



**HAL**  
open science

# Top Incomes, Heavy Tails, and Rank-Size Regressions

Christian Schluter

► **To cite this version:**

Christian Schluter. Top Incomes, Heavy Tails, and Rank-Size Regressions. *Econometrics*, 2018, 6 (1), pp.10. 10.3390/econometrics6010010 . hal-01978497

**HAL Id: hal-01978497**

**<https://amu.hal.science/hal-01978497>**

Submitted on 28 Jun 2023

**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.



Distributed under a Creative Commons Attribution 4.0 International License

Article

# Top Incomes, Heavy Tails, and Rank-Size Regressions

Christian Schluter <sup>1,2</sup>

<sup>1</sup> Aix-Marseille School of Economics, 5 Boulevard Maurice Bourdet CS 50498, 13205 Marseille CEDEX 01, France; christian.schluter@univ-amu.fr; Tel.: +33-413-55-25-72

<sup>2</sup> Department of Economics, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received: 19 November 2017; Accepted: 20 February 2018; Published: 2 March 2018

**Abstract:** In economics, rank-size regressions provide popular estimators of tail exponents of heavy-tailed distributions. We discuss the properties of this approach when the tail of the distribution is regularly varying rather than strictly Pareto. The estimator then over-estimates the true value in the leading parametric income models (so the upper income tail is less heavy than estimated), which leads to test size distortions and undermines inference. For practical work, we propose a sensitivity analysis based on regression diagnostics in order to assess the likely impact of the distortion. The methods are illustrated using data on top incomes in the UK.

**Keywords:** top incomes; heavy tails; rank size regression; extreme value index; regular variation

**JEL Classification:** D31; C13; C14

## 1. Introduction

Income distributions exhibit, like many other size distributions in economics and the natural science, upper tails that decay like power functions (see e.g., [Schluter and Trede 2017](#)). The recent and rapidly growing literature on top incomes focuses on this upper tail, and its presence has important consequences for the measurement of inequality.<sup>1</sup> However, estimating the heaviness of the upper tail is challenging, since real world size distributions usually are Pareto-like (i.e., tails are regularly varying) rather than strictly Pareto.

To be precise, let  $X_1, \dots, X_n$  be a sequence of positive independent and identically distributed random variables (e.g., incomes) with distribution function  $F$  that is regularly varying, so for large  $x$

$$1 - F(x) = x^{-\frac{1}{\gamma}}l(x), \quad \gamma \in (0, \infty), \quad (1)$$

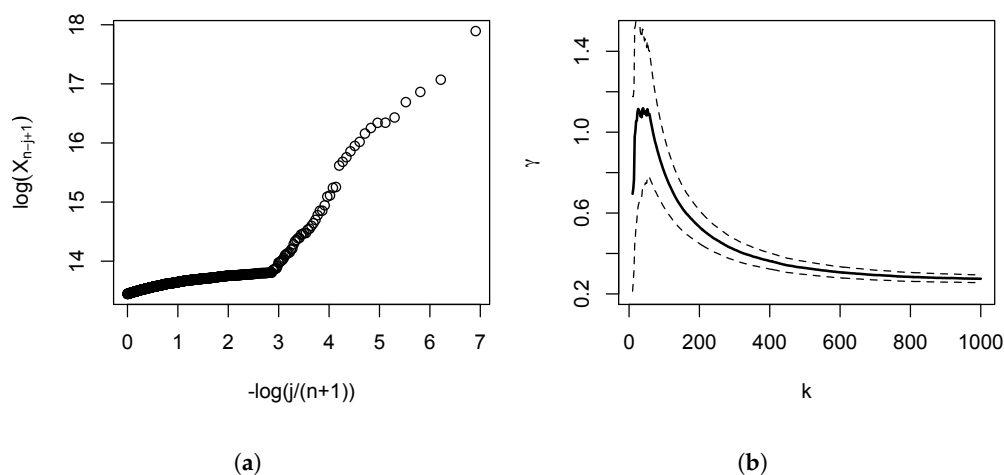
where  $l$  is slowly varying at infinity, i.e.,  $l(tx)/l(x) = 1$  as  $x \rightarrow \infty$ . The parameter  $\gamma$ , usually referred to as extreme value index (and  $1/\gamma$  as the tail exponent), is unknown and needs to be estimated. Many estimators have been proposed in the statistical literature (see e.g., the textbook treatments in [Embrechts et al. 1997](#) or [Beirlant et al. 2004](#)).

An estimator popular among economists is based on a simple ordinary least squares (OLS) regression of log sizes on log ranks (e.g., [Jenkins 2017](#) and [Atkinson 2017](#), and references therein, in the income distribution and top incomes literature, this regression is ubiquitous in the city size literature). The enduring popularity of the OLS estimator is partly due to its simplicity, and partly due

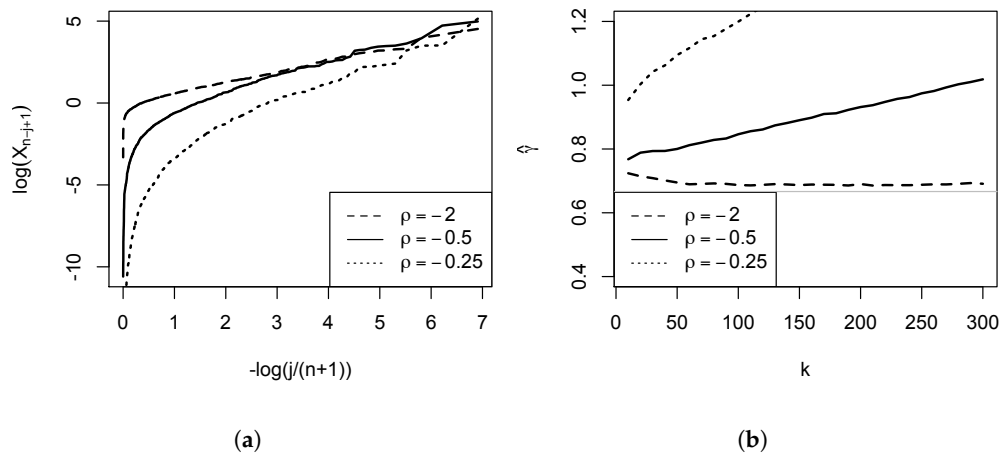
<sup>1</sup> See e.g., [Schluter and Trede \(2002\)](#) in the contexts of Lorenz curves, [Davidson and Flachaire \(2007\)](#) who propose a semi-parametric bootstrap, [Cowell and Flachaire \(2007\)](#) who advocate semi-parametric methods, or [Burkhauser et al. \(2012\)](#) who seek to reconcile survey and tax return data. Also observe that the  $p^{\text{th}}$  moment of the income distribution is finite only if  $p < 1/\gamma$ , so very heavy tails can directly affect the validity of some standard inequality measurement tools. For instance, statistical inference for the Generalised Entropy index with parameter 2 requires the existence of the fourth moment ([Cowell 1989](#)).

to a powerful intuition based on a Pareto quantile-quantile (QQ)-plot, the regression estimating its slope coefficient. However, if the tail of the distribution varies regularly, the Pareto QQ-plot will become linear only *eventually*. In particular, (1) can be expressed equivalently, using the tail quantile function  $U(x) = \inf\{t : \Pr(X > t) = 1/x\}$  where  $x > 1$ , as  $U(x) = x^\gamma \tilde{l}(x)$  where  $\tilde{l}(x)$  is a slowly varying function. Hence, as  $x \rightarrow \infty$ ,  $\log U(x) \sim \gamma \log(x)$  since then  $\log \tilde{l}(x) \rightarrow 0$ . Replacing these population quantities with their empirical counterparts gives the Pareto QQ-plot, and  $\gamma$  is its ultimate slope. This qualification (usually ignored by practitioners in economics) has important consequences for the behaviour of the estimator: Since the OLS estimator estimates the slope parameter of this QQ-plot, deviations from the strict Pareto model -captured by the nuisance function  $l$ - will induce distortions.

The empirical importance of this is illustrated in Figure 1, which depicts the Pareto QQ-plot for our administrative income data for the UK (the subject of our empirical application developed in Section 4 below), using the 1000 largest incomes. The plot exhibits a pronounced kink, and approximate linearity of the QQ-plot only holds for the very highest upper order statistics. Panel (b) shows the consequences for the OLS estimates: As we move in the QQ-plot from the right to the left, the departures from linearity become progressively more severe, and the OLS estimates progressively fall. Based on this first diagnostic QQ-plot, once the lower upper order statistics have been discarded as a source of downward bias, the subsequent analysis can then more clearly focus on the approximate linear part, the remaining distortions, and the choice of the number of order statistics. Figure 2 provides a further illustration for three Burr (Singh-Maddala) distributions (examined in detail in Section 3 below, being the leading parametric income distribution model) possessing the same  $\gamma$ . Here, the speed of decay of the nuisance function  $l$  is parametrised by the absolute value of the parameter  $\rho$ . The smaller the magnitude of  $\rho$ , the greater the initial curvature and steepness of the Pareto QQ-plot, and the larger the induced positive distortions of the OLS estimator of the slope coefficient.



