations from power laws (1.1). The results are provided for dependent heavy-tailed data that follow AR(1) processes $Z_t = \rho Z_{t-1} + u_t$, $t \geq 1$, $Z_0 = 0$, or MA(1) processes $Z_t = u_t + \theta u_{t-1}$, $t \geq 1$, with i.i.d. u_t's. The departures from power laws are modeled using the innovations u_t that have Student t distributions with the number of degree of freedom $m = 2, 3, 4$ (Tables 2 and 4) or distributions exhibiting 2nd order deviations from Pareto tails in the Hall (1982) form

$$P(u > s) = s^{-\zeta}\left(1 + c(s^{-\alpha\zeta} - 1)\right), c \in [0, 1), \alpha > 0, s \geq 1, \tag{4.17}$$

(Tables 1 and 3). The choice of the number of degrees of freedom for Student t distributions is motivated by the recent empirical works on heavy-tailedness that indicate that, for many economic and financial time series, the tail index ζ lies in the interval $(2, 4)$ (see Loretan and Phillips, 1994; Gabaix et al., 2003, 2006). The benchmark case $c = 0$ in (4.17) corresponds to the exact Pareto distributions (2.5), and the values $\rho = 0$ and $\theta = 0$ model i.i.d. observations Z_t. Similar to deviations of γ from $1/2$ in (2.8) and (2.9), the term $c\left(s^{-\alpha\zeta} - 1\right)$ modeling the departures from the power laws in (4.17) creates a bias in the estimators $\hat{b}_n^\gamma$ and $\hat{d}_n^\gamma$ in regressions (1.3) and (1.4).

Tables 1 and 2 present the simulation results for the traditional OLS estimator $\hat{b}_n$ of the tail index using regression (1.3) with $\gamma = 0$. These tables also provide the comparisons of the OLS standard errors of the estimator with its true standard deviation. Tables 3 and 4 present the numerical results on the performance of the OLS estimator $\hat{b}_n^\gamma$ using modified regression (1.3) with $\gamma = 1/2$. In Tables 3 and 4, we also present the standard errors of $\hat{b}_n^\gamma$ with $\gamma = 1/2$ provided by expansion (2.10) and compare them to the true standard deviation of the estimator. The asterics in Tables 1-4 indicate rejection of the true null hypothesis on the tail index $H_0 : \zeta = \zeta_0$ in favor of the alternative hypothesis $H_a : \zeta \neq \zeta_0$ at the 5% significance level using the reported standard errors ($\zeta_0 = 1$ for innovations that follow distributions (4.17) with $\zeta = \alpha = 1$ considered in Tables 1 and 3 and $\zeta = m$ for innovations that have Student t distributions with $m = 2, 3, 4$ degrees of freedom in Tables 2 and 4).

For instance, consider the class of exact Pareto i.i.d. observations, which is the first row in Table 1 and Table 3, with $n = 50$ extreme observations included in estimation. Table 1 (column $n = 50$, the first row) shows that the traditional OLS estimator using regression (1.3) with $\gamma = 0$ yields an average of 0.924 (whereas the true tail index is 1), and the OLS standard error is 0.024, very far from the true standard deviation, 0.185. By contrast, the OLS estimator using regression (1.3) with $\gamma = 1/2$ proposed in this paper (Table 3, column $n = 50$, the first row) and expansion (2.10) yield an average estimate of 1.011, and the standard error of 0.202, very close to the true standard deviation, 0.199.

More generally, the OLS estimates $\hat{b}_n$ of Pareto exponents ζ using traditional regression (1.3) with $\gamma = 0$ reported in Tables 1 and 2 are significantly different from the true tail indices, which means that $\hat{b}_n$ is biased in small samples. According to the same tables, the OLS standard errors in regression (1.3) with $\gamma = 0$

Table 1. Behavior of the usual OLS estimator $\hat{b}_n$ in the regression $\log(\text{Rank}) = a - b\log(\text{Size})$ for innovations deviating from power laws

	n	50	100	200	500
AR(1)			Mean $\hat{b}_n$		
c	ρ		(OLS s.e) (SD $\hat{b}_n$)		
0	0	0.924* (0.024) (0.185)	0.944* (0.014) (0.134)	0.961* (0.008) (0.098)	0.978* (0.004) (0.063)
0	0.5	1.082* (0.021) (0.296)	1.069* (0.012) (0.244)	1.073* (0.007) (0.195)	1.102* (0.004) (0.145)
0	0.8	1.373* (0.034) (0.520)	1.271* (0.019) (0.417)	1.235* (0.011) (0.343)	1.235* (0.006) (0.268)
0.5	0	0.925* (0.024) (0.181)	0.942* (0.014) (0.132)	0.960* (0.008) (0.098)	0.978* (0.004) (0.063)
0.5	0.5	1.082* (0.020) (0.301)	1.067* (0.012) (0.244)	1.074* (0.007) (0.194)	1.104* (0.004) (0.146)
0.5	0.8	1.379* (0.034) (0.512)	1.276* (0.019) (0.412)	1.226* (0.011) (0.343)	1.238* (0.006) (0.266)
0.8	0	0.925* (0.024) (0.186)	0.945* (0.014) (0.134)	0.960* (0.008) (0.097)	0.978* (0.004) (0.063)
0.8	0.5	1.084* (0.020) (0.297)	1.067* (0.012) (0.239)	1.069* (0.007) (0.195)	1.101* (0.004) (0.145)
0.8	0.8	1.378* (0.034) (0.520)	1.270* (0.019) (0.413)	1.227* (0.011) (0.342)	1.238* (0.006) (0.265)
MA(1)			Mean $\hat{b}_n$		
c	θ		(OLS s.e) (SD $\hat{b}_n$)		
0	0.5	0.988 (0.024) (0.261)	0.993 (0.014) (0.193)	1.003 (0.009) (0.142)	1.032* (0.004) (0.094)
0	0.8	0.989 (0.030) (0.275)	0.994 (0.017) (0.198)	1.011 (0.010) (0.146)	1.034* (0.005) (0.098)
0.5	0	0.926* (0.024) (0.182)	0.942* (0.014) (0.133)	0.961* (0.008) (0.099)	0.977* (0.004) (0.063)
0.5	0.5	0.988 (0.024) (0.259)	0.992 (0.014) (0.193)	1.007 (0.009) (0.142)	1.032* (0.004) (0.095)
0.5	0.8	0.988 (0.030) (0.274)	0.992 (0.017) (0.196)	1.005 (0.010) (0.145)	1.034* (0.005) (0.098)
0.8	0	0.925* (0.024) (0.184)	0.944* (0.014) (0.134)	0.960* (0.008) (0.095)	0.978* (0.004) (0.062)
0.8	0.5	0.991 (0.024) (0.258)	0.993 (0.014) (0.192)	1.005 (0.009) (0.140)	1.030* (0.004) (0.095)
0.8	0.8	0.990 (0.030) (0.276)	0.991 (0.017) (0.198)	1.006 (0.010) (0.145)	1.033* (0.005) (0.098)

Notes: The entries are the estimates of the tail index and their standard errors using regression (1.3) with $\gamma = 0$ for the AR(1) and MA(1) processes $Z_t = \rho Z_{t-1} + u_t$, $t \geq 1$, $Z_0 = 0$, and $Z_t = u_t + \theta u_{t-1}$, where i.i.d. u_t follow the distribution $P(u > s) = s^{-\zeta}\left(1 + c(s^{-\alpha\zeta} - 1)\right)$, $s \geq 1$, with $\zeta = \alpha = 1$ and $c \in [0, 1)$. For a general case $\zeta > 0$, one multiplies all the numbers in the table by ζ. "Mean $\hat{b}_n$" is the sample mean of the estimates $\hat{b}_n$ obtained in simulations, and "SD $\hat{b}_n$" is their sample standard deviation. "OLS s.e." is the OLS standard error in regression (1.3) with $\gamma = 0$. The asteric indicates rejection of the true null hypothesis $H_0 : \zeta = 1$ in favor of the alternative hypothesis $H_a : \zeta \neq 1$ at the 5% significance level using the reported OLS standard errors. The total number of observations $N = 2000$. Based on 10000 replications.

