Estimation and Inference about Tail Features with Tail Censored Data*

Yulong Wang[†] Zhijie Xiao[‡]

Syracuse University Boston College

First version: May 2019

This version: March 2020

Abstract

This paper considers estimation and inference about tail features when the observations beyond some threshold are censored. We first show that ignoring such tail censoring could lead to substantial bias and size distortion, even if the censored probability is tiny. Second, we propose a new maximum likelihood estimator (MLE) based on the Pareto tail approximation and derive its asymptotic properties. Third, we provide a small sample modification to the MLE by resorting to Extreme Value theory. The MLE with this modification delivers excellent small sample performance, as shown by Monte Carlo simulations. We illustrate its empirical relevance by estimating (i) the tail index and the extreme quantiles of the US individual earnings with the Current Population Survey dataset and (ii) the tail index of the distribution of macroeconomic disasters and the coefficient of risk aversion using the dataset collected by Barro and Ursúa (2008). Our new empirical findings are substantially different from the existing literature.

Keywords: Extreme Value theory, power law, extreme quantile, tail index

*We thank Ulrich Müller for very inspiring discussions. We also thank Tim Christensen, Bo Honoré, Jia Li, Andrew Patton, Alexis Toda, Mark Watson, and participants at numerous seminar/conference presentations for very helpful comments.

[†]*Address*: Department of Economics, Syracuse University, 127 Eggers Hall, Syracuse, NY 13244. *E-mail*: `ywang402@maxwell.syr.edu`

[‡]*Address*: Department of Economics, Boston College, 380 Maloney Hall, Chestnut Hill, MA 02467. *E-mail*: `zhijie.xiao@bc.edu`

1 Introduction

Tail risk and extreme events are important research topics in economics and finance. In many applications, the features of interest are tail properties such as tail index and extreme quantiles. Existing literature has extensively studied the case with fully observed datasets. In comparison, this article explores the case with tail censoring. We argue that it is important to take into account the censoring if the research interest is in the tail, even when the censoring fraction is small. We provide a new method to construct estimators and confidence intervals for tail features.

Suppose one has a random sample from some underlying distribution F, where the observations larger than some threshold T are replaced with T or simply unobserved. In principle, tail features cannot be even identified if they entirely depend on the right tail part of F that is beyond T. However, we can back out the tail-related features by extrapolation under two assumptions. They are that (i) the tail of F can be well approximated by some suitably chosen parametric distribution, and (ii) T is sufficiently large so that only a small fraction of samples are censored. The first assumption has been thoroughly studied in the statistic literature and is satisfied by many commonly used distributions. The second assumption is also satisfied in many interesting empirical applications, which motivate this article.

Our first motivating example is the Current Population Survey (CPS) dataset, which has been the primary data source used for investigating the distributions of individual earnings and household income in the US. Featured studies using CPS data include Armour, Burkhauser, and Larrimore (2013) and Eika, Mogstad, and Zafar (2019), among many others. In CPS, the individual earnings larger than some threshold T are typically censored (also called topcoded) and replaced with T for confidential reasons.[1] In 2019, the censoring threshold is 310000 USD, leading to an approximately 0.58% censoring fraction in the full sample of individuals between 18 and 70 years old. This quantity is also substantially different across the subsamples defined by race and gender but remains small, as seen in Table 1. Using this dataset, we aim to estimate and construct confidence intervals for the tail index that measures the tail heaviness of the income distribution and the extreme quantiles.

[1] The topcoding has constantly been changing. Description of the topcoding mechanism is available at https://cps.ipums.org/cps/topcodes_tables.

Table 1: Fractions and Numbers of Censored Observations in the 2019 CPS Dataset

	n	cen%	cen#	n	cen%	cen#
full sample	115424	0.582	672			
race\gender		male			female	
all	55553	0.884	491	59871	0.302	181
white	43371	0.966	419	45424	0.310	141
Asian	3676	1.360	50	4099	0.537	22
Hispanic	44420	1.002	445	48192	0.322	155
black	6144	0.195	12	7827	0.115	9

Note: Entries are the sample sizes (n), the fractions in percentage (cen%) and the numbers (cen#) of censored observations in individual earnings from the March CPS variable ERN_VAL. Data are available at https://usa.ipums.org/usa/.

In our second application, we examine the size distribution of macroeconomic disasters, which is investigated by Barro and Ursúa (2008) and Barro and Jin (2011). In particular, Barro and Jin (2011) define a macroeconomic disaster if the annual Gross Domestic Product (GDP) (or consumption) declines by more than 10%. The authors collect the data in 36 countries from 1870 to 2005 and construct a sample of approximately 5000 observations. However, data are missing for four countries during WWII due to government collapse or fighting wars. These observations correspond to the end-of-world case, and hence Barro and Jin (2011) concern that they are the largest observations but censored. In this situation, the tail censoring fraction is about 0.1%, and the parameter of interest is the tail index and the coefficient of risk aversion. Note that the censoring threshold T is unknown here.

One would think that a tiny censoring fraction makes nearly no effect if we ignore it. This is true if the object of interest lies in the mid-sample, such as the median. However, such ignorance could lead to substantial bias and size distortion if some tail features are of interest. To examine this, we conduct an extensive Monte Carlo study and find that even the 0.1% tail censoring could lead to poor finite sample performance in some commonly used methods, including, for example, the classical Hill (1975)'s estimator.

To accommodate the censoring, existing studies typically rely on some parametric assumption of the whole distribution (e.g., Aban, Meerschaert, and Panorska (2006), Jenkins, Burkhauser, Feng, and Larrimore (2010), and Burkhauser, Feng, Jenkins, and Larrimore (2010)). Then, tail features can be expressed as functions of the unknown parameters and estimated by the maximum likelihood estimator (MLE). However, the parametric assumption on the whole distribution may lead to a substantial misspecification error when the object of interest is in the tail. This is because tail features such as very large quantiles are typically on a large scale, and hence small misspecification can be considerably amplified. For example, the standard normal distribution and the Student-t distribution with 20 degrees of freedom share almost the same shape in the mid-sample but exhibit substantially different extreme quantiles. Such misspecification is documented by Brzezinski (2013) in a large-scale simulation study.