**Figure 1.** Pareto quantile-quantile (QQ)-Plot: Top incomes in the UK. Based on administrative income tax return for the UK in 2009/10. The Survey of Personal Incomes (SPI) is described in Section 4. Panel (a): The Pareto QQ-plot (see Section 1.1) is based on the largest 1000 incomes. Panel (b): Estimates of  $\gamma$  for the  $k$  upper order statistics using the OLS regression (solid lines), and pointwise 95% symmetric confidence intervals (dashed lines). The distributional theory is stated in Equation (8).



**Figure 2.** Pareto QQ-Plots: The Burr distribution. Based on the Burr distribution given by  $F_{(\gamma,\rho)}(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$  with  $\gamma = 2/3$ , and  $\rho \in \{-2, -0.5, -0.25\}$ . Panel (a): Pareto QQ-plots for 3 random samples drawn from the Burr distribution. Sample size is 1000. To aid comparison across cases, the points of each QQ-plot have been connected and rendered as lines. Panel (b): Mean of estimates  $\hat{\gamma}$  across 1000 Monte Carlo simulations for given  $\rho$ , drawing samples of size 1000 in each iteration. The faint horizontal line is the population value  $\gamma = 2/3$ .

In this paper, we examine the asymptotic distortions of the OLS estimator that arise in these circumstances, caused by the slow decay of the nuisance function  $l$  and modeled here as higher order regular variation. The theory is presented in Section 2 (proofs are collected in Appendix A), and numerical illustrations and quantifications of the distortions are provided in Section 3, as well as of the stark consequence for inference. More specifically, we show formally that the OLS estimator over-estimates the true value in the leading heavy-tailed model (i.e., the Hall class, which includes the Burr (Singh-Maddala) distribution, as well as the student, Fréchet, and Cauchy distributions). An empirical illustration in the context of top incomes in the UK using data on tax returns is the subject of Section 4.

### 1.1. The Log-Log Rank-Size Regression

We briefly review the rank size regression. Let  $X_{1,n} \leq \dots \leq X_{n,n}$  denote the order statistics of  $X_1, \dots, X_n$ , and consider the  $k$  upper order statistics. Let ranks be shifted by a constant  $\eta < 1$ . The regression of sizes on ranks leads to the minimisation of the least squares criterion

$$\sum_{j=1}^k \left( \log \frac{X_{n-j+1,n}}{X_{n-k,n}} - g \log \frac{k+1}{j-\eta} \right)^2 \tag{2}$$

with respect to  $g$ , where  $\eta < 1$  and  $1 \leq j \leq k < n$ . The classic case is  $\eta = 0$ . However, since the OLS estimator of the slope coefficient is not invariant to shifts in the data, it is conceivable that a purposefully chosen shift could yield an asymptotic refinement (Gabaix and Ibragimov 2011 consider this in the strict Pareto model  $1 - F(x) = cx^{-1/\gamma}$ ). The analysis below allows for this possibility.

The justification of considering regression (2) is based on a Pareto QQ-plot (Beirlant et al. 1996): For a sufficiently high threshold  $X_{n-k,n}$  where  $k < n$ , the Pareto quantile plot in model (1) with coordinates  $(-\log(j/(n+1)), \log X_{n-j+1,n})_{j=1,\dots,k}$  becomes ultimately linear. The line through point  $(-\log((k+1)/(n+1)), \log X_{n-k,n})$  with slope  $g$  is thus given by  $y = \log X_{n-k,n} + g[x + \log((k+1)/(n+1))]$

1)/(n + 1))]] and the data points are  $(x, y) = (-\log(j/(n + 1)), \log X_{n-j+1,n})_{j=1, \dots, k}$ . The regression estimator estimates this slope parameter. In particular, the OLS estimator of the slope coefficient  $g$  is

$$\hat{\gamma} = \frac{\frac{1}{k} \sum_{j=1}^k \log\left(\frac{k+1}{j-\eta}\right) [\log X_{n-j+1,n} - \log X_{n-k,n}]}{\frac{1}{k} \sum_{j=1}^k \left[\log\left(\frac{k+1}{j-\eta}\right)\right]^2} \equiv \frac{N_{n,k}}{D_k}, \quad \eta < 1. \tag{3}$$

Note that the denominator  $D_k$  is a Riemann approximation to  $\int_0^1 \log^2 x dx = 2$ . An asymptotic expansion of the denominator reveals that

$$D_k = 2 + O\left(\frac{\log^2 k}{k}\right) \tag{4}$$

From [Kratz and Resnick \(1996, proof of their Equation 2.4, p. 704\)](#) we know that the numerator  $N_{n,k}$  converges in probability to  $2\gamma$ , hence the estimator is weakly consistent:  $\hat{\gamma} \xrightarrow{P} \gamma$  as  $k \rightarrow \infty$  and  $k/n \rightarrow 0$ . We proceed in the next Section to refine this result by obtaining higher order expansions of the estimator in (3).

The literature contains several variants of regression (2). Rather regressing log sizes on log ranks, one could regress log ranks on log sizes, thus obtaining the ‘dual’ regression. In view of (3), our asymptotic analysis of the numerator carries immediately over to this dual regression. Another variant of (2) includes the additional estimation of a regression constant:  $\log X_{n-j+1,n}$  is regressed on a constant and  $\log j$ . [Kratz and Resnick \(1996\)](#) obtain the distributional theory for this alternative estimator and show that its asymptotic variance is  $2\gamma^2/k$ , which exceeds, as will be shown below, the asymptotic variance of  $\hat{\gamma}$  given by (3). Hence this regression variant is less efficient. [Schultze and Steinebach \(1996\)](#) also prove weak consistency of the estimator in this setting.

## 2. Asymptotic Expansions and Distributional Theory

### 2.1. Preliminaries: Higher Order Regular Variation

In order to obtain our asymptotic expansions, we use an equivalent representation of model (1) based on regular variation and extreme value theory. First we recall the definition of first-order regular variation, and then proceed to model the slowly varying nuisance function  $l$  in (1) by a refinement to second-order regular variation. We then show that most heavy-tailed distributions of interest (in the income, finance and urban literature) satisfy this condition.

It is well known that model (1) has the equivalent (first-order regular variation) representation (e.g., [Dekkers et al. 1989](#))

$$\lim_{t \rightarrow \infty} \frac{\log U(tx) - \log U(t)}{a(t)/U(t)} = \log x, \tag{5}$$

for all  $x > 0$  where  $a$  is a positive norming function with the property  $a(t)/U(t) \rightarrow \gamma$ . The problem for estimating the extreme value index  $\gamma$  is the behaviour of the slowly varying function  $l$  in (1). It is, therefore, common practice in the extreme value literature to model such second-order behaviour, thus strengthening model (1), by strengthening the first-order regular representation (5) to second-order regular variation. Following [De Haan and Stadtmüller \(1996\)](#), we assume that the following refinement of (5) holds

$$\lim_{t \rightarrow \infty} \frac{\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \log x}{A(t)} = H_{\gamma, \rho}(x) \tag{6}$$

for all  $x > 0$ , where  $H_{\gamma > 0, \rho < 0}(x) = \frac{1}{\rho} \left(\frac{x^\rho - 1}{\rho} - \log x\right)$  with  $\rho < 0$ . This parameter  $\rho$  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  $t \rightarrow \infty$ . As  $\rho$  falls in magnitude, the nuisance part of  $l$  in (1) decays more slowly. Our numerical illustrations will thus consider small magnitudes for  $\rho$ .