Table 2. Behavior of the usual OLS estimator $\hat{b}_n$ in the regression $\log(\text{Rank}) = a - b\log(\text{Size})$ for Student t innovations

	n	50	100	200	500
AR(1)		Mean $\hat{b}_n$			
m	ρ	(OLS s.e) (SD $\hat{b}_n$)			
2	0	1.810* (0.045) (0.349)	1.809* (0.026) (0.245)	1.768* (0.014) (0.160)	1.524* (0.010) (0.073)
2	0.5	1.993 (0.042) (0.454)	1.986 (0.024) (0.351)	1.932* (0.014) (0.247)	1.647* (0.011) (0.115)
2	0.8	2.433* (0.053) (0.787)	2.334* (0.031) (0.608)	2.199* (0.019) (0.429)	1.796* (0.015) (0.197)
3	0	2.560* (0.063) (0.473)	2.503* (0.036) (0.312)	2.342* (0.021) (0.192)	1.838* (0.016) (0.079)
3	0.5	2.852* (0.065) (0.589)	2.777* (0.037) (0.414)	2.597* (0.022) (0.262)	1.992* (0.019) (0.107)
3	0.8	3.632* (0.084) (1.021)	3.400* (0.049) (0.722)	3.044 (0.032) (0.448)	2.179* (0.024) (0.186)
4	0	3.151* (0.078) (0.546)	3.002* (0.043) (0.350)	2.729* (0.027) (0.205)	2.017* (0.021) (0.083)
4	0.5	3.523* (0.083) (0.661)	3.358* (0.047) (0.443)	3.024* (0.030) (0.259)	2.162* (0.024) (0.110)
4	0.8	4.546* (0.112) (1.101)	4.096 (0.065) (0.700)	3.516* (0.043) (0.417)	2.334* (0.030) (0.185)
MA(1)		Mean $\hat{b}_n$			
m	θ	(OLS s.e) (SD $\hat{b}_n$)			
2	0.5	1.927 (0.044) (0.446)	1.927* (0.025) (0.325)	1.869* (0.015) (0.220)	1.602* (0.011) (0.097)
2	0.8	1.978 (0.054) (0.524)	1.951 (0.031) (0.368)	1.894* (0.018) (0.242)	1.617* (0.012) (0.104)
3	0.5	2.774* (0.064) (0.569)	2.697* (0.036) (0.400)	2.519* (0.022) (0.245)	1.944* (0.018) (0.099)
3	0.8	2.916 (0.075) (0.707)	2.792* (0.042) (0.464)	2.587* (0.025) (0.283)	1.974* (0.019) (0.106)
4	0.5	3.430* (0.082) (0.649)	3.253* (0.045) (0.428)	2.944* (0.029) (0.244)	2.122* (0.023) (0.099)
4	0.8	3.649* (0.092) (0.790)	3.419* (0.052) (0.510)	3.035* (0.033) (0.287)	2.159* (0.025) (0.106)

Notes: The entries are estimates of the tail index and their standard errors using regression (1.3) with $\gamma = 0$ for the AR(1) and MA(1) processes $Z_t = \rho Z_{t-1} + u_t$, $t \geq 1$, $Z_0 = 0$, and $Z_t = u_t + \theta u_{t-1}$, where i.i.d. u_t have the Student t distribution with m degrees of freedom. “Mean $\hat{b}_n$” is the sample mean of the estimates $\hat{b}_n$ obtained in simulations, and “SD $\hat{b}_n$” is their sample standard deviation. “OLS s.e.” is the OLS standard error in regression (1.3) with $\gamma = 0$. The asteric indicates rejection of the true null hypothesis on the tail index ζ of Z_t $H_0 : \zeta = m$ in favor of the alternative hypothesis $H_a : \zeta \neq m$ at the 5% significance level using the reported OLS standard errors. The total number of observations $N = 2000$. Based on 10000 replications.

Table 3. Behavior of the OLS estimator $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$ in the regression $\log(\text{Rank} - 1/2) = a - b \log(\text{Size})$ for innovations deviating from power laws

	n	50	100	200	500
AR(1)		Mean $\hat{b}_n^{\gamma=1/2}$			
c	ρ	($\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$) (SD $\hat{b}_n^{\gamma=1/2}$)			
0	0	1.011 (0.202) (0.199)	1.001 (0.142) (0.139)	0.998 (0.100) (0.100)	0.998 (0.063) (0.063)
0	0.5	1.179 (0.236) (0.320)	1.131 (0.160) (0.257)	1.112 (0.111) (0.201)	1.124 (0.071) (0.147)
0	0.8	1.487 (0.297) (0.564)	1.340 (0.189) (0.439)	1.277* (0.128) (0.354)	1.258* (0.080) (0.272)
0.5	0	1.013 (0.203) (0.194)	0.999 (0.141) (0.137)	0.997 (0.100) (0.101)	0.998 (0.063) (0.064)
0.5	0.5	1.179 (0.236) (0.326)	1.129 (0.160) (0.257)	1.113 (0.111) (0.200)	1.127 (0.071) (0.147)
0.5	0.8	1.494 (0.299) (0.555)	1.344 (0.190) (0.434)	1.268* (0.127) (0.354)	1.262* (0.080) (0.270)
0.8	0	1.013 (0.203) (0.200)	1.003 (0.142) (0.139)	0.997 (0.100) (0.099)	0.998 (0.063) (0.063)
0.8	0.5	1.181 (0.236) (0.322)	1.129 (0.160) (0.251)	1.109 (0.111) (0.201)	1.123 (0.071) (0.147)
0.8	0.8	1.493 (0.299) (0.565)	1.338 (0.189) (0.435)	1.269* (0.127) (0.353)	1.262* (0.080) (0.269)
MA(1)		Mean $\hat{b}_n^{\gamma=1/2}$			
c	θ	($\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$) (SD $\hat{b}_n^{\gamma=1/2}$)			
0	0.5	1.078 (0.216) (0.281)	1.052 (0.149) (0.202)	1.041 (0.104) (0.146)	1.053 (0.067) (0.095)
0	0.8	1.078 (0.216) (0.296)	1.052 (0.149) (0.207)	1.049 (0.105) (0.149)	1.054 (0.067) (0.099)
0.5	0	1.014 (0.203) (0.195)	1.000 (0.141) (0.138)	0.999 (0.100) (0.101)	0.998 (0.063) (0.064)
0.5	0.5	1.078 (0.216) (0.279)	1.051 (0.149) (0.202)	1.046 (0.105) (0.146)	1.053 (0.067) (0.096)
0.5	0.8	1.076 (0.215) (0.295)	1.050 (0.148) (0.205)	1.043 (0.104) (0.149)	1.055 (0.067) (0.099)
0.8	0	1.013 (0.203) (0.198)	1.002 (0.142) (0.140)	0.998 (0.100) (0.098)	0.998 (0.063) (0.063)
0.8	0.5	1.081 (0.216) (0.277)	1.052 (0.149) (0.201)	1.043 (0.104) (0.144)	1.051 (0.066) (0.096)
0.8	0.8	1.079 (0.216) (0.297)	1.049 (0.148) (0.207)	1.044 (0.104) (0.149)	1.054 (0.067) (0.099)

Notes: The entries are estimates of the tail index and their standard errors using regression (1.3) with $\gamma = 1/2$ for the AR(1) and MA(1) processes $Z_t = \rho Z_{t-1} + u_t$, $t \geq 1$, $Z_0 = 0$, and $Z_t = u_t + \theta u_{t-1}$, where i.i.d. u_t follow the distribution $P(Z > s) = s^{-\zeta}\left(1 + c(s^{-\alpha\zeta} - 1)\right)$, $s \geq 1$, with $\zeta = \alpha = 1$ and $c \in [0, 1)$. For a general case $\zeta > 0$, one multiplies all the numbers in the table by ζ. "Mean $\hat{b}_n^{\gamma=1/2}$" is the sample mean of the estimates $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$ obtained in simulations, and "SD $\hat{b}_n^{\gamma=1/2}$" is their sample standard deviation. The values $\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$ are the standard errors of $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$ provided by Theorem 2. The asteric indicates rejection of the true null hypothesis $H_0 : \zeta = 1$ in favor of the alternative hypothesis $H_a : \zeta \neq 1$ at the 5% significance level using the reported standard errors. The total number of observations $N = 2000$. Based on 10000 replications.

Table 4. Behavior of the OLS estimator $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$
in the regression $\log(\text{Rank} - 1/2) = a - b \log(\text{Size})$ for Student t innovations

	n	50	100	200	500
AR(1)		Mean $\hat{b}_n^{\gamma=1/2}$			
m	ρ	($\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$) (True s.e.)			
2	0	1.981	1.918	1.834	1.552*
		(0.396) (0.374)	(0.271) (0.255)	(0.183) (0.164)	(0.098) (0.074)
2	0.5	2.178	2.104	2.004	1.678*
		(0.436) (0.489)	(0.297) (0.367)	(0.200) (0.253)	(0.106) (0.116)
2	0.8	2.647	2.465	2.277	1.827
		(0.529) (0.854)	(0.349) (0.639)	(0.228) (0.442)	(0.116) (0.200)
3	0	2.798	2.651	2.427*	1.870*
		(0.560) (0.507)	(0.375) (0.325)	(0.243) (0.196)	(0.118) (0.080)
3	0.5	3.118	2.941	2.691	2.026*
		(0.624) (0.633)	(0.416) (0.431)	(0.269) (0.268)	(0.128) (0.108)
3	0.8	3.956	3.592	3.149	2.215*
		(0.791) (1.104)	(0.508) (0.756)	(0.315) (0.459)	(0.140) (0.189)
4	0	3.442	3.177	2.825*	2.051*
		(0.688) (0.585)	(0.449) (0.364)	(0.282) (0.210)	(0.130) (0.084)
4	0.5	3.848	3.553	3.130*	2.198*
		(0.770) (0.710)	(0.502) (0.461)	(0.313) (0.265)	(0.139) (0.112)
4	0.8	4.950	4.323	3.634	2.370*
		(0.990) (1.188)	(0.611) (0.732)	(0.363) (0.427)	(0.150) (0.188)
MA(1)		Mean $\hat{b}_n^{\gamma=1/2}$			
m	θ	($\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$) (True s.e.)			
2	0.5	2.106	2.042	1.939	1.632*
		(0.421) (0.480)	(0.289) (0.339)	(0.194) (0.225)	(0.103) (0.098)
2	0.8	2.157	2.065	1.963	1.647*
		(0.431) (0.564)	(0.292) (0.384)	(0.196) (0.248)	(0.104) (0.105)
3	0.5	3.032	2.856	2.610	1.978*
		(0.606) (0.612)	(0.404) (0.417)	(0.261) (0.251)	(0.125) (0.100)
3	0.8	3.180	2.953	2.679	2.008*
		(0.636) (0.761)	(0.418) (0.483)	(0.268) (0.289)	(0.127) (0.107)
4	0.5	3.747	3.441	3.048*	2.157*
		(0.749) (0.697)	(0.487) (0.446)	(0.305) (0.249)	(0.136) (0.100)
4	0.8	3.977	3.613	3.140*	2.194*
		(0.795) (0.849)	(0.511) (0.531)	(0.314) (0.293)	(0.139) (0.108)