Instead of modeling the whole underlying distribution F, we focus on the *tail* part only and approximate it with the generalized Pareto distribution (GPD). Such a Pareto-tail approximation holds for many commonly used distributions, including, for example, Student-t, F, Beta, and Gaussian distributions. See Chapter 1 of de Haan and Ferreira (2007) for an overview. Given this approximation, we first pick some tail cutoff u, such as some large empirical quantile, and treat the observations larger than u (but still less than T) as stemming from the censored Pareto tail. Let k denote the number of these effective *tail* observations. Then, we can fit them into the classical Tobit model and conduct the MLE of the unknown parameters. Under some mild regularity conditions, we show that the MLE is consistent and asymptotically normal as k diverges, enabling the construction of confidence intervals. Then we can estimate extreme quantiles by expressing them as functions of the Pareto parameters. This is formally studied in Section 2.

The proposed maximum likelihood method can be quite good in finite samples for some applications, but it is also easy to find examples where the asymptotic distributions provide poor approximations. A fundamental limitation of our MLE and many other existing approaches in studying tail features is that they require restrictive conditions for the choice of k (and equivalently u). See, for example, Smith (1987), Chapters 3 and 4 in de Haan and Ferreira (2007), Beirlant, Alves, and Gomes (2016), and Beirlant, Alves, and Reynkens (2017). On the one hand, k has to be sufficiently large to support enough observations stemming from the approximately Pareto tail for the consistency and the asymptotic Gaussianity. On the other hand, k has to be sufficiently small relative to the whole sample size n so that the tail Pareto approximation incurs a negligible bias. Such a delicate balance is technically reflected in the conditions that $k \to \infty$ and $k/n \to 0$. As such, for some combinations of n and F, it is hard to find the k that leads to satisfactory inference. This situation is close in spirit to the bias-variance trade-off in choosing the bandwidth in the standard kernel regressions.

To alleviate the above issue, we propose a small sample modification of the MLE when k is only moderate, say 100. In particular, we follow Müller and Wang (2017) to consider k as a fixed number and study the fixed-k asymptotic embedding by resorting to Extreme Value (EV) theory. Instead of treating the tail observations as *independent* copies from the GPD, EV theory treats them as *dependent* random variables with a joint EV distribution. Such dependence is negligible when k is large but plays an important role when k is only moderate. Using the EV approximation, we propose new confidence intervals for the tail index and extreme quantiles, which have excellent coverage probabilities. This is studied in detail in Section 3.

In summary, the method that we propose is a hybrid approach. We suggest using the MLE for estimation and inference when k (and n) is large enough and switching to the fixed-k intervals otherwise. We choose the switching cutoff to be $k \gtrless 250$ based on our Monte Carlo experiments in Section 4.

Returning to the CPS application, n is large enough in the full sample to support a large k, and hence the MLE is expected to perform well. However, in the Asian male subsample, n is only 3676. Then choosing k as a small fraction, say 5%, of n, leads to only 180 tail observations and triggers the switching. By using the new approach, we make several interesting empirical findings. First, the tail index is substantially different across genders and races, while the existing literature commonly focuses on the full sample and finds the tail index to be approximately 0.5. See Toda and Wang (2019) and references therein. Second, extreme quantiles also considerably vary across genders and races. In particular, the 99.9% quantile of all males can be twice larger than that of the black male group. Third, the tail features are also substantially different across ages. The middle-aged groups exhibit heavier tails and larger extreme quantiles than the groups below 30 or above 60 years old.

In the macroeconomic disaster application, we use the new method to construct confidence intervals for the tail index and the coefficient of risk aversion. We find substantially different results from those in Barro and Jin (2011). In particular, we obtain a significantly heavier tail in the disaster distribution. This further results in a smaller coefficient of relative risk aversion, approximately 0.75 instead of 3. Our Monte Carlo simulation statistically justifies such a vast difference.

The rest of the paper is organized as follows. Section 2 develops the MLE, and Section 3 provides the small sample modification. Section 4 reports Monte Carlo simulations, and Section 5 applies the new approach to the CPS and the macroeconomic disaster examples. Section 6 concludes with some remarks. All proofs and computational details are collected in the Appendix.

2 The Maximum Likelihood Estimator

Consider a random sample $\{Y_i\}_{i=1}^n$ generated from some cumulative distribution function (CDF) F. Due to censoring, the econometrician observes the pair $(Y_i^0, D_i)^\intercal$ such that

$$\begin{aligned} Y_i^0 &= D_i T + (1 - D_i) Y_i \\ D_i &= \mathbf{1}[Y_i > T], \end{aligned} \tag{1}$$

where T denotes some constant censoring threshold and $\mathbf{1}[\cdot]$ the indicator function. Without loss of generality, we focus on the right tail. Define $m = \sum_{i=1}^n D_i$ as the number of censored observations. We assume the density of Y_i, denoted as $f(\cdot)$, is continuous and positive so that $\mathbb{P}(Y_i = T) = 0$.

The model (1) has spawned a vast literature about estimation and inference about the mid-sample features, such as median, non-extreme quantiles, and regression coefficients. See, for example, Powell (1986), Portnoy (2003), and Hong and Tamer (2003). These mid-sample features are typically estimated at the root-n rate. In contrast, the tail features are estimated at a much slower rate since only the largest observations are informative about the right tail.

Define

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

as the conditional CDF given that Y_i is larger than some pre-specified tail cutoff u. We aim to approximate $F_u(y)$ by the generalized Pareto distribution, which is given by

$$G(y; \xi, \sigma) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-y/\sigma) & \xi = 0 \end{cases} \tag{2}$$

with $y \in \mathbb{R}^+$ if $\xi \geq 0$ and $y \in (0, -\sigma/\xi)$ otherwise. Denote y_0 as the right end-point of the support of Y_i. It is well established in the statistic literature (e.g., Balkema and de Haan (1974) and Pickands (1975)) that the GPD is a good approximation of F in the tail, in the sense that

$$\lim_{u \to y_0} \sup_{0 < y < y_0 - u} |F_u(y) - G(y; \xi, \sigma)| = 0 \tag{3}$$

for some scale σ implicitly depending on u, if and only if F is in the domain of attraction of one of the three limit laws. The parameter ξ is referred to as the tail index, which is uniquely determined by F and characterizes its tail heaviness. See Chapter 1 of de Haan and Ferreira (2007) for an overview.