**Examples.** Most heavy-tailed distributions of interest satisfy representation (6). Consider the Hall class of distributions (Hall 1982), given by, for large  $x$ ,

$$F(x) = 1 - ax^{-1/\gamma}[1 + bx^\beta + o(x^\beta)]$$

with  $\gamma, a > 0, b \in R, \beta < 0$ . In this class, the nuisance function  $l$  in model (1) converges to a constant at a polynomial rate. The Hall class nests, for instance, the Burr (Singh-Maddala), Student, Fréchet, and Cauchy distributions.<sup>2</sup> The tail quantile function is  $U(x) = cx^\gamma[1 + dx^\rho + o(x^\rho)]$  where  $c = a^\gamma, d = b\gamma a^{\gamma\beta}$ . This Hall class satisfies the second order representation (6) with  $\rho = \gamma\beta < 0$ , and rate function

$$A(t) = \frac{\rho^2}{\gamma} dt^\rho.$$

Figure 2 illustrates the role of  $\rho$  for the Burr distribution (examined in greater detail in Section 3) in terms of the Pareto QQ-plot, and the implications for the estimator  $\hat{\gamma}$  of its slope parameter. For  $\rho = -2$  the plot is close to linear, and the estimates close to the population value. However, as  $\rho$  falls in magnitude, the initial curvature increases, and the slope estimates consequently becomes more positively distorted as the number of upper order statistics  $k$  entering the estimator increases.

### 2.2. The Main Results

We first state the higher order asymptotic expansion of the numerator  $N_{n,k}$ . We then obtain the distributional theory for our estimator  $\hat{\gamma}$ , before returning to the distortions induced by deviations from the strict Pareto model (captured by second order regular variation).

**Asymptotic expansion.** In the Appendix A we prove the following higher order expansion of the numerator  $N_{n,k}$  under the assumption of second-order regular variation (6). Throughout, we will consider an intermediate sequence  $k = k_n$  of positive integers such that  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . It is then true that, for  $\gamma > 0$  and  $\rho < 0$ ,

$$\begin{aligned} N_{n,k}/\gamma &= 2 - \left(\frac{1}{2} - \eta\right) \frac{\log(k - \eta)}{k} - \left(\frac{1}{2} - \eta\right) \frac{\log^2 k}{2k} \\ &+ O_p\left(\frac{1}{k^{1/2}}\right) + O\left(\frac{1}{k}\right) + O_p\left(\frac{\log k}{k^{1/2}}\right) \\ &+ A\left(\frac{n}{k}\right) \frac{1}{\rho} \left[\frac{2 - \rho}{(1 - \rho)^2}\right] + O_p\left(\frac{\log k}{k}\right) + o_p(A(n/k)) \end{aligned} \tag{7}$$

A few comments are in order. The first two lines of this expression characterise the first-order behaviour of the numerator. It can be seen that setting the regression shift factor  $\eta$  to 1/2 eliminates the second and third term. However, the term  $O_p(\log k/k^{1/2})$  is still present. The asymptotic refinement due to second-order regular variation is given by the terms of line 3. Although  $A(t) \rightarrow 0$  as  $t \rightarrow \infty$ , this decay might be slow:  $A(t)$  is regularly varying with index  $\rho$ , and as  $\rho$  falls in magnitude the nuisance part of  $l$  in (1) decays more slowly. A slow decay then introduces a noticeable distortion in finite samples. We examine these distortions after stating the distributional theory for the estimator.

---

<sup>2</sup> The Burr distribution  $F_{(\gamma,\rho)}(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$  is a member of the Hall class with parameters  $\gamma$  and  $\rho < 0, c = 1$  and  $d = \gamma/\rho$ , as is the Student  $t_\delta$  distribution with  $\delta$  degrees of freedom where  $\gamma = 1/\delta, \rho = -2/\delta, d = \gamma B C^{-2\gamma}, B = -0.5\delta^2(\delta + 1)/(\delta + 2)$ , and  $C = \Gamma((\delta + 1)/2)\delta^{(\delta-1)/2}/(\delta\pi)^{1/2}\Gamma(\delta/2)$  (valid for  $\delta > 2$ ); so is the Fréchet distribution  $F_\gamma(x) = \exp(-x^{-1/\gamma})$  with  $\rho = -1, c = 1$ , and  $d = -0.5\gamma$ , and the Cauchy distribution with  $\gamma = 1, \rho = -2, c = 1/\pi$ , and  $d = -0.5\pi^2$ .

**Distributional theory.** Beirlant et al. (1996) observe that our slope estimator  $\hat{\gamma}$ , given by (3), is (to first order) a member of the class of kernel estimators discussed in Csorgo et al. (1985) with kernel  $K(t) = 1 - \log t$ . Since  $\int_0^1 K(t)dt = 2$  and not unity, a scale correction is required. Since  $\int_0^1 K^2(t)dt = 5$ , the following result obtains as  $k \rightarrow \infty$  and  $k/n \rightarrow 0$ , and if  $\sqrt{k}A(n/k) \rightarrow 0$

$$\sqrt{k}(\hat{\gamma} - \gamma) \rightarrow^d N\left(0, \frac{5}{4}\gamma^2\right) \quad (8)$$

**Higher order distortions.** Asymptotically, the estimator is thus unbiased if  $\sqrt{k}A(n/k) \rightarrow 0$ . If this decay is slow, however, the estimator will suffer from a higher order distortion in finite samples. By (7), this distortion equals, for  $\gamma > 0$  and  $\rho < 0$ ,

$$b_{k,n} \equiv \frac{1}{2} \frac{\gamma}{\rho} \frac{2 - \rho}{(1 - \rho)^2} A(n/k) \quad (9)$$

In particular, in the Hall model,  $A(t) = (\rho^2/\gamma)dt^\rho$ . The sign of the higher order distortion of  $N_{n,k}$  and hence  $\hat{\gamma}$  is, since  $\rho < 0$ , then given by  $-\text{sgn}(d)$ . For the Burr (Singh-Maddala), student, Fréchet, and Cauchy distributions it can be shown that  $d < 0$ , leading to a positive higher order distortion. We conclude that the higher order distortion induced by higher order regular variation is positive for many popular distribution -i.e., for which the nuisance function  $l$  in model (1) converges to a constant at a polynomial rate- leading to an overestimation of  $\gamma$ .<sup>3</sup>

Simulation evidence for these theoretical results is presented next. We also quantify the higher order distortions and the consequences for statistical inference about  $\gamma$ .

### 3. Numerical Illustrations

We illustrate numerically several of our results in a Monte Carlo study. First, we verify the distributional theory, then show that most of the empirical distortion is captured by the bias function  $b_{k,n}$ . At the same time, we show that the distortions can be sizeable, leading to substantial test size distortions, while a bias correction using  $b_{k,n}$  would reconcile nominal and actual test sizes.

Our Monte Carlo study is based on the Burr distribution, a member of the Hall class, parametrised here as  $F_{(\gamma,\rho)}(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$  with parameters  $\gamma$  and  $\rho < 0$ . In the income distribution and inequality literature, this distribution is also known as the Singh-Maddala distribution, and used frequently in parametric income models. Specifically, we set  $\gamma = 2/3$ , and  $\rho = -1/2$  to begin with. Qualitatively similar results are obtained for the student, Fréchet, and Cauchy distributions, all of which are members of the Hall class, and therefore not reported here. Since  $1 < 1/\gamma < 2$  we consider a situation of fairly heavy tails (as second moments of the distribution do not exist). However, the qualitative insights depend little on the actual choice of  $\gamma$ . We have chosen  $\rho = -1/2$  as our leading example since we are interested in the consequences of deviating from a strict Pareto model. As  $\rho$  falls in magnitude the nuisance part of  $l$  in (1) decays more slowly. This is illustrated in Figure 2, where we depict three Pareto QQ-plots for different  $\rho$ . For  $\rho = -2$ , the plot is almost linear throughout. The deviations from the strict Pareto model become increasingly more pronounced in the left part of the plot as  $\rho$  falls in magnitude.