Notes: The entries are estimates of the tail index and their standard errors using regression (1.3) with $\gamma = 1/2$ for the AR(1) and MA(1) processes $Z_t = \rho Z_{t-1} + u_t$, $t \geq 1$, $Z_0 = 0$, and $Z_t = u_t + \theta u_{t-1}$, where i.i.d. u_t have the Student t distribution with m degrees of freedom. For a general case $\zeta > 0$, one multiplies all the numbers in the table by ζ. "Mean $\hat{b}_n^{\gamma=1/2}$" is the sample mean of the estimates $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$ obtained in simulations, and "SD $\hat{b}_n^{\gamma=1/2}$" is their sample standard deviation. The values $\sqrt{2/n}\times$Mean $\hat{b}_n^{\gamma=1/2}$ are the standard errors of $\hat{b}_n^{\gamma}$ with $\gamma = 1/2$ provided by Theorem 2. The asteric indicates rejection of the true null hypothesis on the tail index ζ of Z_t $H_0 : \zeta = m$ in favor of the alternative hypothesis $H_a : \zeta \neq m$ at the 5% significance level using the reported standard errors. The total number of observations $N = 2000$. Based on 10000 replications.

are consistently smaller than the true standard deviations. In most of the numerical results presented in Tables 1 and 2, the true null hypothesis on the tail index $H_0 : \zeta = \zeta_0$ is rejected in favor of the alternative hypothesis $H_a : \zeta \neq \zeta_0$ at the 5% significance level using the OLS standard errors.

In most of the entries in Tables 3 and 4, including dependence and deviations from power tail distributions, the standard errors in the regression with shifts $\gamma = 1/2$ are much closer to the true standard deviations than in the case of the OLS standard errors reported in Tables 1 and 2. Comparing to the traditional regression in Tables 1 and 2, the approach illustrated by Tables 3 and 4 rejects the true null hypothesis on the tail index $H_0 : \zeta = \zeta_0$ significantly less often.

Additional simulation results show that regression (1.3) with $\gamma = 1/2$ also performs well and dominates the choice $\gamma = 0$ for GARCH processes. At the same time, it performs very similar to (3.12) and thus may be preferable due to simplicity.

5 An empirical application: Zipf's law for cities

As an example, we study the distribution of city populations (see also Gabaix and Landier, 2008, where the estimation procedures proposed in this paper are used to to confirm a Zipf's law for market capitalization of large firms). This example is, historically, the first economic example of Zipf's law (Zipf, 1949), which is the name of power law (1.1) with the tail exponent ζ equal to 1. Zipf's law is a regularity that has been exerting an enduring interest, because it appears to describe such diverse phenomena as the frequency of words, the popularity of Internet sites, the magnitude of earthquakes (see Li, 2003) and the size of firms (see Axtell, 2001; Gabaix and Landier, 2008).

As a U.S. example of a study of Zipf's law for the cities (in the upper tail at least, see Eeckhout, 2004), we take, following Krugman (1996) and Gabaix (1999), all 135 American metropolitan areas listed in the Statistical Abstract of the United States in the year 1991, which includes all agglomerations with size above 250,000 inhabitants. The advantage is that "metropolitan area" represents the agglomeration of the cities (e.g., the metropolitan area of Boston includes Cambridge), which is commonly viewed as the correct economic definition. We rank cities from largest (rank 1) to smallest (rank $n = 135$), and denote their sizes

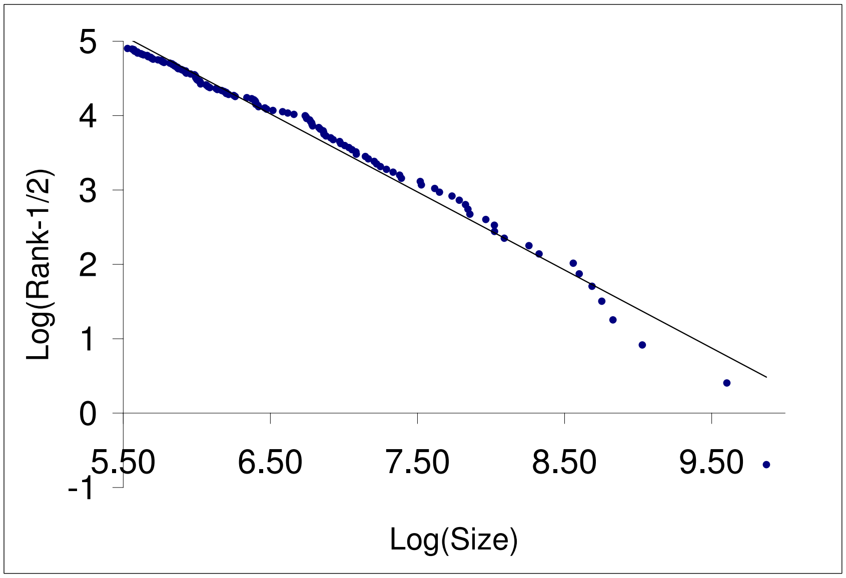

Figure 1: *Log(Population) vs. Log(Rank$-1/2$) for the 135 metropolitan areas in the Statistical Abstract in the US, 1991. The slope of the graph corresponds to the estimate of the slope in regression (1.3) with the optimal shift $\gamma = 1/2$, and is 1.050 (s.e. 0.128). It is consistent with a Zipf's law, i.e. a power law distribution with the tail index equal to 1.*

$S_{(1)} \geq ... \geq S_{(n)}$.

Regression (1.3) with $\gamma = 1/2$ estimated for the data is

$$\log\ (t - 0.5) = \underset{(0.128)}{10.846 - 1.050}\ \log\ S_{(t)}.$$

The number in the bracket is the standard error for the tail index (the slope coefficient $\hat{b}_n^\gamma$) given by $\sqrt{\frac{2}{n}}\hat{b}_n$ by Theorem 2. Figure 1 shows the corresponding plot.

Regression (1.4) with $\gamma = 1/2$ estimated for the data is

$$\log\ S_{(t)} = 10.244 - 0.930\ \log\ (t - 0.5),$$

producing the estimate of the tail index equal to $1/\hat{d}_n^\gamma \approx 1.075$ with the standard error given by $\sqrt{\frac{2}{n}}\frac{1}{\hat{d}_n} \approx 0.131$

by Theorem 2. The estimates of the tail index are not statistically different from 1 at the 10% significance level, so that Zipf's law for cities is confirmed in this dataset.

6 Conclusion and suggestions for future research

The OLS log-log rank-size regression $\log(\text{Rank}) = a - b\log(\text{Size})$ and related procedures are some of the most popular approaches to Pareto exponent estimation, with b taken as an estimate of the tail index. Unfortunately, these procedures are strongly biased in small samples. We provide a simple approach to bias reduction based on the modified log-log rank-size regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$. The shift of 1/2 is optimal and reduces the bias to a leading order. We further show that the standard error on the Pareto exponent ζ in this regression is asymptotically $(2/n)^{1/2}\zeta$, and obtain similar results for the regression $\log(\text{Size}) = c - d\log(\text{Rank} - 1/2)$. The proposed estimation procedures are illustrated using an empirical analysis of the U.S. city size distribution. Simulation results indicate that the proposed tail index estimation procedures perform well under dependence and deviations from power law distributions. They further demonstrate the advantage of the new methods over the standard OLS log-log rank-size regressions.

An important open problem concerns asymptotic expansions for the OLS tail index estimators and their biases for dependent processes, including the autocorrelated time series considered in simulations. Combining the modified OLS estimation approach with block-bootstrap and GARCH filters may be useful in developing tail index estimation procedures under dependence. In addition, unreported preliminary results suggest that the OLS approaches to tail index estimation are more robust than Hill's estimator of a tail index under deviations from power laws. Other important problems include the analysis of the optimal choice of the number n of extreme observations used in estimation and the study of the asymptotic bias of the OLS estimators when n is determined by minimizing the asymptotic mean square error. Analysis of these issues and comparisons of the OLS tail index estimators with other procedures are left for further research.

7 Appendix. Proof of Theorems 1 and 2

Let Z_t follow distribution (2.5), and let $Z'_t = Z_t^{\zeta}$. As in (1.2), denote by $Z'_{(1)} \geq ... \geq Z'_{(n)}$ decreasingly ordered variables Z'_t. We have $P(Z'_t > s) = P(Z_t > s^{1/\zeta}) = 1/s$, $s \geq 1$. Consequently, Z'_t follow distribution (2.5) with $\zeta = 1$. Evidently, for the logarithms of ordered observations $x_t = \log(Z_{(t)})$ and $x'_t = \log(Z'_{(t)})$ one has $x_t = x'_t/\zeta$. Therefore, we get that the OLS estimators $\hat{b}_n^{\gamma}$ and $\hat{d}_n^{\gamma}$ in (2.6) and (2.7) satisfy

$$\hat{b}_n^{\gamma}/\zeta = -\frac{\sum_{t=1}^{n}(x'_t - \overline{x}'_n)(y_t - \overline{y}_n)}{\sum_{t=1}^{n}(x'_t - \overline{x}'_n)^2}, \quad \zeta\hat{d}_n^{\gamma} = -\frac{\sum_{t=1}^{n}(x'_t - \overline{x}'_n)(y_t - \overline{y}_n)}{\sum_{t=1}^{n}(y_t - \overline{y}_n)^2}.$$

This implies that it suffices to prove Theorems 1 and 2 for the case $\zeta = 1$. This will be assumed throughout the rest of the appendix.