The tail approximation (3) is a mild assumption as it is satisfied by many commonly used distributions. In particular, the positive ξ case covers distributions with a Pareto-type tail such as Pareto, Student-t, and F distributions.[2] The case with $\xi = 0$ covers the distributions with finite moments of any order. Leading examples are normal and log-normal distributions. The situation with a negative ξ includes the distributions with a finite right end-point. For expositional simplicity, we focus our discussion on the case with $\xi > 0$, which covers the empirical applications with heavy tails.

In practice, we usually choose u as some large order statistic of Y_i, say the 95% empirical quantile. We let $T = T_n$ and $u = u_n$ depend on the sample size n and assume $T_n > u_n$ (otherwise there is no observation). Also, denote k as the number of the observations between u_n and T_n and $\{Y_{(1)} \geq Y_{(2)} \geq, \ldots, \geq Y_{(n)}\}$ the order statistics[3] by descending sorting. Then effectively the available observations are the censored largest $m + k$ order statistics

$$\left(\underbrace{T_n, \ldots, T_n}_{m}, Y_{(m+1)}, \ldots, Y_{(m+k)}\right)^\intercal, \tag{4}$$

where the largest m order statistics, $\{Y_{(1)}, \ldots, Y_{(m)}\}$ are censored. Using (3), we can write the conditional log-likelihood of the tail observations as

$$\begin{aligned}
\mathcal{L}_n(\xi, \sigma) &= \sum_{i=1}^{m+k} \left\{ D_i \log\left(1 - F_u\left(T_n - u_n\right)\right) + (1 - D_i) \log \frac{f\left(Y_{(i)} - u_n\right)}{1 - F\left(u_n\right)} \right\} \\
&\approx \sum_{i=1}^{m+k} \left\{ D_i \log\left(1 - G\left(T_n - u_n\right)\right) + (1 - D_i) \log g\left(Y_{(i)} - u_n; \xi, \sigma\right) \right\} \\
&= \sum_{i=1}^{m+k} \left\{ -\frac{D_i}{\xi} \log\left(1 + \frac{\xi\left(T_n - u_n\right)}{\sigma}\right) - (1 - D_i) \log \sigma \right. \\
&\qquad \left. - (1 - D_i)\left(1 + \frac{1}{\xi}\right) \log\left(1 + \frac{\xi\left(Y_{(i)} - u_n\right)}{\sigma}\right) \right\},
\end{aligned}$$

[2] In the standard Pareto distribution with the CDF $\mathbb{P}(Y > y) \propto y^{-\alpha}$, the tail index ξ equals $1/\alpha$. We focus on ξ instead of α for notational ease.

[3] This is different from the conventional notation for order statistics, that is, $\{Y_{n:n} \geq Y_{n:n-1} \geq, \ldots, \geq Y_{n:1}\}$. We think this alternative is more intuitive in our setup, especially in Section 3.

where $g(y;\xi,\sigma) = \partial G(y;\xi,\sigma)/\partial y$. Then the MLE of ξ and σ are constructed as

$$\left(\hat{\xi},\hat{\sigma}\right)^{\intercal} = \arg\max_{(0,\infty)^2} \mathcal{L}_n(\xi,\sigma). \tag{5}$$

To derive the asymptotic properties of the MLE, we make the following assumptions. To simplify notations, we write $\alpha = 1/\xi$ when convenient and define $L(y) = y^{\alpha}(1-F(y))$.

Condition 1 *$(Y_i^0, D_i)^{\intercal}$ is independently and identically generated from (1).*

Condition 2 *$F(\cdot)$ is continuously differentiable with $0 < f(\cdot) < \bar{f}$ for some constant $\bar{f} < \infty$ and satisfies $L(y) = C(1+\delta y^{-\beta} + o\left(y^{-\beta}\right))$ for some constants $\beta > 0$, $C \neq 0$ and $\delta \in \mathbb{R}$.*

Condition 3 *$T_n \to \infty$ and $T_n/u_n \to \kappa \in (1,\infty)$.*

Condition 4 *$k \to \infty$ and $k = o\left(n^{2\beta/(\alpha+2\beta)}\right)$.*

Condition 1 assumes a random sample generated with censoring. Condition 2 is imposed by Hall (1982), which states that the Pareto tail approximation involves a second-order bias of the order $y^{-\beta}$. This is imposed to avoid technical complexity and can be relaxed with other weaker conditions (cf. Goldie and Smith (1987)). Consider the Student-t distribution with zero mean, unit variance, and v degrees of freedom, for example. The CDF is given by

$$1 - F_{t(v)}(y) = Cy^{-v}(1+\delta y^{-2} + O(y^{-4})) \text{ as } y \to \infty.$$

Then Condition 2 holds with $\xi = 1/v$ and $\beta = 2$.

Condition 3 assumes the censoring threshold is larger than the tail cutoff. Specifically, the censoring is asymptotically negligible if $\kappa = \infty$, and leads to no tail observation if $\kappa = 1$. Condition 4 specifies the choice of the tail cutoff and equivalently the number of tail observations k. This asymptotic framework has been commonly used in the literature about extreme value theory. In particular, our Condition 4 corresponds to the condition on r in Theorem 1 in Hall (1982) and Condition (3.2) in Smith (1987). In addition, different versions of this condition are extensively studied and used in Chapters 3 and 4 in de Haan and Ferreira (2007). The very last assumption that $k = o\left(n^{2\beta/(\alpha+2\beta)}\right)$ satisfies $k/n \to 0$ and is imposed for expositional simplicity. We can relax it into $kn^{-2\beta/(\alpha+2\beta)} \to \mu$ for some $\mu \in \mathbb{R}$. Doing so leads to a non-zero mean in the asymptotic normal distribution, which further depends on μ, β, and δ. These nuisance parameters are hardly estimable in practice, and hence researchers typically choose a sufficiently small k to retain the asymptotic zero mean. This is similar in spirit to the undersmoothing in standard kernel regressions.

In practice, it makes no difference, at least asymptotically, between treating T_n as constant and the number of censored observations m as stochastic and the opposite treatment. To see this, our Condition 2 is sufficient for the SR2 condition in Smith (1987) with his $\phi(u_n) = u_n^{-\beta}$. Our Condition 3 assumes that u_n and T_n are of the same order of magnitude as $n \to \infty$, and hence m/k still converges to some positive constant in probability. Then it suffices to consider the no censoring case and establish the asymptotic equivalence between thinking u_n or k as constant and the other is random. The argument is given by Smith (1987), pp.1180-1181.