For the simulation study, we draw  $R = 1000$  samples of size  $n = 10,000$  at first (then  $n = 1000$ ), and consider the upper  $k$  order statistics. In order to choose a particular  $k$ , we follow standard practice and

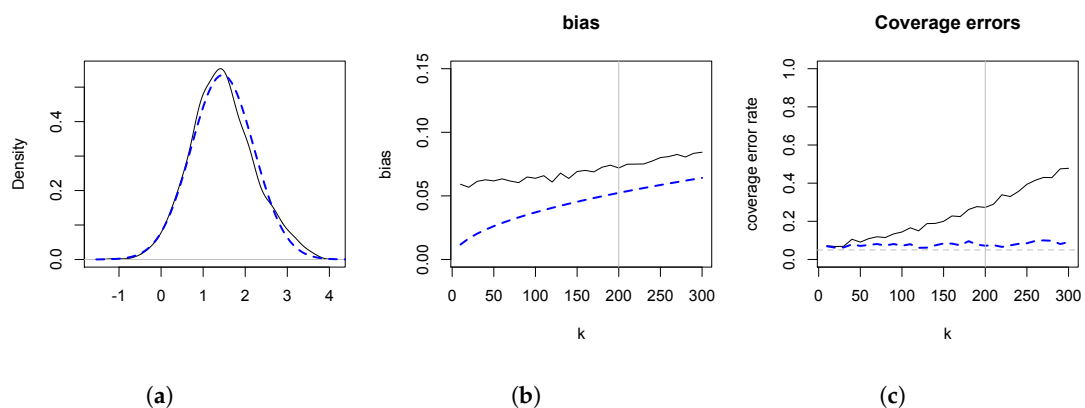
<sup>3</sup> De Haan and Ferreira (2006) consider the merit of shifting the tail for tail quantile functions  $U(t) = c_0 + c_1 t^\gamma + c_2 t^{\gamma+\tau} + o(t^{\gamma+\tau})$  where  $c_0$  and  $c_2$  are not zero,  $c_1 > 0$ , and  $\gamma > 0$  and  $\tau < 0$ . It can then be shown that if  $\tau < -\gamma$ , the second order parameter satisfies  $\rho = -\gamma$ . A data shift that eliminates  $c_0$  then results in  $\rho = \tau$ , so the post shift second order parameter has increased in magnitude, leading to a decrease in the induced distortion. However, the reverse reasoning also applies. In particular, the Hall model is  $U(x) = cx^\gamma[1 + dx^\rho + o(x^\rho)]$ . A data shift by  $c_0$  yields  $U(x) + c_0 = cx^\gamma[1 + (c_0/c)x^{-\gamma} + dx^\rho + o(x^\rho)]$ , and increases the distortion if  $\gamma \leq |\rho|$ .

minimise the theoretical asymptotic Mean Squared Error (AMSE) (e.g., Hall 1982, or Beirlant et al. 1996), given by  $b_{k,n}^2 + (1/k)(5/4)\gamma^2$ , trading off distortion and dispersion. The theoretical higher order bias in  $\hat{\gamma}$  induced by higher order regular variation in this Burr case is

$$b_{k,n} = \frac{1}{2}\gamma \frac{2-\rho}{(1-\rho)^2} \left(\frac{n}{k}\right)^\rho$$

which is, of course, increasing in  $k$ . The theoretical AMSE is minimised around  $k^* = 200$ , which also corresponds to the minimiser of the empirical AMSE based on the  $R$  samples. The mean of  $\hat{\gamma}$  at this  $k^*$  is 0.739, and exceeds, as predicted by the theory, the population value  $\gamma = 2/3$ .

Figure 3 depicts the results. In panel (a) we illustrate the distributional theory, given by (8), for  $k^*$ , by plotting a kernel density estimate of  $\sqrt{k^*}\hat{\gamma}$  (solid line), as well as a normal density with variance  $(5/4)\gamma^2$ , centered on the empirical mean of the simulated data. The two are in close agreement. The figure also implies that any inferential problems are due to location shifts. In panel (b) we contrast the empirical distortions (solid line) with  $b_{k,n}$  (dashed line).  $\hat{\gamma}$  overestimates  $\gamma$ , and the distortion increases in  $k$ . It is evident that most of the distortion is captured by  $b_{k,n}$ . In panel (c) we illustrate the consequences of the distortions for statistical inference, by plotting the empirical coverage error rates of the usual 95% symmetric confidence intervals. The higher order distortions lead to undermining inference because of the considerable size distortions. For instance, at  $k^*$ , the empirical coverage error rate is 30% for a nominal 5% rate. Shifting the estimate by  $b_{k^*,n}$  reduces the coverage error rate to 7%.



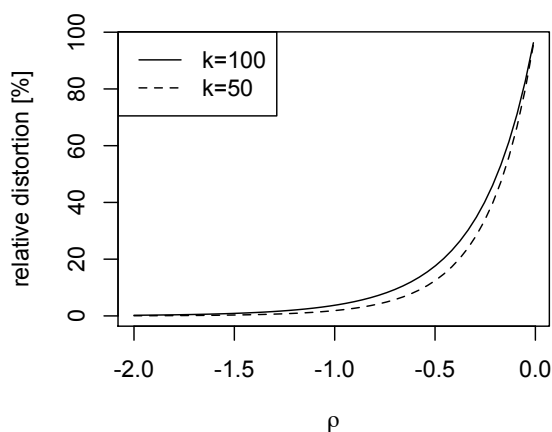
**Figure 3.** Bias and Inference: Burr. Monte Carlo study for the Burr distribution with parameters  $\gamma = 2/3$  and  $\rho = -0.5$ . Based on samples of size  $n=10,000$  and  $R=1000$  repetitions.  $k^* = 200$  minimises the asymptotic Mean Squared Error (AMSE), and is depicted by the vertical lines in panels b and c. Panel (a): Density plot of  $\sqrt{k^*}\hat{\gamma}$  (solid line) and shifted normal density with variance  $(5/4)\gamma^2$  (dashed line). Panel (b): empirical bias (solid line) and higher order bias function  $b_{k,n}$  (dashed line). Panel (c): Coverage error rate of the usual 95% symmetric confidence intervals for nominal rate of 5%, with no bias correction (solid line) and correction by the theoretical  $b_{k,n}$  (dashed line).

Next, we consider the role of the sample size  $n$ . Reducing the sample sizes in the Monte Carlo to  $n = 1000$  yields results that are in line with the above theory, and therefore not depicted. The bias of  $\hat{\gamma}$  increases by a factor predicted by the theory, namely  $b_{k,1000}/b_{k,10,000} = 10^{1/2} = 3.16$ . The optimal  $k^*$  shrinks by a factor of 4, as now  $k^* = 50$ . The density of  $\sqrt{k^*}\hat{\gamma}$  is in good agreement with the theory, and empirical coverage error rates at this  $k^*$  are 32% for the uncorrected and 11% for the corrected estimator. The empirical coverage error rate for the uncorrected estimator rises steeply after  $k^*$ , reaching 64% at  $k = 100$ . Reducing the sample sizes further to 100 results in  $k^* = 20$ , and an empirical coverage error rate for the uncorrected estimator of 46% at this  $k^*$ . Biases are increased by a factor  $b_{k,100}/b_{k,10,000} = 10$ .

Finally, we illustrate the importance of the speed of decay in the nuisance function  $l$  of model (1). As  $\rho$  falls in magnitude, the nuisance function  $l$  decays more slowly. For the Burr case with  $\gamma = 2/3$ ,



we depict in Figure 4  $b_{k,n}$  as  $\rho$  falls in magnitude for  $n = 1,000$  and selected  $k$ . While for  $\rho = -2$  the distortions are negligible (in line with Figure 2, it is evident that for small magnitudes of  $\rho$  the higher order distortions cannot be ignored).



**Figure 4.** Relative distortions in the Burr model. Burr model with  $\gamma = 2/3$  and  $n = 1000$ . Depicted is  $100 * b_{k,n} / \gamma$  as  $\rho$  varies.

As the purpose of our simulation study is the provision of numerical evidence for our theory, we have used the theoretical bias function  $b_{k,n}$  in the Burr case. When no such external knowledge is available, estimating the bias function requires non-parametric estimates of the second order parameter  $\rho$  and the function  $A(\cdot)$ . However, existing methods perform poorly, yielding excessively volatile estimates. The theory then informs a sensitivity analysis which is described in Section 4.1 in the context of our empirical application.

#### 4. Empirical Illustration: Top incomes in the UK

Our empirical application uses administrative income tax return data are from the public-release files of the Survey of Personal Incomes (SPI) for the year 2009/10 (see e.g., Jenkins 2017 for a detailed description, and an analysis that includes rank size regressions). The SPI data underlie the UK top income share estimates in the World Top Incomes Database (WTID), and is a stratified sample of the universe of tax returns. The unit of taxation is the individual, and we use total taxable income as the income variable. The file contains 674,715 individuals, and we consider the  $n$  largest incomes.

In Figure 1 panel (a), we have depicted the Pareto QQ-plot for the 1000 largest incomes. It is evident that the data clearly reject a strict Pareto model: The plot exhibits a pronounced kink, and approximate linearity of the QQ plot only holds for the very highest upper order statistics. The function  $l$  in (1) captures this significant departure from the strict Pareto model. The Pareto QQ-plot thus conveys crucial information that is usually ignored by practitioners in economics, making it a key diagnostic device. For instance, a common mechanical approach is to set  $k$  by choosing ‘blindly’ (i.e., without reference to the Pareto QQ-plot) e.g., the top 1% or the top 1000 observations. Since the approximate linearity only obtains for about the 70 largest observations, the estimate of the slope parameter of the Pareto QQ-plot, i.e., the OLS estimator (3), will be severely biased if  $k$  is set to 1000 or higher. This is illustrated in panel (b) of the figure: The estimates fall for higher values of  $k$ , since the estimation procedure then attributes increasing weights to the left of the kink in the Pareto QQ-plot.