For the proof, we will need the following well-known results provided by Lemmas 1-4. Lemma 1 gives the strong approximation to partial sums of independent r.v.'s that holds under the assumption of the existence of a moment generating function in a neighborhood of zero. It is provided by, e.g., the results in Komlós, Major and Tusnády (1975) (see also Komlós, Major and Tusnády, 1976) and by Theorem 2.6.1 on p. 107 in Csörgő and Révész (1981).

In Lemma 1, the notation $\{\tilde{S}_n; \quad n = 1, 2, ...\} =^d \{S_n; \quad n = 1, 2, ...\}$ means that $\{S_n\}$ and $\{\tilde{S}_n\}$ are distributionally equivalent in the sense that all finite-dimensional distributions of $\{S_n\}$ and $\{\tilde{S}_n\}$ are the same, that is, the distribution of the random vector $(S_{t_1}, ..., S_{t_k})$ is the same as that of $(\tilde{S}_{t_1}, ..., \tilde{S}_{t_k})$ for all $1 \leq t_1 < t_2 < ... < t_k$, $k \geq 1$.

Lemma 1 *Let X_t, $t \geq 1$, be a sequence of i.i.d. r.v.'s with $EX_t = 0$, $EX_t^2 = 1$ such that $R(z) = E\exp(zX_t)$ exists in a neighborhood of $z = 0$. Further, let $S_n = \sum_{t=1}^{n} X_t$, $S_0 = 0$, stand for the partial sums of X'_ts. A probability space $(\Omega, \Im, P)$ with a sequence $\{\tilde{S}_n\}$ and a standard Brownian motion $W = (W(s), s \geq 0)$ on it can be so constructed that $\{\tilde{S}_n; \; n = 1, 2, ...\} =^d \{S_n; \; n = 1, 2, ...\}$ and $|\tilde{S}_n - W(n)| = O_{a.s.}(\log\ n)$.*

Similar to Lemma 1, throughout the rest of the appendix, $W = (W(s), s \geq 0)$ denotes a standard Brownian motion. Lemma 2 concerns the modulus of continuity for Brownian sample paths due to P. Lévy.

The asymptotic relation in the lemma is provided, for instance, by Theorem 9.25 on p. 114 in Karatzas and Shreve, 1991, and by the results in Borodin and Salminen, 2002, p. 53.

Lemma 2 *The following relation holds:*

$$\limsup_{\delta\to+0} \frac{1}{\sqrt{2\delta\log(1/\delta)}} \sup_{\substack{0\le t_1,t_2\le 1\\ 0<|t_2-t_1|<\delta}} |W(t_2)-W(t_1)| = 1 \ \text{(a.s.)}. \tag{7.18}$$

Lemma 3 provides an estimate of the rate of growth of sums of independent r.v.'s in terms of their variances. The lemma is a consequence of Theorem 6.17 and the discussion following it on p. 222 in Petrov (1995).

In what follows, for a r.v. X with $EX^2<\infty$, $Var(X)$ denotes its variance.

Lemma 3 *If u_t, $t\ge 1$, are independent r.v.'s such that $Eu_t^2<\infty$, $t\ge 1$, and $V_n = Var(\sum_{t=1}^n u_t) = \sum_{t=1}^n Var(u_t)\to\infty$ as $n\to\infty$, then $\sum_{t=1}^n (u_t - Eu_t) = o_{a.s.}\big(V_n^{1/2}\log\ V_n\big)$.*

Lemma 4 below is provided by Theorem 6.7 in Petrov (1995).

Lemma 4 *Let a_t, $t\ge 1$, be positive numbers such that $a_1\le a_2\le a_3\le \dots$ and $a_t\to\infty$ as $t\to\infty$. If u_t, $t\ge 1$, are independent r.v.'s such that $\sum_{t=1}^\infty Var(u_t)/a_t^2<\infty$, then $\sum_{t=1}^n (u_t-Eu_t)/a_n\to 0$ a.s. as $n\to\infty$.*

The arguments for the following Lemmas 5-9 are provided at the end of this appendix. We first formulate, in Lemma 5, several asymptotic relations involving sums of logarithms. Denote

$$M_n = \sum_{t=1}^{n-1}\Big[\frac{1}{t}\sum_{i=1}^t \log\,(i-\gamma) - \frac{1}{n}\sum_{i=1}^n \log\,(i-\gamma) - \log(t-\gamma) + \log(n-\gamma)\Big]^2, \tag{7.19}$$

$$G_n = \frac{1}{\sqrt{n}}\Big[n + \sum_{t=1}^n \frac{1}{t}\Big(\sum_{i=1}^t \log\,(i-\gamma)\Big) - \Big(\sum_{t=1}^n \log\,(t-\gamma)\Big)\Big], \tag{7.20}$$

$$H_n = \frac{1}{\sqrt{n}}\Big[\sum_{t=1}^n \log^2\,(t-\gamma) - \frac{1}{n}\Big(\sum_{t=1}^n \log\,(t-\gamma)\Big)^2 + \sum_{t=1}^n\frac{1}{t}\sum_{i=1}^t \log\,(i-\gamma) - \sum_{t=1}^n \log\,(t-\gamma)\Big].$$

Lemma 5 *For all $\gamma < 1$, the following relations hold:*

$$\sum_{t=1}^{n} \log\,(t-\gamma) = n \log\,(n-\gamma) - n + \Big(\frac{1}{2} - \gamma\Big) \log\,(n-\gamma) + O(1), \tag{7.21}$$

$$\sum_{t=1}^{n} \log^2\,(t-\gamma) = (n-\gamma)\log^2\,(n-\gamma) - 2(n-\gamma)\log\,(n-\gamma) + \\ 2n + \frac{\log^2\,(n-\gamma)}{2} + O(1). \tag{7.22}$$

$$\sum_{t=1}^{n} \frac{\log\,(t-\gamma)}{t} = \frac{\log^2 n}{2} + o(\log^2 n), \tag{7.23}$$

$$M_n = O(1), \tag{7.24}$$

$$G_n = \frac{(1-2\gamma)\log^2 n}{4\sqrt{n}} + o\Big(\frac{\log^2 n}{\sqrt{n}}\Big). \tag{7.25}$$

$$H_n = \frac{(2\gamma-1)\log^2 n}{4\sqrt{n}} + o\Big(\frac{\log^2 n}{\sqrt{n}}\Big). \tag{7.26}$$

Relations (2.8) and (2.10) for $\zeta = 1$ are consequences of (2.6) and the asymptotic expansions for the statistics A_n^{γ} and B_n under $\zeta = 1$ provided by Lemmas 6 and 7.

Lemma 6 *The following asymptotic expansions hold for $\zeta = 1$:*

$$E(A_n^{\gamma} + B_n) = \frac{(1-2\gamma)\log^2 n}{4} + o\Big(\log^2 n\Big), \tag{7.27}$$

$$\frac{1}{\sqrt{n}}(A_n^{\gamma} + B_n) = \mathcal{N}(0,2) + O_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big). \tag{7.28}$$

Lemma 7 *The following asymptotic relation holds for $\zeta = 1$:*

$$\frac{B_n}{n} = 1 + O_{a.s.}\Big(\frac{\log\, n}{\sqrt{n}}\Big). \tag{7.29}$$

Similar to (2.8) and (2.10), asymptotic expansions (2.9) and (2.11) for $\zeta = 1$ follow from (2.7) and the asymptotic expansions for the statistics A_n^{γ} and D_n under $\zeta = 1$ provided by Lemmas 8 and 9.

Lemma 8 *The following asymptotic expansions hold for $\zeta = 1$:*

$$E(A_n^\gamma + D_n) = \frac{(2\gamma - 1)\log^2 n}{4} + o\Big(\log^2 n\Big), \tag{7.30}$$

$$\frac{1}{\sqrt{n}}(A_n^\gamma + D_n) = \mathcal{N}(0, 2) + O_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big). \tag{7.31}$$

Lemma 9 *The following asymptotic relation holds for $\zeta = 1$:*

$$\frac{D_n}{n} = 1 + O\Big(\frac{\log^2\ n}{n}\Big). \tag{7.32}$$

Proof of Lemma 5. Relations (7.21) and (7.22) follow from Euler-Maclaurin summation formula with the remainder terms that are $O(1)$ for the sums in them (see, e.g., Havil, 2003, p. 86). Using again Euler-Maclaurin summation formula in a similar way (or first-order integral approximations to partial sums), we obtain (7.23). Denote $L_t = \frac{1}{t}\sum_{i=1}^{t} \log\ (i - \gamma) - \log(t - \gamma) + 1 - \Big(\frac{1}{2} - \gamma\Big)\frac{\log\ (t - \gamma)}{t}$. From (7.21) it follows that

$$M_n = \sum_{t=1}^{n-1}\Big[L_t - L_n + \Big(\frac{1}{2} - \gamma\Big)\frac{\log\ (t-\gamma)}{t} - \Big(\frac{1}{2} - \gamma\Big)\frac{\log\ (n-\gamma)}{n}\Big]^2 \leq$$
$$C\sum_{t=1}^{n-1} L_t^2 + CnL_n^2 + C\sum_{t=1}^{n-1}\Big[\frac{\log\ (t-\gamma)}{t}\Big]^2 + C\Big[\frac{\log^2\ (n-\gamma)}{n}\Big] \leq$$
$$C\sum_{t=1}^{n-1}\frac{1}{t^2} + \frac{C}{n} + C\sum_{t=1}^{n-1}\Big[\frac{\log\ (t-\gamma)}{t}\Big]^2 + C\Big[\frac{\log^2\ (n-\gamma)}{n}\Big] \leq C.$$