Under Conditions 1-4, the following proposition establishes the asymptotic normality of the MLE.

Proposition 1 *Suppose Conditions 1-4 hold. Then*

$$k^{1/2}\begin{pmatrix} \hat{\xi} - \xi \\ \frac{\hat{\sigma}}{\sigma} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, M^{-1}\right),$$

where the elements of M are given by

$$\begin{aligned} M_{11} &= \frac{2}{(1+\xi)(1+2\xi)} + \frac{\kappa^{-2-1/\xi}}{(1+\xi)(1+2\xi)\xi^2} \times \\ &\quad \left\{-1-\xi+\kappa(2+4\xi)-\kappa^2(1+\xi)(1+2\xi)\right\} \\ M_{22} &= \frac{1}{(1+2\xi)} - \frac{\kappa^{-2-1/\xi}}{(1+2\xi)} \\ M_{12} &= \frac{1}{(1+\xi)(1+2\xi)} + \frac{\kappa^{-2-1/\xi}}{(1+\xi)(1+2\xi)\xi^2} \times \\ &\quad \left\{-(1+\xi)^2 + (1-2\kappa)(1+\xi)(1+2\xi) + \kappa(2+\xi)(1+2\xi)\right\}. \end{aligned}$$

The tail censoring complicates the asymptotic variance substantially as compared with the no censoring case (cf. Smith (1987)). In particular, when the censoring is asymptotically negligible ($\kappa = \infty$), the information matrix reduces to

$$M = \begin{bmatrix} \frac{2}{(1+\xi)(1+2\xi)} & \frac{1}{(1+\xi)(1+2\xi)} \\ \frac{1}{(1+\xi)(1+2\xi)} & \frac{1}{(1+2\xi)} \end{bmatrix}.$$

Given the MLE of ξ and σ, we can further estimate the extreme quantile $Q(1-p) \equiv$

$\inf\{y : 1-p \leq F(y)\}$. We set $p = p_n \to 0$ to capture the extremeness. The estimator can be constructed by inverting (2), that is,

$$\hat{Q}(1-p_n) = u_n + \frac{\hat{\sigma}}{\hat{\xi}}\left(d_n^{\hat{\xi}} - 1\right),$$

where $d_n = (m+k)/(p_n n)$. To derive a non-trivial asymptotic result, we let $d_0 \equiv \lim_{n\to\infty} d_n > 0$ so that the target quantile is of the same or larger magnitude of u_n (otherwise it can be estimated by the corresponding empirical quantile). The following proposition derives the asymptotic distribution of $\hat{Q}(1-p_n)$.

Proposition 2 *Suppose Conditions 1-4 hold. If $d_0 > 0$, then*

$$k^{1/2}\frac{\hat{Q}(1-p_n) - Q(1-p_n)}{\sigma q_\xi(d_n)} \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where $q_\xi(t) = \xi^{-1}t^\xi \log t$ and

$$\Sigma = \left(1, \frac{d_0^\xi - 1}{\xi q_\xi(d_0)}\right) M^{-1} \left(1, \frac{d_0^\xi - 1}{\xi q_\xi(d_0)}\right)^\intercal + q_\xi(d_0)^{-2}.$$

Proposition 2 establishes the asymptotic normality of the extreme quantile estimator. Then the confidence intervals for ξ and $Q(1-p_n)$ can be constructed by plugging in the estimators for the asymptotic variance.

3 Small Sample Modification under the Fixed-k Asymptotics

The results in Section 2 suggest that the asymptotic normal approximation can be used for inference about the tail features as k goes to ∞. In practice, however, the choice of the tail sample size k is widely accepted as a challenging question even without censoring. This is because a good selection of k has to balance the tail approximation bias and the variance delicately. Ultimately, the underlying distribution has to be reasonably close to the Pareto distribution in the tail to guarantee a satisfactory finite sample performance.

The asymptotic approximation can be quite accurate for some cases, but it is also easy to find examples where the limiting normal distribution provides a poor approximation. Consider the example that F is a mixture of the standard normal distribution with probability 0.8 and some Pareto distribution with probability 0.2. Such a mixture structure implies that only the very few largest observations are informative about the true tail. In this case, choosing a large k means including too many contaminating observations from the mid-sample, while choosing a small k invalidates the asymptotic Gaussianity. In principle, there is no such a procedure that consistently justifies whether a given k is appropriate when F is entirely unknown. See Theorem 5.1 in Müller and Wang (2017) for a discussion on the non-censored case.

Therefore, in this article, we do not focus on the choice of k but instead treat it as given. In some cases, k is determined by some economic theory or empirical guidance. For example, in the macroeconomic disaster application, the economic definition of disasters for more than 10% of GDP decline yields the choice of k. In other cases, we may employ some data-driven algorithms that balance the Pareto approximation bias and the variance. See, for example, Hall (1982), Drees (2001), and Clauset, Shalizi, and Newman (2009).

When k and n/k are both sufficiently large, we would expect the MLE in (5) based on the increasing-k asymptotics to work well. Nevertheless, k is only moderate in some situations, including our macroeconomic disaster application and the Asian male subsample in CPS. This causes a small sample issue that the asymptotic Gaussianity is questionable. To find a better alternative, we resort to the asymptotic embedding that requires n diverges, but k remains a fixed constant. Under this fixed-k asymptotic framework, the consistent estimation of the tail index and the extreme quantiles are out of the question since the tail sample size is fixed. Fortunately, inference about these tail features is still implementable, as we discuss in this section.