In the light of these observations, we restrict our subsequent analysis to the range of  $k$  in which the Pareto QQ-plot is approximately linear. We confirm this in Figure 5 panel (a), having restricted the plot to the  $n = 70$  highest incomes. The plot now appears fairly linear. In panel (b), we depict the regression estimates  $\hat{\gamma}$  and the 95% symmetric pointwise confidence intervals. One first visual way of

choosing an estimate is to consider an area of the plot where the estimate is fairly stable (as is done by inspecting Hill or so-called alternative Hill plots) and picking the largest such  $k$  since the variance of the estimate falls in  $k$ . Such subjective choice would be around  $k = 60$  with an estimate of  $\hat{\gamma} = 1.070$  (indicated by the horizontal faint line in the figure).<sup>4</sup> Overall, the visual method would suggest an estimate of  $\gamma$  between 0.9 and 1, implying very heavy tails. Taking into consideration the variability of the estimate, one cannot reject the hypothesis that the tail index be unity, i.e., Zipf’s law. Returning to panel (a) we have also plotted the line with slope 1. This line does well in describing the data. We turn to a method that permits an objective choice of a particular  $k$ , and examine the remaining distortions in the estimate of  $\gamma$ .

#### 4.1. Sensitivity Analysis, and the Choice of $k$

The preceding analysis has shown that  $\hat{\gamma}$  is likely to suffer from positive higher order distortions, captured by  $b_{k,n}$ . Estimating this bias function requires non-parametric estimates of the second order parameter  $\rho$  and the function  $A(\cdot)$ , but existing methods perform poorly, yielding excessively volatile estimates. Hence we limit ourselves to a sensitivity analysis, taking  $\rho$  as a sensitivity parameter, whose objective is to gauge plausible values of the potential distortions based on diagnostics of the rank size regression. This approach is sketched next.

Following Beirlant et al. (1996), we observe that the mean weighted theoretical squared deviation

$$\frac{1}{k} \sum_{j=1}^k w_{j,k} E \left( \log \left( \frac{X_{n-j+1,n}}{X_{n-k,n}} \right) - \gamma \log \left( \frac{k+1}{j} \right) \right)^2$$

equals, to first order,

$$c_k \text{Var}(\hat{\gamma}) + d_k(\rho) b_{k,n}^2 \tag{10}$$

for some coefficients  $c_k$  depending only on  $k$ , and  $d_k(\rho)$  depending on  $k$  and  $\rho$  (these are stated explicitly in the Appendix A). Set  $w_{j,k} \equiv 1$ . An estimate of the mean theoretical deviation is the mean of the squared residuals  $k^{-1} SSR_k$  of the rank size regression. In view of the usual bias-variance trade-off for our estimator  $\hat{\gamma}$  for fixed  $n$ , we ascribe all the measured deviation  $k^{-1} SSR_k$  to the bias, thereby defining a very conservative bound, and let

$$\tilde{b}_{k,n}(\rho) = [k^{-1} SSR_k / d_k(\rho)]^{1/2}$$

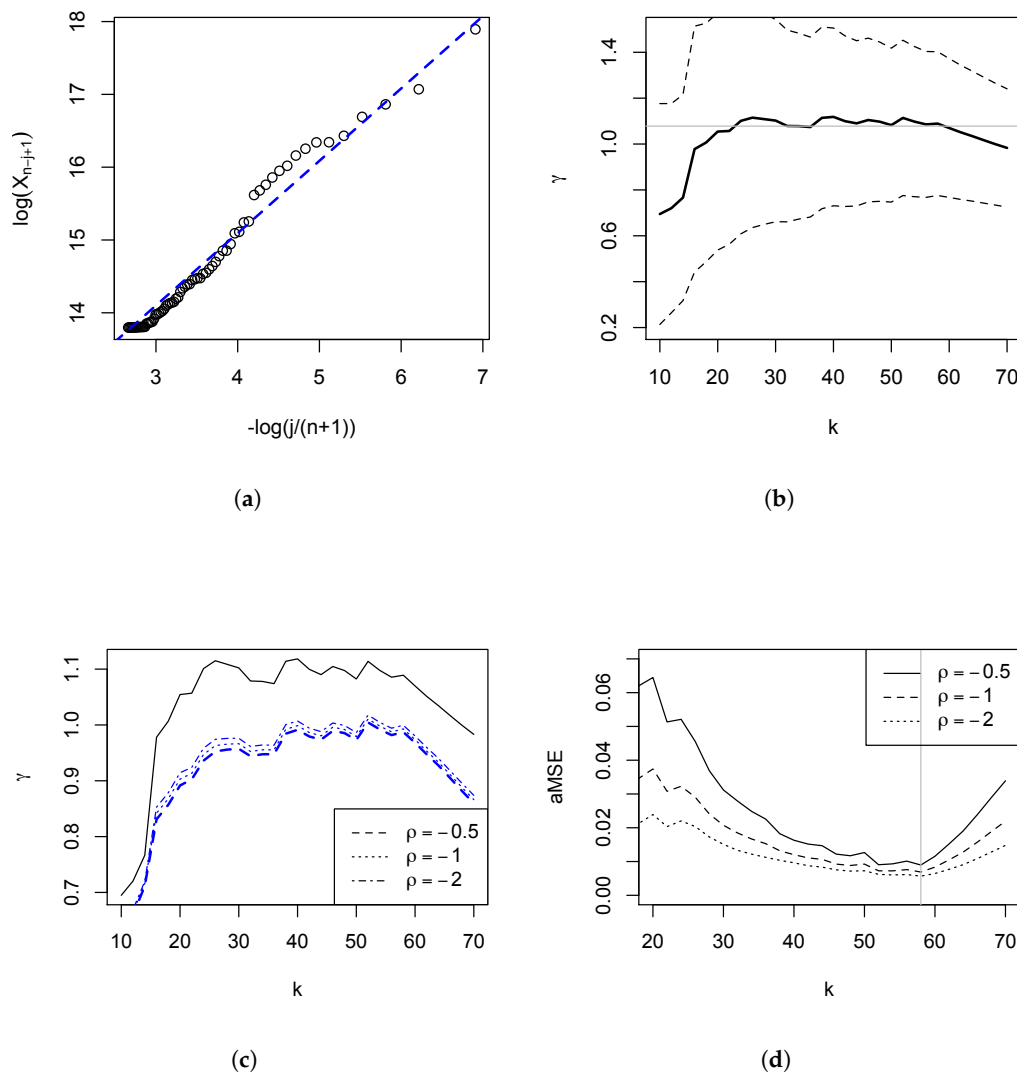
This conservative sensitivity analysis then consists of examining  $\hat{\gamma} - \tilde{b}_{k,n}(\rho)$  for a range of values of  $\rho$ .

Figure 5 panel (c) reports the results of such a sensitivity analysis for  $k$  being restricted to the  $n = 70$  highest incomes. Since under this restriction the Pareto QQ-plot is approximately linear, we expect that the remaining distortions are fairly modest. This is borne out in the sensitivity plot, as the precise value of  $\rho$  now plays only a minor role.

Should a researcher wish to choose a particular  $k$  by minimising an approximation to the AMSE, Equation (10) is the basis of the procedure proposed in Beirlant et al. (1996): Apply two weighting schemes  $w_{j,k}^{(i)}$  ( $i = 1, 2$ ), estimate the corresponding two mean weighted theoretical deviations using the residuals, and compute a linear combination thereof such that  $\text{Var}(\hat{\gamma}) + b_{k,n}^2$  obtains. We have carried out this programme (see Appendix A for further details) for weights  $w_{j,k}^{(1)} \equiv 1$  and  $w_{j,k}^{(2)} = j / (k + 1)$  for given  $\rho$ , and Figure 5 panel (d) depicts the results. Minimising this approximation to the AMSE yields  $k^*(\rho)$ , which, for  $\rho \in \{-2, -1, -0.5\}$ , resulted in  $k^* = 58$  across the selected  $\rho$ , for which  $\hat{\gamma}_{k^*} = 1.089$  obtains. In view of the results depicted in panel (c) it is not surprising that changing  $\rho$  has only a small

<sup>4</sup> Alternative estimators lead to similar conclusions. For instance, using the classic Hill estimator, at  $k = 60$  an estimate of  $\gamma$  of 1.017 is obtained. The plot (not displayed here) is fairly stable around this value for  $k \in [20, 60]$ .

effect. This estimate of  $\gamma$  is very close to the subjective visual choice of  $\hat{\gamma}$  of 1.075, reported above, based on Figure 5b.