Thus, (7.24) indeed holds. From (7.21) we further get

$$G_n = \frac{1}{\sqrt{n}}\Big[n + \sum_{t=1}^{n}\log\ (t-\gamma) - n + \Big(\frac{1}{2} - \gamma\Big)\sum_{t=1}^{n}\frac{\log\ (t-\gamma)}{t} -$$
$$n\log\ (n-\gamma) + n - \Big(\frac{1}{2} - \gamma\Big)\log\ (n-\gamma) + O(\log\ n)\Big] =$$
$$\frac{1}{\sqrt{n}}\Big[n\log\ (n-\gamma) - n + \Big(\frac{1}{2} - \gamma\Big)\log\ (n-\gamma) + \Big(\frac{1}{2} - \gamma\Big)\sum_{t=1}^{n}\frac{\log\ (t-\gamma)}{t} -$$
$$n\log\ (n-\gamma) + n - \Big(\frac{1}{2} - \gamma\Big)\log\ (n-\gamma) + O(\log\ n)\Big] =$$
$$\frac{1}{\sqrt{n}}\Big(\frac{1}{2} - \gamma\Big)\sum_{t=1}^{n}\frac{\log\ (t-\gamma)}{t} + O\Big(\frac{\log\ n}{\sqrt{n}}\Big).$$

This, together with (7.23), implies (7.25). In a similar way, relation (7.26) follows from (7.21), (7.22) and (7.23). ∎

Proof of Lemma 6. By the Rényi representation theorem (see Beirlant et al., 2004, Sections 4.2.1 (iii) and 4.4), one has that, for the logarithms $x_t = \log\ Z_{(t)}$ of ordered observations from a population with the distribution satisfying power law (2.5), the transformations

$$\tau_t = t\big(x_t - x_{t+1}\big), \quad t = 1, ..., n-1,$$

are i.i.d. exponential r.v.'s with parameter 1: $P(\tau_t > s) = \exp(-s)$, $s \geq 0$. That is, one can represent the regressors in (1.3) as weighted sums of exponential r.v.'s in the following way:

$$x_t = x_n + z_t, \ t = 1, ..., n,$$

where $z_n = 0$ and $z_t = \sum_{i=t}^{n-1} \frac{\tau_i}{i}$, $t = 1, ..., n-1$. We, therefore, get

$$B_n = \sum_{t=1}^{n} (x_t - \overline{x}_n)^2 = \sum_{t=1}^{n} (x_n + z_t - x_n - \overline{z}_n)^2 = \sum_{t=1}^{n} (z_t - \overline{z}_n)^2 = \sum_{t=1}^{n-1} z_t^2 - n\overline{z}_n^2, \tag{7.33}$$

and, similarly,

$$A_n^{\gamma} = \sum_{t=1}^{n} (x_t - \overline{x}_n)(y_t - \overline{y}_n) = \sum_{t=1}^{n} (z_t - \overline{z}_n)(y_t - \overline{y}_n) = \sum_{t=1}^{n-1} z_t y_t - n\overline{z}_n \overline{y}_n. \tag{7.34}$$

We further have

$$\sum_{t=1}^{n-1} z_t^2 = \sum_{t=1}^{n-1} \Big(\sum_{i=t}^{n-1} \frac{\tau_i}{i} \Big)^2 = \sum_{t=1}^{n-1} \sum_{i=t}^{n-1} \frac{\tau_i^2}{i^2} + 2 \sum_{t=1}^{n-1} \sum_{i=t}^{n-2} \frac{\tau_i}{i} \sum_{j=i+1}^{n-1} \frac{\tau_j}{j}. \tag{7.35}$$

Using a change of summation indices, we get

$$\sum_{t=1}^{n-1} \sum_{i=t}^{n-1} \frac{\tau_i^2}{i^2} = \sum_{i=1}^{n-1} \frac{\tau_i^2}{i^2} \sum_{t=1}^{i} 1 = \sum_{i=1}^{n-1} \frac{i\tau_i^2}{i^2} = \sum_{i=1}^{n-1} \frac{\tau_i^2}{i}, \tag{7.36}$$

$$\sum_{t=1}^{n-1} \sum_{i=t}^{n-2} \frac{\tau_i}{i} \sum_{j=i+1}^{n-1} \frac{\tau_j}{j} = \sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \frac{\tau_i}{i} \sum_{t=1}^{i} 1 = \sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \frac{i\tau_i}{i} = \sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \tau_i. \tag{7.37}$$

Relations (7.36) and (7.37), together with (7.35), imply

$$\sum_{t=1}^{n-1} z_t^2 = \sum_{i=1}^{n-1} \frac{\tau_i^2}{i} + 2 \sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \tau_i. \tag{7.38}$$

In addition,

$$n\overline{z}_n^2 = \frac{1}{n} \Big(\sum_{t=1}^{n-1} \sum_{i=t}^{n-1} \frac{\tau_i}{i} \Big)^2 = \frac{1}{n} \Big(\sum_{i=1}^{n-1} \tau_i \Big)^2 = \frac{1}{n} \sum_{i=1}^{n-1} \tau_i^2 + \frac{2}{n} \sum_{i=2}^{n-1} \tau_i \sum_{j=1}^{i-1} \tau_j, \tag{7.39}$$

with the second equality obtained by a change of summation indices similar to (7.36). Using (7.33), (7.38) and (7.39), we get

$$B_n = \sum_{i=1}^{n-1} \frac{\tau_i^2}{i} + 2\sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \tau_i - \frac{1}{n}\sum_{i=1}^{n-1} \tau_i^2 - \frac{2}{n}\sum_{i=2}^{n-1} \tau_i \sum_{j=1}^{i-1} \tau_j. \tag{7.40}$$

Similar to the above derivations, we have, using a change of summation indices,

$$\sum_{t=1}^{n-1} z_t y_t = \sum_{t=1}^{n-1} \log\,(t-\gamma)\Big(\sum_{i=t}^{n-1} \frac{\tau_i}{i}\Big) = \sum_{t=1}^{n-1} \frac{\tau_t}{t}\Big(\sum_{i=1}^{t} \log\,(i-\gamma)\Big), \tag{7.41}$$

$$n\overline{z}_n\overline{y}_n = \Big(\sum_{t=1}^{n-1}\sum_{i=t}^{n-1} \frac{\tau_i}{i}\Big)\Big(\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma)\Big) = \Big(\sum_{t=1}^{n-1} \tau_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma)\Big). \tag{7.42}$$

Relations (7.34), (7.41) and (7.42) imply

$$A_n^\gamma = \sum_{t=1}^{n-1} \frac{\tau_t}{t}\Big(\sum_{i=1}^{t} \log\,(i-\gamma)\Big) - \Big(\sum_{t=1}^{n-1} \tau_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma)\Big). \tag{7.43}$$

From (7.40) and (7.43) we get

$$\begin{aligned}\frac{1}{\sqrt{n}}(A_n^\gamma + B_n) = \frac{1}{\sqrt{n}}\Big[&\sum_{t=1}^{n-1} \frac{\tau_t}{t}\Big(\sum_{i=1}^{t} \log\,(i-\gamma)\Big) - \Big(\sum_{t=1}^{n-1} \tau_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma)\Big) + \\ &\sum_{i=1}^{n-1} \frac{\tau_i^2}{i} + 2\sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \tau_i - \frac{1}{n}\sum_{i=1}^{n-1} \tau_i^2 - \frac{2}{n}\sum_{i=2}^{n-1} \tau_i \sum_{j=1}^{i-1} \tau_j\Big].\end{aligned} \tag{7.44}$$

Since τ_t, $t \geq 1$, are i.i.d. r.v.'s with $E\tau_t = 1$, $t \geq 1$, we obtain

$$E\Big(2\sum_{j=2}^{n-1} \frac{\tau_j}{j} \sum_{i=1}^{j-1} \tau_i - \frac{2}{n}\sum_{i=2}^{n-1} \tau_i \sum_{j=1}^{i-1} \tau_j\Big) = n + O(\log\ n), \tag{7.45}$$

$$E\Big[\sum_{i=1}^{n-1} \frac{\tau_i^2}{i}\Big] = O(\log\ n), \tag{7.46}$$

$$E\Big[\frac{1}{n}\sum_{i=1}^{n-1} \tau_i^2\Big] = O(1). \tag{7.47}$$

Since, by (7.21),

$$\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma) = O(\log\ n), \tag{7.48}$$

from (7.44)-(7.47), it follows that

$$E(A_n^\gamma + B_n) = \sum_{t=1}^{n-1} \frac{1}{t}\Big(\sum_{i=1}^{t} \log\,(i-\gamma)\Big) - (n-1)\Big(\frac{1}{n}\sum_{t=1}^{n} \log\,(t-\gamma)\Big) + n + O(\log\ n) =$$

$$\sqrt{n}G_n + o\Big(\log^2 n\Big),$$

where G_n is defined in (7.20). This, together with (7.25), implies (7.27).