We first study the tail index ξ. EV theory (the Fisher–Tippett–Gnedenko theorem) suggests that when the underlying distribution is within the maximum domain of attraction (e.g., Chapter 1 of de Haan and Ferreira (2007)), the sample maximum is asymptotically distributed as the EV distribution, which is parametric and entirely characterized by ξ. Specifically, our Condition 2 is sufficient for the maximum domain of attraction assumption. Then, EV theory implies that there exist sequences of constants a_n and b_n such that, up to some location and scale normalization,

$$\frac{Y_{(1)} - b_n}{a_n} \xrightarrow{d} X_1, \tag{6}$$

where the CDF of X_1 is given by

$$V_{\xi}(x)=\begin{cases}1-\exp\left(-\left(1+\xi x/\sigma\right)^{-1/\xi}\right) & \xi\neq 0\\ 1-\exp\left(-\exp\left(-x/\sigma\right)\right) & \xi=0.\end{cases} \tag{7}$$

We can subsume σ into a_n so that σ is always 1 in (7). In additional to the sample maximum, EV theory also extends to the first $m+k$ order statistics such that if (6) holds, then for any fixed m and k,

$$\left(\frac{Y_{(1)}-b_n}{a_n},...,\frac{Y_{(m+k)}-b_n}{a_n}\right)^{\intercal}\xrightarrow{d}\left(X_1,...,X_{m+k}\right)^{\intercal}. \tag{8}$$

The joint density of $(X_1,\ldots,X_{m+k})^{\intercal}$ is given by $V_{\xi}(x_{m+k})\prod_{i=1}^{m+k}v_{\xi}(x_i)/V_{\xi}(x_i)$ on $x_{m+k}\leq x_{m+k-1}\leq\ldots\leq x_1$, where $v_{\xi}(x)=dV_{\xi}(x)/dx$.

Since the first m elements are censored, the effective tail observations asymptotically reduce to

$$\mathbf{X}=\left(X_{m+1},...,X_{m+k}\right)^{\intercal},$$

whose density is derived in the following proposition.

Proposition 3 *Suppose Conditions 1 and 2 hold. Then for any fixed positive integers m and k, there exist sequences of constants a_n and b_n such that*

$$\frac{\left(Y_{(m+1)},\ldots,Y_{(m+k)}\right)^{\intercal}-b_n\iota_k}{a_n}\xrightarrow{d}\mathbf{X},$$

where ι_k denotes the $k\times 1$ vector of ones, and the joint density of $\mathbf{X}$ is given by

$$\begin{aligned}f_{\mathbf{X}|\xi}\left(x_{m+1},...,x_{m+k}\right) &= \frac{1}{m!}\left(-\log V_{\xi}\left(x_{m+1}\right)\right)^{m}V_{\xi}\left(x_{m+k}\right)\prod_{i=m+1}^{m+k}v_{\xi}\left(x_i\right)/V_{\xi}\left(x_i\right) \qquad (9)\\ &= \frac{1}{m!}\exp\left(\begin{array}{c}-\frac{m}{\xi}\log\left(1+\xi x_{m+1}\right)-\left(1+\xi x_{m+k}\right)^{-1/\xi}\\ -\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k}\log\left(1+\xi x_{m+i}\right)\end{array}\right).\end{aligned}$$

It is clear that the elements of $\mathbf{X}$ are dependent as captured by the term $\left(-\log V_{\xi}\left(x_{m+1}\right)\right)^{m}V_{\xi}\left(x_{m+k}\right)$, which is negligible if k is large but plays a vital role when k is only moderate. This is the fundamental difference between the increasing-k and the fixed-k asymptotic embeddings, which are respectively used by the MLE and its small sample modification. From now on, we use bold letters to denote vectors.

If the constants a_n and b_n were known, the vector

$$\mathbf{Y} = \left(Y_{(m+1)}, \ldots, Y_{(m+k)}\right)^{\intercal}$$

is then approximately distributed as $\mathbf{X}$, and the limiting problem is reduced to the small sample parametric one: constructing a confidence interval based on one draw $\mathbf{X}$ whose density $f_{\mathbf{X}|\xi}$ is known up to ξ. However, a_n and b_n respectively correspond to the scale σ and the tail location u. Therefore, they ultimately depend on F and are challenging to estimate. Consider the standard Pareto distribution, for example. The Pareto exponent α is simply $1/\xi$. Then the fact that $a_n = n^{\xi}$ implies that a small estimation bias in ξ could be amplified by the n-power and lead to a poor inference.

To avoid the knowledge (and estimation) of a_n and b_n, we consider the following self-normalized statistics:

$$\begin{aligned}\mathbf{Y}^* &= \frac{\mathbf{Y} - Y_{(m+k)}\iota_k}{Y_{(m+1)} - Y_{(m+k)}} \\ &= \left(1, \frac{Y_{(m+2)} - Y_{(m+k)}}{Y_{(m+1)} - Y_{(m+k)}}, \ldots, \frac{Y_{(m+k-1)} - Y_{(m+k)}}{Y_{(m+1)} - Y_{(m+k)}}, 0\right)^{\intercal}.\end{aligned} \tag{10}$$

It is easy to establish that $\mathbf{Y}^*$ is maximal invariant with respect to the group of location and scale transformations (cf. Chapter 6 of Lehmann and Romano (2005)). In words, the statistic constructed as a function of $\mathbf{Y}^*$ remains unchanged if data are shifted and multiplied by any non-zero constant. This invariance is also intuitive since the tail shape should preserve no matter how data are linearly transformed.

The continuous mapping theorem and Proposition 3 yield that

$$\mathbf{Y}^* \xrightarrow{d} \mathbf{X}^* = \left(1, \frac{X_{m+2} - X_{m+k}}{X_{m+1} - X_{m+k}}, \ldots, \frac{X_{m+k-1} - X_{m+k}}{X_{m+1} - X_{m+k}}, 0\right)^{\intercal}, \tag{11}$$

which is again invariant to location and scale transformation. By change of variables, the density of $\mathbf{X}^*$ is given by

$$f_{\mathbf{X}^*|\xi}\left(\mathbf{x}^*\right) = \frac{\Gamma\left(k+m\right)}{m!} \int_0^{\infty} s^{k-2} \exp\left(-\frac{m}{\xi} \log\left(1+\xi s\right)\right) e\left(\mathbf{x}^*, s\right) ds, \tag{12}$$

where $e\left(\mathbf{x}^*, s\right) = \exp\left(-(1+1/\xi)\sum_{i=1}^{k} \log(1+\xi x_i^* s)\right)$ and x_i^* denotes the ith component of $\mathbf{x}^*$. See Appendix A.1 for more details.