**Figure 5.** Estimates of  $\gamma$ . Based on Survey of Personal IncomeS (SPI) data for 2009/10. Panel (a) Pareto QQ plot for the largest 70 incomes. The dashed line has slope 1. Panel (b): Estimates of  $\hat{\gamma}$  (solid line) and the 95% symmetric pointwise confidence interval (dashed line). The faint horizontal line at 1.075 is subjectively chosen. Panel (c): Sensitivity analysis. Plot of  $\hat{\gamma}$  (solid line) and  $\hat{\gamma} - \tilde{b}_{k,n}(\rho)$  for a different values of  $\rho$ . Panel (d): Approximation to the AMSE for different values of  $\rho$ . Minimising AMSE yields  $k^* = 58$  (vertical line) across the selected  $\rho$ , for which  $\hat{\gamma}_{k^*} = 1.089$  obtains.

## 5. Conclusions

The OLS estimator of the slope coefficient in the rank size regression (shifted or unshifted) can suffer significant higher order distortions that arise from the slow decay of the nuisance function  $l$  in the model  $1 - F(x) = x^{-\frac{1}{\gamma}}l(x)$  for  $\gamma > 0$ . Modeling the tail as second order regular variation, we have shown that the estimator over-estimates the true value in models in which  $l$  converges to a constant at a polynomial rate (i.e., in the leading heavy-tailed distributions). Our numerical illustrations have shown that these distortions can be dramatic, leading to test size distortions in which actual error rates are multiples of nominal error rates. The empirical illustration based on the Pareto QQ-plot has revealed a further distortion, namely the presence of a pronounced kink. Figure 1 has revealed that

using the common rule to choose 1% of the observation for tail estimation would lead to a severe under-estimation of how heavy the tail is.

The higher order distortions are functions of  $A(\cdot)$  and the second order regular variation parameter  $\rho$ . Since existing methods usually result in poor estimates of these, reliable bias corrections are not feasible. In view of this we have proposed a sensitivity analysis based on diagnostics from the rank size regression. When applied to our data on top incomes, we still cannot reject the hypothesis  $\gamma$  be unity, a situation often described in several fields as Zipf's law (e.g., Schluter and Trede 2017).

The simplicity of the regression estimator is undoubtedly the principal reason for its popularity among practitioners in economics. This paper has shown that in many situations the naive (i.e., 'blind') use of this estimator should be considered with care: Pareto QQ-plot, the sensitivity plot and the AMSE plot convey jointly important information about the behaviour of the estimator.

**Acknowledgments:** I thank the referees for their constructive comments that have helped to improve the paper.

**Conflicts of Interest:** The author declares no conflicts of interest.

## Appendix A. Proofs

Before proving the main result given by (7), we consider first the behaviour of the numerator  $N_{n,k}$  under first-order regular variation (5). We then refine the asymptotic expansion by assuming that the second-order regular variation (6) holds.

**First-order asymptotic expansion of the numerator  $N_{n,k}$ .** Assume that (5) holds, and consider an intermediate sequence  $k = k_n$  of positive integers such that  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . It will be shown that

$$\begin{aligned} N_{n,k}/\gamma &= 2 - \left(\frac{1}{2} - \eta\right) \frac{\log(k - \eta)}{k} - \left(\frac{1}{2} - \eta\right) \frac{\log^2 k}{2k} \\ &+ O_p\left(\frac{1}{k^{1/2}}\right) + O\left(\frac{1}{k}\right) + O_p\left(\frac{\log k}{k^{1/2}}\right) \end{aligned} \quad (\text{A1})$$

**Remark:** The term  $O_p\left(\frac{\log k}{k^{1/2}}\right)$  dominates  $(\log k)/k$ , and is not eliminated by setting the shift factor  $\eta$  to  $1/2$ .

In the proof of (A1) we will make use of the following Euler Maclaurin formulae (e.g., Gabaix and Ibragimov 2011, Equations A.4 and A.5)

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^k \log^2(i - \eta) &= 2 + \frac{k - \eta}{k} \log^2(k - \eta) - 2 \frac{k - \eta}{k} \log(k - \eta) + \frac{\log^2(k - \eta)}{2k} + O\left(\frac{1}{k}\right) \\ &= 2 + \log(k - \eta)(\log(k - \eta) - 2) + O\left(\frac{\log^2 k}{k}\right) \end{aligned} \quad (\text{A2})$$

and

$$\frac{1}{k} \sum_{i=1}^k \log(i - \eta) = -1 + \log(k - \eta) + \left(\frac{1}{2} - \eta\right) \frac{\log(k - \eta)}{k} + O\left(\frac{1}{k}\right) \quad (\text{A3})$$

**Proof of (A1).** We adapt the proofs of Kratz and Resnick (1996) (KR henceforth) of their Equations 2.4 and 2.8. The key is the use of Renyi's representation of exponential order statistics, which implies (e.g., KR, p. 705)

$$E_{n-k+i,n} - E_{n-k,n} =^d \sum_{j=k-i+1}^k \frac{E_j}{j}$$

where  $E_j$  ( $j = 1, \dots, n$ ) denote iid unit exponential random variables, and  $E_{n-k,n}$  denotes the  $(n - k)$ -th order statistic. We obtain an asymptotic refinement by using, instead of KR's Lemmas 2.2 and 2.3, the above Euler Maclaurin formulae, and Lyapunov's central limit theorem (CLT). Our numerator is denoted there by  $A_n$ , and the indices are mapped by setting  $i = k + 1 - j$ . From KR (pp. 704-707), we have

$$\begin{aligned} N_{n,k}/\gamma &=^d \frac{1}{k} \sum_{i=1}^k -\log\left(\frac{i-\eta}{k+1}\right) \sum_{j=i}^k \frac{E_j}{j} + o_p(1/k) \\ &= \log(k+1)\bar{E}_k - \frac{1}{k} \sum_{j=1}^k E_j \left[ \frac{1}{j} \sum_{i=1}^j \log(i-\eta) \right] + o_p(1/k) \end{aligned} \tag{A4}$$

where  $\bar{E}_k = (1/k) \sum_j^k E_j$ .

We first show that KR's result (A4) can also be derived from the first order regular variation condition under the stated assumptions. Let  $Y$  denote a standard Pareto random variable, and denote the  $(n - k)$ -th order statistic by  $Y_{n-k,n}$ . Consider the scaled log excesses

$$\frac{\log X_{n-i+1,n} - \log X_{n-k,n}}{a(Y_{n-k,n})/U(Y_{n-k,n})}$$

where  $a(\cdot)$  and  $U(\cdot)$  are defined in representation (5). Then, noting that  $X_{i,n} =^d U(Y_{i,n})$ , and using (5) with  $t = Y_{n-k,n}$  and  $x = Y_{n-i+1,n}/Y_{n-k,n}$ , the scaled log excesses satisfy as  $n \rightarrow \infty$  and  $n/k \rightarrow \infty$

$$\begin{aligned} \frac{\log X_{n-i+1,n} - \log X_{n-k,n}}{a(Y_{n-k,n})/U(Y_{n-k,n})} &=^d \frac{\log U(Y_{n-i+1,n}) - \log U(Y_{n-k,n})}{a(Y_{n-k,n})/U(Y_{n-k,n})} \\ &= \log\left(\frac{Y_{n-i+1,n}}{Y_{n-k,n}}\right) + o_p(1) \end{aligned}$$

By Renyi's representation of exponential order statistics, we have  $Y_{n-i+1,n}/Y_{n-k,n} =^d Y_{k-i+1,k}$ , so

$$\log\left(\frac{Y_{n-i+1,n}}{Y_{n-k,n}}\right) =^d \log(Y_{k-i+1,k}) =^d E_{k-i+1,k} =^d \sum_{i=j}^k \frac{E_i}{i}$$

since, using Renyi's representation again,  $E_{k-j+1,k} =^d E_{1,k} + \sum_{i=j}^{k-1} \frac{E_i}{i} = \sum_{i=j}^k \frac{E_i}{i}$ . From Wellner (1978), we know that  $\frac{k}{n} Y_{n-k,n} \rightarrow^p 1$ , so  $a(Y_{n-k,n})/U(Y_{n-k,n}) \rightarrow \gamma$ . Using the definition of  $N_{n,k}$ , on combining the results we thus obtain

$$N_{n,k}/\gamma =^d \left[ \frac{1}{k} \sum_{j=1}^k (-1) \log\left(\frac{j-\eta}{k+1}\right) \sum_{i=j}^k \frac{E_i}{i} \right] + o_p(1/k).$$

as claimed.