Consider

$$\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i\tau_j}{j} - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} \tau_i\tau_j = \frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i(\tau_j-1)}{j} + \\
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i}{j} - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} \tau_i(\tau_j-1) - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} \tau_i = \\
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{(\tau_i-1)(\tau_j-1)}{j} + \frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_j-1}{j} + \frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i-1}{j} + \\
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{1}{j} - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_i-1)(\tau_j-1) - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_j-1) - \\
&\frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_i-1) - \frac{(n-1)(n-2)}{n^{3/2}}. \qquad (7.49)
\end{aligned}$$

Using a change of summation indices, we have that

$$\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{1}{j} - \frac{(n-1)(n-2)}{n^{3/2}} = \sqrt{n} + O\Big(\frac{\log\ n}{\sqrt{n}}\Big)$$

and

$$\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_j-1}{j} - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_i-1) - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_j-1) = \\
&\frac{2}{\sqrt{n}} \sum_{j=1}^{n-1} (\tau_j-1) - \frac{2}{n^{3/2}} \sum_{j=1}^{n-1} (\tau_j-1)(n-j) - \frac{2}{n^{3/2}} \sum_{j=1}^{n-1} (\tau_j-1)j + O_P\Big(\frac{1}{\sqrt{n}}\Big) = \\
&\frac{2}{\sqrt{n}} \sum_{j=1}^{n-1} (\tau_j-1) - \frac{2}{\sqrt{n}} \sum_{j=1}^{n-1} (\tau_j-1) + \frac{2}{n^{3/2}} \sum_{j=1}^{n-1} (\tau_j-1)j - \frac{2}{n^{3/2}} \sum_{j=1}^{n-1} (\tau_j-1)j + O_P\Big(\frac{1}{\sqrt{n}}\Big) = O_P\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}$$

From (7.49) it thus follows that

$$\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i\tau_j}{j} - \frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} \tau_i\tau_j = \frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{(\tau_i-1)(\tau_j-1)}{j} - \\
&\frac{2}{n^{3/2}} \sum_{1\le i<j\le n-1} (\tau_i-1)(\tau_j-1) + \frac{2}{\sqrt{n}} \sum_{1\le i<j\le n-1} \frac{\tau_i-1}{j} + \sqrt{n} + O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big).
\end{aligned}$$

Using this relation, from (7.44) we now obtain

$$\frac{1}{\sqrt{n}}(A_n^\gamma + B_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\ (i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{i=1}^{n}\log\ (i-\gamma)\Big) +$$
$$2\frac{1}{t}\sum_{i=1}^{t-1}(\tau_i - 1) - \frac{2}{n}\sum_{i=1}^{t-1}(\tau_i - 1) + 2\sum_{j=t+1}^{n-1}\frac{1}{j}\Big] + G_n +$$
$$\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n-1}\frac{\tau_i^2}{i} - \frac{1}{n^{3/2}}\sum_{i=1}^{n-1}\tau_i^2\Big] + O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big), \tag{7.50}$$

where G_n is defined in (7.20). Relations (7.46), (7.47) and Chebyshev's inequality imply

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-1}\frac{\tau_i^2}{i} = O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big), \tag{7.51}$$

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n-1}\tau_i^2 = O_P\Big(\frac{1}{\sqrt{n}}\Big). \tag{7.52}$$

In addition, it is not difficult to see that $Var\Big[\sum_{t=1}^{n-1}\frac{\tau_t-1}{t}\sum_{i=1}^{t-1}(\tau_i-1)\Big] = O\Big(\sum_{t=1}^{n}\frac{1}{t}\Big) = O(\log\ n)$. This implies that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}\frac{\tau_t - 1}{t}\sum_{i=1}^{t-1}(\tau_i - 1) = O_P\Big(\sqrt{\frac{\log\ n}{n}}\Big). \tag{7.53}$$

Similarly, since $Var\Big[\sum_{t=1}^{n-1}(\tau_t-1)\sum_{i=1}^{t-1}(\tau_i-1)\Big] = O(n^2)$, we get

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n-1}(\tau_t - 1)\sum_{i=1}^{t-1}(\tau_i - 1) = O_P\Big(\frac{1}{\sqrt{n}}\Big). \tag{7.54}$$

Using relations (7.25) and (7.50)-(7.54), we obtain

$$\frac{1}{\sqrt{n}}(A_n^\gamma + B_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\ (i-\gamma)\Big) -$$
$$\Big(\frac{1}{n}\sum_{i=1}^{n}\log\ (i-\gamma)\Big) + 2\sum_{j=t+1}^{n-1}\frac{1}{j}\Big] + \frac{2}{\sqrt{n}}\sum_{t=1}^{n-1}\frac{\tau_t - 1}{t}\sum_{i=1}^{t-1}(\tau_i - 1) -$$
$$\frac{2}{n^{3/2}}\sum_{t=1}^{n-1}(\tau_t - 1)\sum_{i=1}^{t-1}(\tau_i - 1) + G_n + O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big) =$$
$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\ (i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{i=1}^{n}\log\ (i-\gamma)\Big) + 2\sum_{j=t+1}^{n-1}\frac{1}{j}\Big] +$$

$$G_n + O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) -$$

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\log\,(i-\gamma)\Big) + 2\sum_{j=t+1}^{n-1}\frac{1}{j}\Big] + \frac{(1-2\gamma)\log^2 n}{4\sqrt{n}} + o_P(\frac{\log^2 n}{\sqrt{n}}) =$$

$$-\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\log\,(t/n) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) -$$

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\log\,(i-\gamma)\Big) + 2\sum_{j=t+1}^{n-1}\frac{1}{j} + \log\,(t/n)\Big] + \frac{(1-2\gamma)\log^2 n}{4\sqrt{n}} + o_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big). \tag{7.55}$$

Let us show that

$$U_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) -$$

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\log\,(i-\gamma)\Big) + 2\sum_{j=t+1}^{n-1}\frac{1}{j} + \log\,(t/n)\Big] = O_P\Big(\frac{1}{\sqrt{n}}\Big). \tag{7.56}$$

We have

$$Var(\sqrt{n}U_n) = \sum_{t=1}^{n-1}\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{i=1}^{n}\log\,(i-\gamma)\Big) - \log(t-\gamma) +$$

$$\log(n-\gamma) + 2\sum_{j=t+1}^{n-1}\frac{1}{j} + 2\log\,(t/n) + \log\big(1-\gamma/t\big) - \log\big(1-\gamma/n\big)\Big]^2 \le$$

$$C\Big(M_n + \sum_{t=1}^{n-1}\Big[\log\big(1-\gamma/t\big) - \log\big(1-\gamma/n\big)\Big]^2 + \sum_{t=1}^{n-1}\Big[\sum_{j=t+1}^{n-1}\frac{1}{j} + \log\,(t/n)\Big]^2\Big) =$$

$$C(M_n + Q_n + R_n),$$

where M_n is defined in (7.19), $R_n = \sum_{t=1}^{n-1}\Big[\sum_{j=t+1}^{n-1}\frac{1}{j} + \log\,(t/n)\Big]^2$, and

$$Q_n = \sum_{t=1}^{n-1}\Big[\log\big(1-\gamma/t\big) - \log\big(1-\gamma/n\big)\Big]^2. \tag{7.57}$$

Using the inequality $|\log\,(1-x)| \le 2|x|$, $-1/2 < x < 1/2$, one easily obtains that

$$Q_n = O(1). \tag{7.58}$$

Since, by integral approximations to partial sums (or by (3.16)), $\Big|\sum_{j=t+1}^{n-1}\frac{1}{j} + \log\,(t/n)\Big| \le \frac{C}{t}$ for all t and n, we also get that $R_n = O(1)$. Using (7.24) and the above relations, we conclude that $Var(\sqrt{n}U_n) = O(1)$. Thus, (7.56) indeed holds. We now provide the argument for the relation

$$-\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t - 1)\log\,(t/n) = \sqrt{2}\mathcal{N}(0,1) + O_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big) \tag{7.59}$$

using strong approximations to partial sums of independent r.v.'s by Brownian motion.

Using partial summation similar to the proof of Lemma 2.3 in Phillips (2007), we get (below, $S_t = \sum_{i=1}^t u_i$ and $u_i = \tau_i - 1$)

$$-\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t \log\,(t/n) = -\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t \log\ t + \log\ n\frac{1}{\sqrt{n}}\sum_{t=1}^n u_t =$$
$$\Big[-\log\ n\frac{S_n}{\sqrt{n}} + \sum_{t=2}^n \big(\log\ t - \log\,(t-1)\big)\frac{S_{t-1}}{\sqrt{n}}\Big] + \log\ n\frac{S_n}{\sqrt{n}} =$$
$$\sum_{t=2}^n \big(\log\ t - \log\,(t-1)\big)\frac{S_{t-1}}{\sqrt{n}}. \tag{7.60}$$

By Lemma 1, one can expand the probability space as necessary to set up a partial sum process that is distributionally equivalent to S_t and the standard Brownian motion $W(\cdot)$ on the same space such that

$$\sup_{1\le t\le n}\Big|\frac{S_{t-1}}{\sqrt{n}} - W\big(\frac{t-1}{n}\big)\Big| = O_{a.s.}\Big(\frac{\log\ n}{\sqrt{n}}\Big). \tag{7.61}$$