The density (12) can be used to conduct inference about ξ. In particular, consider the hypothesis testing problem

$$H_0 : \xi = \xi_0 \text{ against } H_1 : \xi \in \Xi \backslash \{\xi_0\},$$

where Ξ denotes the parameter space of ξ. We set $\Xi = (0, 1)$ in the Monte Carlo simulations to cover the distributions with an unbounded support and a finite mean, which can be easily extended. Since the alternative hypothesis is composite, we follow Andrews and Ploberger (1994) and Elliott, Müller, and Watson (2015) to consider the weighted average alternative

$$\int_{\Xi} f_{\mathbf{X}^*|\xi} \left(\mathbf{x}^*\right) dW(\xi),$$

where the weighting measure W reflects the importance a researcher attaches to different alternative values of ξ. In practice, we set $W(\cdot)$ to be the CDF of the standard uniform distribution for simplicity.

Given the density of $\mathbf{X}^*$, we construct the likelihood-ratio test as

$$\varphi\left(\mathbf{x}^*\right) = \mathbf{1}\left[\frac{\int_{\Xi} f_{\mathbf{X}^*|\xi}\left(\mathbf{x}^*\right) dW(\xi)}{f_{\mathbf{X}^*|\xi_0}\left(\mathbf{x}^*\right)} > \mathrm{cv}\left(\xi_0, k, m\right)\right], \tag{13}$$

where $\mathrm{cv}\left(\xi_0, k, m\right)$ denotes the critical value that depends on the significance level, the null value ξ_0, the tail sample size k, and the number of the censored observations m. The critical value is obtained by simulation, and the test is implemented by replacing $\mathbf{X}^*$ with $\mathbf{Y}^*$ in finite samples. The continuous mapping theorem and Proposition 3 yield that $\mathbb{E}\left[\varphi\left(\mathbf{Y}^*\right)\right] \rightarrow \mathbb{E}\left[\varphi\left(\mathbf{X}^*\right)\right]$, which equals the nominal level under the null hypothesis. The confidence interval is constructed by inverting the test. Note that this method does not require the knowledge of the censoring threshold T, which applies to cases such as the macroeconomic disaster.

Following Müller and Wang (2017), we can also construct the confidence intervals of the extreme quantiles under the fixed-k asymptotics. To this end, we focus on the $Q(1 - p_n)$ quantile with $p_n = O(n^{-1})$, which captures the fact that the object of interest is of the same order of magnitude as the sample maximum. In particular, we consider $p_n = h/n$ for some fixed $h > 0$. Then EV theory implies that $\left(Q\left(1 - h/n\right) - b_n\right)/a_n$ converges to the e^{-h} quantile of X_1, denoted $q(\xi, h) = \left(h^{-\xi} - 1\right)/\xi$. Again, the research problem would become inference about $q(\xi, h)$ based on the $k \times 1$ vector of observations $\mathbf{X}$ if a_n and b_n were known. Without loss of generality, we construct a confidence set $S(\mathbf{Y}) \subset \mathbb{R}$ such that

$\mathbb{P}\left(Q\left(1-p_n\right) \in S(\mathbf{Y})\right) \geq 1-\mathrm{lv}$, at least as $n \rightarrow \infty$, where lv denotes the significance level.

To eliminate a_n and b_n, we use the self-normalized vector $\mathbf{Y}^*$ as in (12). Besides, we also impose location and scale equivariance on our confidence interval S. Specifically, we impose that for any constants $a>0$ and b, our interval S satisfies that $S(a\mathbf{Y}+b)=aS(\mathbf{Y})+b$, where $aS(\mathbf{Y})+b=\{y:(y-b)/a \in S(\mathbf{Y})\}$. Under this equivariance constraint, we can write

$$\begin{aligned} \mathbb{P}(Q\left(1-p_n\right) \in S(\mathbf{Y})) &= \mathbb{P}\left(\frac{Q\left(1-p_n\right)-b_n}{a_n} \in S\left(\frac{\mathbf{Y}-b_n\iota_k}{a_n}\right)\right) \\ &= \mathbb{P}\left(\frac{Q\left(1-p_n\right)-Y_{(m+k)}}{Y_{(m+1)}-Y_{(m+k)}} \in S\left(\mathbf{Y}^*\right)\right) \\ &\rightarrow \mathbb{P}_\xi\left(\frac{q(\xi,h)-X_{m+k}}{X_{m+1}-X_{m+k}} \in S(\mathbf{X}^*)\right), \end{aligned}$$

where the notation $\mathbb{P}_\xi$ (and $\mathbb{E}_\xi$ below) indicates that the randomness is entirely characterized by ξ asymptotically. The asymptotic problem then is the construction of a location and scale equivariant S that satisfies

$$\mathbb{P}_\xi\left(\frac{q(\xi,h)-X_{m+k}}{X_{m+1}-X_{m+k}} \in S(\mathbf{X}^*)\right) \geq 1-\mathrm{lv} \text{ for all } \xi \in \Xi \tag{14}$$

since any S that satisfies Proposition 3 and the equivariance constraint also satisfies

$$\liminf_{n\rightarrow\infty} \mathbb{P}(Q\left(1-p_n\right) \in S(\mathbf{Y})) \geq 1-\mathrm{lv}.$$

This problem involves a single observation $\mathbf{X} \in \mathbb{R}^k$ from a parametric distribution indexed only by the scalar parameter $\xi \in \Xi$.

In principle, there could still be many solutions that satisfy the asymptotic size constraint. To obtain the optimal one, we consider the weighted average expected length criterion

$$\int \mathbb{E}_\xi[\mathrm{lgth}(S(\mathbf{X}))]dW(\xi), \tag{15}$$

where W again denotes some weighting measure on Ξ, and $\mathrm{lgth}(A)=\int \mathbf{1}[y \in A]dy$ for any Borel set $A \subset \mathbb{R}$.