We proceed to examine (A4). Using (A3) yields

$$\begin{aligned} N_{n,k}/\gamma &= \log(k+1)\bar{E}_k + \bar{E}_k - \frac{1}{k} \sum_{j=1}^k E_j \log(j-\eta) \\ &\quad - \left(\frac{1}{2} - \eta\right) \frac{1}{k} \sum_{j=1}^k E_j \frac{\log(j-\eta)}{j} - \frac{1}{k} \sum_{j=1}^k E_j O\left(\frac{1}{j}\right) \end{aligned}$$

By Lyapunov’s CLT,

$$\frac{k^{1/2}}{\log k} \left[ \frac{1}{k} \sum_{j=1}^k (E_j - 1) \log(j - \eta) \right] \rightarrow^d N(0, 1)$$

so  $(1/k) \sum_{j=1}^k E_j \log(j - \eta) = O_p\left(\frac{\log k}{k^{1/2}}\right) + (1/k) \sum_{j=1}^k \log(j - \eta)$ . Using again (A3) and substituting the result, we obtain

$$\begin{aligned} N_{n,k}/\gamma &= \bar{E}_k + 1 + \log(k + 1)\bar{E}_k - \log(k - \eta) \\ &- \left(\frac{1}{2} - \eta\right) \frac{\log(k - \eta)}{k} + O\left(\frac{1}{k}\right) + O_p\left(\frac{\log k}{k^{1/2}}\right) \\ &- \left(\frac{1}{2} - \eta\right) \frac{1}{k} \sum_{j=1}^k \frac{\log(j - \eta)}{j} - \frac{1}{k} \sum_{j=1}^k O\left(\frac{1}{j}\right) \\ &- \left(\frac{1}{2} - \eta\right) \frac{1}{k} \sum_{j=1}^k (E_j - 1) \frac{\log(j - \eta)}{j} - \frac{1}{k} \sum_{j=1}^k (E_j - 1) O\left(\frac{1}{j}\right) \end{aligned}$$

Note that  $\bar{E}_k + 1 = 2 + O_p(k^{-1/2})$  and  $\log(k + 1)(\bar{E}_k - 1) = O_p\left(\frac{\log k}{k^{1/2}}\right)$ .  $(1/k) \sum_{j=1}^k \frac{\log(j - \eta)}{j}$  is a Riemann approximation to the integral  $k^{-1} \int_1^k (\log x - \eta)/x dx = (\log^2 k)/2k$ , and  $(1/k) \sum_{j=1}^k \frac{1}{j}$  is a Riemann approximation to the integral  $k^{-1} \int_1^k (1/x) dx = (\log k)/k$ . By Lyapunov’s CLT,  $(k/\sqrt{2})[(1/k) \sum_{j=1}^k (E_j - 1) \frac{\log(j - \eta)}{j}] \rightarrow^d N(0, 1)$ , so  $(1/k)(E_j - 1) \frac{\log(j - \eta)}{j} = O_p(1/k)$ . Similarly, the last term is  $O_p(1/k^2)$ . Hence

$$\begin{aligned} N_{n,k}/\gamma &= 2 - \left(\frac{1}{2} - \eta\right) \frac{\log(k - \eta)}{k} - \left(\frac{1}{2} - \eta\right) \frac{\log^2 k}{2k} \\ &+ O_p\left(\frac{1}{k^{1/2}}\right) + O\left(\frac{1}{k}\right) + O_p\left(\frac{\log k}{k^{1/2}}\right), \end{aligned}$$

which is Equation (A1), as claimed.

Before refining the asymptotic expansion, we briefly consider:

**Proof of (4).**  $D_k = 2 + O\left(\frac{\log^2 k}{k}\right)$ . Expanding the quadratic in the definition of  $D_k$

$$D_k = \log^2(k + 1) - 2(\log k + 1) \frac{1}{k} \sum_{i=1}^k \log(i - \eta) + \frac{1}{k} \sum_{i=1}^k \log^2(i - \eta)$$

and using the Euler Maclaurin formulae yields the stated result.

We are now in a position to examine the behaviour of the numerator  $N_{n,k}$  under second order regular variation.

**Proof of the higher order expansion (7).** Consider the scaled log excesses again, this time using representation (6) instead of (5). Set again  $t = Y_{n-k,n}$  and  $x = Y_{n-i+1,n}/Y_{n-k,n}$ , and recall  $Y_{n-i+1,n}/Y_{n-k,n} =^d Y_{k-i+1,k}$ . Hence we obtain the higher order expression

$$\begin{aligned} \frac{\log X_{n-i+1,n} - \log X_{n-k,n}}{a(Y_{n-k,n})/U(Y_{n-k,n})} &=^d \sum_{i=j}^k \frac{E_i}{i} \\ &+ A(Y_{n-k,n})H_{\gamma,\rho}(Y_{k-i+1,k}) + o_p(1). \end{aligned} \tag{A5}$$

The role of the first term on the right for  $N_{n,k}$  has already been described above. In what follows, we consider the higher order term. Since  $H_{\gamma>0,\rho<0}(x) = \frac{1}{\rho}(\frac{x^\rho-1}{\rho} - \log x)$ , the higher order expansion of  $N_{n,k}$  requires the analysis of

$$\frac{1}{k} \sum_{i=1}^k (-1) \log \left( \frac{i-\eta}{k+1} \right) \left[ \frac{Y_{k-i+1,k}^\rho - 1}{\rho} \right] = \log(k+1)\bar{Y}_\rho - \frac{1}{k} \sum_{i=1}^k (\log i - \eta) \left[ \frac{Y_{k-i+1,k}^\rho - 1}{\rho} \right]$$

where  $\bar{Y}_\rho = (1/k) \sum_{i=1}^k \frac{Y_i^\rho - 1}{\rho}$ .  $\bar{Y}_\rho$  has expectation  $(1-\rho)^{-1}$ , so by the CLT  $\bar{Y}_\rho = (1-\rho)^{-1} + O_p(k^{-1/2})$ . To handle the last sum, note that  $Y_{k-j+1,k} =^d \exp(E_{k-j+1,k}) =^d (V_{j,k})^{-1}$  where  $V$  denotes a standard uniform random variable, and we replace the order statistic  $V_{j,k}$  by its expectation,  $V_{j,k} = j/(k+1) + O_p(k^{-1/2})$ . A Taylor series expansion then gives  $(V_{j,k}^{-1})^\rho = (k+1/j)^\rho + O_p(k^{-1/2})$ . Then

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^k (-1) \log \left( \frac{i-\eta}{k+1} \right) \left[ \frac{Y_{k-i+1,k}^\rho - 1}{\rho} \right] &= \frac{\log(k+1)}{1-\rho} - \frac{1}{\rho} \frac{(k+1)^\rho}{k} \sum_{i=1}^k (\log i - \eta) j^{-\rho} \\ &+ \frac{1}{\rho} \frac{1}{k} \sum_{i=1}^k \log(i-\eta) + O_p \left( \frac{\log k}{k^{1/2}} \right) \end{aligned}$$

For the third term on the rhs, we use the Euler Maclaurin (A3), for the second term on the rhs we have the following Euler Maclaurin

$$\frac{1}{k} \sum_{j=1}^k j^{-\rho} \log(j-\eta) = \frac{1}{1-\rho} k^{-\rho} \log(k-\eta) - \left( \frac{1}{1-\rho} \right)^2 k^{-\rho} + o(k^{-\rho})$$

Combing these two Euler Maclaurin formulae, we can simplify to get

$$\begin{aligned} \frac{1}{\rho} \frac{(k+1)^\rho}{k} \sum_{i=1}^k (\log i - \eta) j^{-\rho} - \frac{1}{\rho} \frac{1}{k} \sum_{i=1}^k \log(i-\eta) &= \frac{\log k - \eta}{1-\rho} - \frac{1}{\rho} \frac{1}{(1-\rho)^2} + \frac{1}{\rho} + O \left( \frac{\log k}{k} \right) \\ &= \frac{\log k - \eta}{1-\rho} - \frac{2-\rho}{(1-\rho)^2} + O \left( \frac{\log k}{k} \right) \end{aligned}$$

Therefore<sup>5</sup>

$$\frac{1}{k} \sum_{i=1}^k (-1) \log \left( \frac{i-\eta}{k+1} \right) \left[ \frac{Y_{k-i+1,k}^\rho - 1}{\rho} \right] = \frac{2-\rho}{(1-\rho)^2} + O_p \left( \frac{\log k}{k} \right) \tag{A6}$$

We are now in a position to combine the results. In order to simplify notation, denote the first order expansion of the numerator  $N_{n,k}/\gamma$  by  $N_{1,n,k}/\gamma$ , given by the rhs of (A1). Then substituting the higher order expression for the scaled excesses (A5) into the formula for  $N_{n,k}$ , recalling that  $\frac{k}{n} Y_{n-k,n} \rightarrow^p 1$  (Wellner 1978), and using (A6) yields

$$N_{n,k}/\gamma = N_{1,n,k}/\gamma + A \left( \frac{n}{k} \right) \frac{1}{\rho} \left[ \frac{2-\rho}{(1-\rho)^2} \right] + O_p \left( \frac{\log k}{k} \right) + o_p(A(n/k)).$$

**Proof of (8).** The class of kernel estimator considered in Csorgo et al. (1985) is of the form

$$\hat{\gamma}_{kernel} = \frac{\sum_{j=1}^k (j/k) K(j/k) [\log X_{n-j+1} - \log X_{n-j}]}{\int_0^1 K(t) dt}$$

<sup>5</sup> To support this key expression, numerical evidence from a Monte Carlo with  $k = 1000$ , 1000 samples, and  $\eta = 0$  yielded for the lhs of (15) v.  $(2-\rho)/(1-\rho)^2$  the following:  $\rho = -1/2$ : 1.105 v. 1.111,  $\rho = -1$ : .746 v. 0.75,  $\rho = -2$ : 0.443 v. 0.444.