As conventional, throughout the rest of the proof we suppose that that the probability space on which the random sequences considered are defined has been appropriately enlarged so that relation (7.61) holds. From (7.61) we get

$$\sum_{t=2}^n \Big(\log\ t - \log\,(t-1)\Big)\frac{S_{t-1}}{\sqrt{n}} = \sum_{t=2}^n \Big(\log\ t - \log\,(t-1)\Big)W\Big(\frac{t-1}{n}\Big) +$$
$$O_{a.s.}\Big(\frac{\log\ n}{\sqrt{n}}\Big)\sum_{t=2}^n \Big(\log\ t - \log\,(t-1)\Big) =$$
$$\sum_{t=2}^n \Big(\log\ t - \log\,(t-1)\Big)W\Big(\frac{t-1}{n}\Big) + O_{a.s.}\Big(\frac{\log^2\ n}{\sqrt{n}}\Big). \tag{7.62}$$

Let us consider the difference between

$$\sum_{t=2}^n \Big(\log\ t - \log\,(t-1)\Big)W\Big(\frac{t-1}{n}\Big) = \sum_{t=2}^n \Big[\log\ \Big(n\frac{t}{n}\Big) - \log\ \Big(n\frac{t-1}{n}\Big)\Big]W\Big(\frac{t-1}{n}\Big)$$

and $\int_0^1 W(r)d\log(nr)$. We have

$$\Big|\sum_{t=2}^{n}\Big(\log\ t-\log\ (t-1)\Big)W\Big(\frac{t-1}{n}\Big)-\int_{1/n}^{1}W(r)d\log\ (nr)\Big|=$$

$$\Big|\sum_{t=2}^{n}\Big[\Big(\log\ t-\log\ (t-1)\Big)W\Big(\frac{t-1}{n}\Big)-\int_{(t-1)/n}^{t/n}W(r)d\log\ (nr)\Big]\Big|\leq$$

$$\sum_{t=2}^{n}\int_{(t-1)/n}^{t/n}\Big|W(r)-W\Big(\frac{t-1}{n}\Big)\Big|d\log(nr)\leq$$

$$\sup_{\substack{0\leq t_1,t_2\leq 1\\ 0<|t_2-t_1|\leq 1/n}}\big|W(t_2)-W(t_1)\big|\sum_{t=2}^{n}\int_{(t-1)/n}^{t/n}d\log(nr)=$$

$$\sup_{\substack{0\leq t_1,t_2\leq 1\\ 0<|t_2-t_1|\leq 1/n}}\big|W(t_2)-W(t_1)\big|\sum_{t=2}^{n}\big(\log\ t-\log\ (t-1)\big)=$$

$$\log\ n\sup_{\substack{0\leq t_1,t_2\leq 1\\ 0<|t_2-t_1|\leq 1/n}}\big|W(t_2)-W(t_1)\big|. \tag{7.63}$$

From Lemma 2 it follows that

$$\sup_{\substack{0\leq t_1,t_2\leq 1\\ 0<|t_2-t_1|\leq 1/n}}\big|W(t_2)-W(t_1)\big|=O_{a.s.}\Big(\frac{\sqrt{\log\ n}}{\sqrt{n}}\Big). \tag{7.64}$$

In addition, using integration by parts, it is not difficult to see that

$$\int_0^{1/n}W(r)d\log\ (nr)=O_P\Big(\frac{\log\ n}{\sqrt{n}}\Big). \tag{7.65}$$

From (7.62)-(7.65) and integration by parts it follows that

$$-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\log\ (t/n)u_t=\int_0^1 W(r)d\log\ (nr)+O_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big)=$$

$$-\int_0^1\log\ s\ dW(s)+O_P\Big(\frac{\log^2 n}{\sqrt{n}}\Big).$$

Since $\int_0^1\log\ sdW(s)=_d W\Big(\int_0^1\log^2\ sds\Big)=W(2)$, we get that (7.59) indeed holds (this relation also follows from (7.62), Lemma 2, the relation $\frac{1}{n}\sum_{t=1}^n\log^2\ (t/n)=2+O\big(\frac{\log^2 n}{n}\big)$ implied by (7.21) and (7.22), and the property that, similar to (7.60), $\sum_{t=2}^n\Big(\log\ t-\log\ (t-1)\Big)W\Big(\frac{t-1}{n}\Big)=-\sum_{t=1}^n\Big(W\Big(\frac{t}{n}\Big)-W\Big(\frac{t-1}{n}\Big)\Big)\log\ (t/n)$). Relations (7.55), (7.56) and (7.59) imply (7.28). ■

Proof of Lemma 7. By (7.33), (7.38) and (7.39),

$$\frac{B_n}{n}=\frac{1}{n}\sum_{t=1}^{n-1}z_t^2-\overline{z}_n^2=\frac{1}{n}\sum_{t=1}^{n-1}\frac{\tau_t^2}{t}+\frac{2}{n}\sum_{t=2}^{n-1}\frac{\tau_t}{t}\sum_{i=1}^{t-1}\tau_i-\frac{1}{n^2}\Big(\sum_{t=1}^{n-1}\tau_t\Big)^2. \tag{7.66}$$

Using Lemma 3 for i.i.d. exponential r.v.'s τ_t, $t \geq 1$, with $V_n = \sum_{t=1}^n Var(\tau_t) = n$, we conclude that

$$\frac{1}{n}\sum_{i=1}^{n-1} \tau_i = 1 + o_{a.s.}\Big(\frac{\log\ n}{\sqrt{n}}\Big), \tag{7.67}$$

and, consequently,

$$\frac{1}{n^2}\Big(\sum_{t=1}^{n-1} \tau_t\Big)^2 = 1 + o_{a.s}\Big(\frac{\log\ n}{\sqrt{n}}\Big). \tag{7.68}$$

Using (7.67), we also obtain

$$\frac{2}{n}\sum_{t=2}^{n-1} \frac{\tau_t}{t}\sum_{i=1}^{t-1} \tau_i = \frac{2}{n}\sum_{t=2}^{n-1} \tau_t\Big(1 + O_{a.s.}\Big(\frac{\log\ t}{\sqrt{t}}\Big)\Big). \tag{7.69}$$

As is easy to see, $\frac{1}{n}\sum_{t=2}^{n-1} \frac{\log\ t}{\sqrt{t}} = O\big(\frac{\log\ n}{\sqrt{n}}\big)$. Using Lemma 3 for independent r.v.'s $u_t = \tau_t \frac{\log\ t}{\sqrt{t}}$, $t \geq 1$, with $V_n = \sum_{t=1}^n Var(u_t) = \sum_{t=1}^n \frac{\log^2\ t}{t} = O(\log^3 n)$, we also have $\frac{1}{n}\sum_{t=2}^{n-1}(\tau_t - 1)\frac{\log\ t}{\sqrt{t}} = o_{a.s.}\big(\frac{\log^2\ n}{n}\big)$. Thus, $\frac{1}{n}\sum_{t=2}^{n-1} \tau_t \frac{\log\ t}{\sqrt{t}} = O_{a.s.}\big(\frac{\log\ n}{\sqrt{n}}\big)$. This, together with (7.67) and (7.69), implies that

$$\frac{2}{n}\sum_{t=2}^{n-1} \frac{\tau_t}{t}\sum_{i=1}^{t-1} \tau_i = 2 + O_{a.s.}\Big(\frac{\log\ n}{\sqrt{n}}\Big). \tag{7.70}$$

We further have

$$\frac{1}{n}\sum_{t=1}^{n-1} \frac{\tau_t^2}{t} = \frac{1}{n}\sum_{t=1}^{n-1} \frac{E\tau_t^2}{t} + \frac{1}{n}\sum_{t=1}^{n-1} \frac{\tau_t^2 - E\tau_t^2}{t} = \frac{1}{n}\sum_{t=1}^{n-1} \frac{\tau_t^2 - E\tau_t^2}{t} + O_{a.s.}\Big(\frac{\log\ n}{n}\Big). \tag{7.71}$$

Taking $a_t = \log\ t$ and $u_t = \frac{\tau_t^2 - E\tau_t^2}{t}$, $t \geq 1$, we have $\sum_{t=1}^\infty Var(u_t)/a_t^2 = \sum_{t=1}^\infty \frac{Var(\tau_1^2)}{t^2 \log^2\ t} < \infty$. Therefore, by Lemma 4, $\sum_{t=1}^{n-1} u_t/a_n = \sum_{t=1}^{n-1} \frac{\tau_t^2 - E\tau_t^2}{t \log\ n} \to 0$ a.s. as $n \to \infty$ and, consequently, $\frac{1}{n}\sum_{t=1}^{n-1} \frac{\tau_t^2 - E\tau_t^2}{t} = o_{a.s.}(\frac{\log\ n}{n})$. This, together with (7.71), implies that

$$\frac{1}{n}\sum_{t=1}^{n-1} \frac{\tau_t^2}{t} = O_{a.s.}\Big(\frac{\log\ n}{n}\Big). \tag{7.72}$$

From (7.66), (7.68), (7.70) and (7.72) it follows that (7.29) indeed holds. ■

Proof of Lemma 8. We have

$$D_n = \sum_{t=1}^n y_t^2 - n\overline{y}_n^2 = \sum_{t=1}^n \log^2(t-\gamma) - \frac{1}{n}\big(\sum_{t=1}^n \log\ (t-\gamma)\big)^2. \tag{7.73}$$