To solve the program of minimizing (15) subject to (14) among all equivariant set estimators S, we introduce

$$Y^*(\xi)=\frac{q(\xi,h)-X_{m+k}}{X_{m+1}-X_{m+k}},$$

and write $\mathbb{E}_\xi[\mathrm{lgth}(S(\mathbf{X}))] = \mathbb{E}_\xi[(X_{m+1} - X_{m+k})\,\mathrm{lgth}(S(\mathbf{X}^*))] = \mathbb{E}_\xi[\kappa_\xi(\mathbf{X}^*)\,\mathrm{lgth}(S(\mathbf{X}^*))]$ with $\kappa_\xi(\mathbf{X}^*) = \mathbb{E}_\xi[X_{m+1} - X_{m+k}|\mathbf{X}^*]$. Thus, our problem becomes

$$\begin{array}{l} \min_{S(\cdot)} \int_\Xi \mathbb{E}_\xi[\kappa_\xi(\mathbf{X}^*)\,\mathrm{lgth}(S(\mathbf{X}^*))]dW(\xi) \\ \text{s.t. } \mathbb{P}_\xi\left(Y^*(\xi) \in S(\mathbf{X}^*)\right) \geq 1 - \mathrm{lv} \text{ for all } \xi \in \Xi. \end{array} \tag{16}$$

There are two advantages to translate the asymptotic problem into (16). First, (16) does not require the knowledge of the censoring threshold T but only the number of censored observations m. Second, (16) only involves S evaluated at $\mathbf{X}^*$ and hence $\mathbf{Y}^*$ in practice. This means the knowledge of a_n and b_n is asymptotically unnecessary as long as n is sufficiently larger than $m + k$. Note that any solution to (16) also provides the form of S, that is, $S(\mathbf{X}) = (X_{m+1} - X_{m+k})S(\mathbf{X}^*) + X_{m+k}$. So once $S(\cdot)$ is determined, the confidence interval can be constructed in practice by plugging in

$$(Y_{(m+1)} - Y_{(m+k)})S(\mathbf{Y}^*) + Y_{(m+k)}.$$

To make further progress in solving (16), we write the problem in the following Lagrangian form:

$$\min_{S(\cdot)} \int_\Xi \mathbb{E}_\xi[\kappa_\xi(\mathbf{X}^*)\,\mathrm{lgth}(S(\mathbf{X}^*))]dW(\xi) + \int_\Xi \mathbb{P}_\xi\left(Y^*(\xi) \in S(\mathbf{X}^*)\right) d\Lambda(\xi),$$

where the non-negative measure Λ denotes the Lagrangian weights that guarantee the asymptotic size constraint. By writing the expectations above as integrals over the densities $f_{\mathbf{X}^*|\xi}$ of $\mathbf{X}^*$ and $f_{Y^*(\xi),\mathbf{X}^*|\xi}$ of $(Y^*(\xi), \mathbf{X}^*)$, the solution of the above problem is given by

$$S(\mathbf{x}^*) = \left\{ y : \int_\Xi \kappa_\xi(\mathbf{x}^*) f_{\mathbf{X}^*|\xi}(\mathbf{x}^*)dW(\xi) < \int_\Xi f_{Y^*(\xi),\mathbf{X}^*|\xi}(y, \mathbf{x}^*)d\Lambda(\xi) \right\}. \tag{17}$$

The integrals can be numerically calculated by Gaussian quadrature, and then the only remaining challenge is to find some suitable Lagrangian weights Λ. We solve this challenge by the numerical approach developed in Elliott, Müller, and Watson (2015). The MATLAB program and the weights Λ are available at the author's website. Note that Λ only needs to be computed once by the author instead of empirical researchers. Then the most time-consuming part in solving the program (16) is the numerical integration, which costs only a few seconds in a modern PC. Further details are provided in Appendix A.1.

4 Monte Carlo Simulations

This section examines the finite sample performance of the proposed method and compares it with several popular existing methods. We generate random samples from four commonly used distributions: the generalized Pareto distribution with $\xi = 0.5$ and $\sigma = 1$ (GPD), the absolute value of the Student-t distribution with 2 degrees of freedom (|t(2)|), the F distribution with parameters 4 and 4 (F(4,4)), and the double Pareto-lognormal distribution (dPlN), that is,

$$Y = \exp\left(c_1 + c_2 Z_1 + \xi Z_2 - c_3 Z_3\right),$$

where Z_1, Z_2, Z_3 are independent and $Z_1 \sim N(0,1)$, and $Z_2, Z_3 \sim Exp(1)$. For parameter values, we set $c_1 = 0$, $c_2 = 0.5$, $\xi = 0.5$, and $c_3 = 1$, which are typical values for income data as documented in Toda (2012). In particular, the dPlN distribution is the product of independent double Pareto and lognormal variables. It has been documented to fit well to size distributions of economic variables including income (Reed (2003)), city size (Giesen, Zimmermann, and Suedekum (2010)), and consumption (Toda (2017)). In all four DGP's, the true value of the tail index is 0.5. Regarding the tail censoring, we set the censoring threshold T as the 99% and 99.9% quantiles of the underlying distributions, implying that the censored probability (cen_p) is either 1% or 0.1%.

We first consider some widely used estimators in empirical studies. Due to space limitations, we only report the results of Hill (1975)'s estimator and the bias-corrected estimator proposed by Gabaix and Ibragimov (2011) (denoted GI). The confidence intervals are based on their asymptotic normality and the plug-in estimators of their asymptotic variances. The sample size n is 1000, 2000, and 5000, and k is set as $[0.05n]$ for both methods, where $[A]$ denotes the closest integer of A. All results are based on 1000 simulations.

Table 2 depicts the mean biases and the coverage probabilities of these two methods. Several key findings can be summarized as follows. First, both the Hill and the GI estimators suffer from severe biases, and the confidence intervals based on them exhibit substantial undercoverage. This holds even if the censoring probability is only 0.1%. Second, ignoring the upper tail censoring tends to underestimate the tail index, which implies a misleadingly thin tail. This is seen in Section 5.2 when we study the macroeconomic disasters. Finally, unreported results show that other methods reviewed in Chapter 3 of de Haan and Ferreira (2007) also suffer from substantial undercoverage. Therefore, it is crucial to take the censoring into account, even if the censoring probability is tiny.