Their Theorem 2 (or Theorem 1.1 in [Beirlant et al. 1996](#)) states the following. Under general conditions on kernel  $K$  and distribution function  $F$  so that there exists a nonrandom sequence  $C_n$  such that  $C_n(\hat{\alpha} - \gamma)$  converges weakly to a limiting  $N(0, 1)$  distribution for some sequence  $k = k_n \rightarrow \infty$  with  $k_n/n \rightarrow 0$ , it is necessary and sufficient that

$$\lim_{n \rightarrow \infty} \sqrt{k} \int_0^1 b(kw/n)K(w)dw = 0$$

where  $b$  is a function such that  $b(1/x) \rightarrow 0$  as  $x \rightarrow \infty$ , and that for the tail quantile function  $U(x) = x^\gamma \tilde{I}(x)$  with

$$\tilde{I}(x) = c(x) \exp \left( \int_{1/x}^1 \frac{b(u)}{u} du \right)$$

where  $c(x) \rightarrow c$  as  $x \rightarrow \infty$ . If this condition is satisfied, then as  $k = k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$

$$\sqrt{k} \left( \int_0^1 K^2(v)dv \right)^{-1/2} (\hat{\gamma}_{kernel} - \gamma) \rightarrow^d N(0, \gamma^2).$$

[Beirlant et al. \(1996\)](#) observe that our slope estimator  $\hat{\gamma}$ , given by (3), is (to first order) a member of the class of kernel estimators  $\hat{\gamma}_{kernel}$  with kernel  $K(t) = 1 - \log t$ , and that the above condition holds under the regular variation hypothesis. Turning to the specific kernel  $K(t) = 1 - \log t$ , since  $\int_0^1 K(t)dt = 2$  and not unity, a scale correction is required. As  $\int_0^1 K^2(t)dt = 5$ , the stated result (8) follows.

**Proof of (10).** Consider the mean weighted theoretical squared deviation

$$\frac{1}{k} \sum_{j=1}^k w_{j,k} E \left( \log \left( \frac{X_{n-j+1,n}}{X_{n-k,n}} \right) - \gamma \log \left( \frac{k+1}{j} \right) \right)^2$$

for some weights  $w_{j,n}$ . Using (A5) this equals, to first order,

$$\frac{\gamma^2}{k} \sum_{j=1}^k E \left( \left( \sum_{i=j}^k \frac{E_i}{i} - \log \left( \frac{k+1}{j} \right) \right) + A(Y_{n-k,n})H_{\gamma,\rho}(Y_{k-i+1,k}) \right)^2$$

Then, recalling that  $Y_{k-j+1,k} \stackrel{d}{=} (V_{j,k})^{-1}$  and proceeding as in [Beirlant et al. \(1996, Section 4\)](#), which involves approximating expectations  $E(f(V_{j,k}))$  by the leading term  $f(j/(k+1))$  when applying the delta method yields, to first order,

$$\frac{\gamma^2}{k} \tilde{c}_k + d_k(\rho) b_{k,n}^2$$

with

$$\tilde{c}_k = \sum_{j=1}^k w_{j,k} \left( \sum_{l=1}^{k-j+1} \left( \frac{1}{k-l+1} \right)^2 + \left( \sum_{l=1}^{k-j+1} \frac{1}{k-l+1} - \log \left( \frac{k+1}{j} \right) \right)^2 \right)$$

and  $d_k(\rho) = \left( \frac{1}{2} \frac{2-\rho}{(1-\rho)^2} \right)^{-2} \tilde{d}_k(\rho)$  with

$$\tilde{d}_k(\rho) = \frac{1}{k} \sum_{j=1}^k w_{j,k} \left( \frac{(j/(k+1))^{-\rho} - 1}{\rho} \right)^2$$

Finally, we set  $c_k = (4/5)\tilde{c}_k$  and  $w_{j,k} \equiv 1$  to arrive at (10).

**Remark:** In order to obtain an estimate of the AMSE, [Beirlant et al. \(1996\)](#) use two weighting schemes,



namely  $w_{j,k} \equiv 1$  leading to coefficients, say,  $c_k^1$  and  $d_k^1$  and mean weighted squared residuals  $k^{-1}SSR_k^1$ , and  $w_{j,k} = j/(k+1)$  leading to  $c_k^2$ ,  $d_k^2$ , and  $k^{-1}SSR_k^2$ . Then a linear combination of two approximate MSE expressions (with coefficients, say,  $x$  and  $y$ ) is sought that yields  $Var(\hat{\gamma}) + b_{k,n}^2$ , which is achieved by solving simultaneously the equations

$$\begin{aligned} xc_k^1 + yc_k^2 &= 1 \\ xd_k^1 + yd_k^2 &= 1. \end{aligned}$$

## References

- Atkinson, Anthony Barnes. 2017. Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica* 84: 129–56.
- Beirlant, Jan, Petra Vynckier, and Jozef L. Teugels. 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* 9: 1659–67.
- Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef L. Teugels. 2004. *Statistics of Extremes*. Wiley Series in Probability and Statistics. Chichester: Wiley.
- Burkhauser, Richard V., Shuaiz hang Feng, Stephen Jenkins, and Jeff Larrimore. 2012. Recent trends in top income shares in the USA: Reconciling estimates from March CPS and IRS tax return data. *Review of Economics and Statistics* 94: 371–88.
- Cowell, Frank A. 1989. Sampling variances and decomposable inequality measures. *Journal of Econometrics* 42: 27–41.
- Cowell, Frank A., and Emmanuel Flachaire. 2007. Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics* 141: 1044–72.
- Csorgo, Sandor, Paul Deheuvels, and David Mason. 1985. Kernel Estimates of the Tail Index of a Distribution. *The Annals of Statistics* 13: 1050–77.
- Davidson, Russell, and Emmanuel Flachaire. 2007. Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics* 141: 141–66.
- De Haan, Laurens, and Ana Ferreira. 2006. *Extreme Value Theory*. New York: Springer.
- De Haan, Laurens, and Ulrich Stadtmüller. 1996. Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)* 61: 381–95.
- Dekkers, Arnold L. M., John H. J. Einmahl, and Laurens de Haan. 1989. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17: 1833–55.
- Ebrechts, Paul, Claudia Kluppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events*. Berlin: Springer.
- Gabaix, Xavier, and Rustam Ibragimov. 2011. Rank - 1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business and Economic Statistics* 29: 24–39.
- Hall, Peter. 1982. On some simple estimate of an exponent of regular variation. *Journal of the Royal Statistical Society Ser. B* 44: 37–42.
- Kratz, Marie, and Sidney I. Resnick. 1996. The QQ-estimator and heavy tails. *Communications in Statistics. Stochastic Models* 12: 699–724.
- Jenkins, Stephen P. 2017. Pareto models, top incomes and recent trends in UK income inequality. *Economica* 84: 261–89.
- Schluter, Christian, and Mark Trede. 2002. Tails of Lorenz curves. *Journal of Econometrics* 109: 151–66.
- Schluter, Christian, and Mark Trede. 2017. Size distributions reconsidered. *Econometric Reviews*, forthcoming.
- Schultze, J., and J. Steinebach. 1996. On least squares estimates of an exponential tail coefficient. *Statistics and Decisions* 14: 353–72.
- Wellner, Jon A. 1978. Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Zeitschrift fuer Wahrscheinlichkeitstheorie und verwandte Gebiete* 45: 73–88.