Using (7.43) and (7.73) we get, as in (7.55),

$$\frac{1}{\sqrt{n}}(A_n^\gamma + D_n) = \frac{1}{\sqrt{n}}\Big[\sum_{t=1}^{n-1}\frac{\tau_t}{t}\Big(\sum_{i=1}^{t}\log(i-\gamma)\Big) - \Big(\sum_{t=1}^{n-1}\tau_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n}\log(t-\gamma)\Big) + \sum_{t=1}^{n}\log^2(t-\gamma) - \frac{1}{n}\Big(\sum_{t=1}^{n}\log(t-\gamma)\Big)^2\Big] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t-1)\log(t/n) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t-1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{t=1}^{n}\log\,(t-\gamma)\Big) - \log(t/n)\Big] + \frac{1}{\sqrt{n}}\Big[\sum_{t=1}^{n}\log^2\,(t-\gamma) - \frac{1}{n}\Big(\sum_{t=1}^{n}\log\,(t-\gamma)\Big)^2 + \sum_{t=1}^{n}\frac{1}{t}\sum_{i=1}^{t}\log\,(i-\gamma) - \frac{n-1}{n}\sum_{t=1}^{n}\log\,(t-\gamma)\Big]. \tag{7.74}$$

This implies

$$E(A_n^\gamma + D_n) = \sum_{t=1}^{n}\log^2\,(t-\gamma) - \frac{1}{n}\Big(\sum_{t=1}^{n}\log\,(t-\gamma)\Big)^2 + \sum_{t=1}^{n}\frac{1}{t}\sum_{i=1}^{t}\log\,(i-\gamma) - \frac{n-1}{n}\sum_{t=1}^{n}\log\,(t-\gamma).$$

From (7.26) and (7.48) it follows that (7.30) indeed holds. Let us show that

$$V_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}(\tau_t-1)\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{t=1}^{n}\log\,(t-\gamma)\Big) - \log(t/n)\Big] = O_P\Big(\frac{1}{\sqrt{n}}\Big). \tag{7.75}$$

Similar to the arguments for (7.56), we get that the variance of V_n satisfies

$$Var(\sqrt{n}V_n) = \sum_{t=1}^{n-1}\Big[\frac{1}{t}\Big(\sum_{i=1}^{t}\log\,(i-\gamma)\Big) - \Big(\frac{1}{n}\sum_{i=1}^{n}\log\,(i-\gamma)\Big) - \log\,(t/n)\Big]^2 \leq C(M_n + Q_n),$$

where M_n is defined in (7.19) and Q_n is defined in (7.57). Using (7.24) and (7.58), we thus get that $Var(\sqrt{n}V_n) = O(1)$. Consequently, (7.75) indeed holds. Relations (7.26), (7.59), (7.74) and (7.75) imply (7.31). ■

Proof of Lemma 9. Relation (7.32) follows from (7.21), (7.22) and representation (7.73). ■

References

Aban, Inmaculada B., and Meerschaert, Mark M. (2004), "Generalized least-squares estimators for the thickness of heavy tails," *Journal of Statistical Planning and Inference*, 119, 341–352.

Alperovich, G. (1989), "The distribution of city size: A sensitivity analysis," *Journal of Urban Economics,* 25, 93–102.

Axtell, R. L. (2001), "Zipf distribution of U.S. firm sizes," *Science,* 293, 1818–1820.

Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999), "Tail index estimation and an exponential regression model," *Extremes,* 2, 177–200.

Beirlant, Jan, Goegebeur, Yuri, Teugels, Jozef, and Segers, Johan (2004) *Statistics of extremes* Wiley Series in Probability and Statistics, Chichester: John Wiley & Sons Ltd.

Borodin, Andrei N., and Salminen, Paavo (2002) *Handbook of Brownian motion—facts and formulae* Probability and its Applications, second ed., Basel: Birkhäuser Verlag.

Bosker, M., Brakman, S., Garretsen, H., de Jong, H., and Schramm, M. (2007), "The development of cities in Italy 1300-1861," *CESinfo Working Paper No. 1893.*

Brakman, S., Garretsen, H., van den Berg, M., and van Marrewijk, C. (1999), "The return of Zipf: Towards a further understanding of the rank-size distribution," *Journal of Regional Science,* 39, 182–213.

Čížek, P., W. Härdle, and R. Weron, eds (2005) *Statistical tools for finance and insurance,* Springer Verlag, Heidelberg.

Csörgő, M., and Révész, P. (1981) *Strong approximations in probability and statistics,* Academic Press, New York.

Csörgő, S., and Viharos, L. (1997), "Asymptotic normality of least-squares estimators of tail indices," *Bernoulli,* 3, 351–370.

— (2006), "Testing for small bias of tail index estimators," *Journal of Computational and Applied Mathematics,* 186, 232–252.

Davis, D. R., and Weinstein, D. E. (2002), "Bones, bombs, and break points: The geography of economic activity," *American Economic Review,* 92, 1269–1289.

Dobkins, L. H., and Ioannides, Y. M. (2000), "Dynamic evolution of the U.S. city size distribution," In *The economics of cities,* ed. J. Huriot and J. Thisse, Cambridge: Cambridge University Press pp. 217–260.

Eaton, J., and Eckstein, Z. (1997), "Cities and growth: Theory and evidence from France and Japan," *Regional Science and Urban Economics,* 27, 443–474.

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

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling extremal events for insurance and finance,* Springer, New York.

Fama, E. (1965), "The behavior of stock market prices," *Journal of Business,* 38, 34–105.

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

Gabaix, X., and Ioannides, Y. (2004), "The evolution of city size distributions," In *Handbook of regional and urban economics,* ed. V. Henderson and J.-F. Thisse, vol. 4, North-Holland pp. 2341–2378.

Gabaix, X., and Landier, A. (2008), "Why has CEO pay increased so much?," *Quarterly Journal of Economics,* 123, 49–100.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003), "A theory of power-law distributions in financial market fluctuations," *Nature,* 423, 267–270.

— (2006), "Institutional investors and stock market volatility," *Quarterly Journal of Economics,* 121, 267–270.

Hall, Peter (1982), "On some simple estimates of an exponent of regular variation," *Journal of the Royal Statistical Society. Series B.,* 44, 37–42.

Havil, J. (2003) *Gamma: Exploring Euler's constant,* Princeton University Press, Princeton, NJ.

Helpman, E., Melitz, M. J., and Yeaple, S. R. (2004), "Export versus FDI with heterogeneous firms," *American Economic Review,* 94, 300–316.

Hinloopen, J., and van Marrewijk, C. (2006), "Comparative advantage, the rank-size rule, and Zipf's law," *Tinbergen Institute Discussion Paper 06-100/1.*

Karatzas, I., and Shreve, S. E. (1991) *Brownian motion and stochastic calculus,* Springer-Verlag, New York.

Klass, O. S., Biham, O., Levy, M., Malcai, O., and Solomon, S. (2006), "The Forbes 400 and the Pareto wealth distribution," *Economics Letters,* 90, 290–295.

Komlós, J., Major, P., and Tusnády, G. (1975), "An approximation of partial sums of independent RV's and the sample DF. I.," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 32, 111–131.

— (1976), "An approximation of partial sums of independent RV's and the sample DF. II.," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 34, 33–58.

Kratz, Marie, and Resnick, Sidney I. (1996), "The QQ-estimator and heavy tails," *Communications in Statistics. Stochastic Models,* 12, 699–724.

Krugman, P. (1996) *The self-organizing economy,* Blackwell Publishers, Oxford, UK and Cambridge, MA.

Levy, M. (2003), "Are rich people smarter?," *Journal of Economic Theory,* 110, 42–64.

Levy, M., and Levy, H. (2003), "Investment talent and the pareto wealth distribution: Theoretical and experimental analysis," *Review of Economics and Statistics,* 85, 709–725.

Li, W. (2003), "Zipf's law everywhere," *Glottometrics,* 5, 14–21.

Loretan, M., and Phillips, P. C. B. (1994), "Testing the covariance stationarity of heavy-tailed time series," *Journal of Empirical Finance,* 1, 211–248.

Mandelbrot, B. (1960), "The Pareto-Levy law and the distribution of income," *International Economic Review,* 1, 79–106.

— (1963), "The variation of certain speculative prices," *Journal of Business,* 36, 394–419.

— (1997) *Fractals and scaling in finance. Discontinuity, concentration, risk,* Springer-Verlag, New York.

Nishiyama, Y., Osada, S., and Sato, Y. (2008), "OLS estimation and the t test revisited in rank-size rule regression," *Journal of Regional Science,* 48, 691–715.

Persky, J. (1992), "Retrospectives: Pareto law," *Journal of Economic Perspectives,* 6, 181–192.

Petrov, Valentin V. (1995) *Limit theorems of probability theory,* vol. 4 of *Oxford Studies in Probability*, New York: The Clarendon Press Oxford University Press. Sequences of independent random variables, Oxford Science Publications.

Phillips, P. C. B. (2007), "Regression with slowly varying regressors," *Econometric Theory,* 23, 557–614.

Rachev, S. T., Menn, C., and Fabozzi, F. J. (2005) *Fat-tailed and skewed asset return distributions: Implications for risk management, portfolio selection, and option pricing,* Wiley, Hoboken, NJ.

Rosen, K. T., and Resnick, M. (1980), "The size distribution of cities: An examination of the Pareto law abd primacy," *Journal of Urban Economics,* 8, 165–186.

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

Schultze, J., and Steinebach, J. (1999), "On least squares estimates of an exponential tail coefficient," *Statistcs & Extremes,* 2, 177–200.

Soo, K. T. (2005), "Zipf's Law for cities: a cross-country investigation," *Regional Science and Urban Economics,* 35, 239–263.

Viharos, L. (1999), "Weighted least-squares estimators of tail indices," *Probability and Mathematical Statistics,* 19, 249–265.

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