Table 2: Small Sample Properties of Estimation and Inference about Tail Index, Igorning Tail Censoring

cen_p	1%				0.1%			
	Bias		Cov		Bias		Cov	
n=1000	Hill	GI	Hill	GI	Hill	GI	Hill	GI
GPD	-0.18	-0.24	0.03	0.00	-0.04	-0.07	0.88	0.89
t(2)	-0.16	-0.23	0.10	0.00	-0.02	-0.06	0.93	0.93
F(4,4)	-0.12	-0.20	0.39	0.05	0.03	-0.03	0.98	0.98
dPlN	-0.17	-0.24	0.05	0.00	-0.04	-0.04	0.89	0.90
n=2000	Hill	GI	Hill	GI	Hill	GI	Hill	GI
GPD	-0.18	-0.24	0.00	0.00	-0.04	-0.07	0.85	0.81
t(2)	-0.16	-0.23	0.00	0.00	-0.02	-0.06	0.93	0.86
F(4,4)	-0.12	-0.20	0.12	0.00	0.03	-0.03	0.97	0.97
dPlN	-0.17	-0.24	0.00	0.00	-0.04	-0.04	0.88	0.82
n=5000	Hill	GI	Hill	GI	Hill	GI	Hill	GI
GPD	-0.18	-0.24	0.00	0.00	-0.04	-0.08	0.73	0.45
t(2)	-0.16	-0.23	0.00	0.00	-0.02	-0.07	0.90	0.63
F(4,4)	-0.12	-0.20	0.00	0.00	0.03	-0.03	0.93	0.95
dPlN	-0.17	-0.24	0.00	0.00	-0.03	-0.08	0.77	0.47

Note: Entries are the biases and coverage probabilities (Cov) of the 95% confidence intervals based on Hill's estimator (Hill) and Gabaix and Ibragimov (2010)'s estimator (GI). Data are generated from the Pareto(0.5), the absolute value of Student-t(2), the F(4,4), and the dPlN distributions with the censored probability (cen_p) being 1% or 0.01%. The results are based on 1000 simulations.

Now we implement the new method proposed in Sections 2 and 3. Table 3 depicts the coverage and length of the 95% maximum likelihood confidence intervals (denoted ml) based on Proposition 1 and those of the fixed-k intervals (denoted fk) by inverting (13). Several interesting findings can be made as follows. First, the maximum likelihood confidence intervals are substantially longer than the fixed-k ones when the sample size is not large. Besides, the coverage probability is smaller than the nominal level when the censoring is at the 99.9% quantile. This is because the asymptotic normality cannot perform well when k is not large. Second, in comparison, the fixed-k ones always deliver the nominal size with shorter length, especially when the sample size is not large. Finally, when n reaches

5000 (and k reaches 250), the maximum likelihood intervals are comparable with the fixed-k ones. Hence a simple rule-of-thumb choice of the switching cutoff is $k \lesseqgtr 250$, provided n is sufficiently large.

Table 3: Small Sample Properties of Inference about Tail Index

cen_p	1%				0.1%			
	Cov		Lgth		Cov		Lgth	
n=1000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.98	0.93	1.39	0.73	0.91	0.95	0.88	0.70
t(2)	0.98	0.94	1.40	0.73	0.90	0.95	0.87	0.70
F(4,4)	0.99	0.93	1.39	0.73	0.90	0.95	0.87	0.70
dPlN	0.99	0.93	1.30	0.73	0.90	0.94	0.87	0.70
n=2000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.96	0.94	0.99	0.69	0.93	0.94	0.63	0.58
t(2)	0.96	0.93	0.99	0.69	0.92	0.93	0.62	0.58
F(4,4)	0.96	0.94	0.99	0.69	0.93	0.94	0.63	0.58
dPlN	0.96	0.94	0.99	0.69	0.92	0.93	0.62	0.58
n=5000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.97	0.94	0.63	0.54	0.95	0.94	0.40	0.39
t(2)	0.96	0.94	0.62	0.54	0.93	0.92	0.40	0.38
F(4,4)	0.97	0.95	0.63	0.54	0.94	0.93	0.40	0.38
dPlN	0.96	0.93	0.62	0.54	0.94	0.94	0.40	0.38

Note: Entries are the coverage probabilities (Cov) and the averaged length (Lgth) of the maximum likelihood intervals (ml) and the fixed-k intervals (fk) for the tail index. Data are generated from the Pareto(0.5), the absolute value of Student-t(2), the F(4,4), and the dPlN distributions with the censored probability (cen_p) being 1% or 0.01%. The results are based on 1000 simulations. The level of significance is 5%.

Tables 4 depicts the coverage probabilities and lengths of the confidence intervals of the 99% quantile, using either the maximum likelihood method as in Proposition 2 or the fixed-k method (17). Both methods deliver satisfactory size and length properties, although the maximum likelihood intervals suffer from slight undercoverage. However, as we target the more extreme 99.9% quantile as in Table 5, such undercoverage is substantial when k is less than 250. In contrast, the fixed-k ones always perform excellently. These results reinforce our switching cutoff at $k = 250$.

Table 4: Small Sample Properties of Inference about the 0.99 Quantile

cen_p	1%				0.1%			
	Cov		Lgth		Cov		Lgth	
n=1000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.95	0.94	6.71	7.09	0.91	0.96	4.83	5.79
t(2)	0.95	0.94	6.73	7.13	0.91	0.94	4.87	5.89
F(4,4)	0.94	0.94	11.67	12.17	0.91	0.95	8.32	9.91
dPlN	0.95	0.94	4.92	5.25	0.91	0.95	3.54	4.37
n=2000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.96	0.94	4.60	4.86	0.92	0.96	3.38	3.88
t(2)	0.96	0.94	4.60	4.76	0.93	0.96	3.43	3.89
F(4,4)	0.96	0.95	8.04	8.09	0.92	0.95	5.85	6.72
dPlN	0.96	0.94	3.44	3.50	0.93	0.96	2.52	2.90
n=5000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.97	0.96	2.85	2.93	0.93	0.94	2.12	2.38
t(2)	0.97	0.95	2.84	2.91	0.94	0.96	2.15	2.40
F(4,4)	0.97	0.95	4.93	5.05	0.93	0.94	3.68	4.10
dPlN	0.97	0.95	2.08	2.12	0.93	0.96	1.57	1.77

Note: Entries are the coverage probabilities (Cov) and the averaged length (Lgth) of the maximum likelihood intervals (ml) and the fixed-k intervals (fk) for the 99% quantiles. Data are generated from the Pareto(0.5), the absolute value of Student-t(2), the F(4,4), and the dPIN distributions with the censored probability (cen_p) being 1% or 0.01%. The results are based on 1000 simulations. The level of significance is 5%.

Table 5: Small Sample Properties of Inference about the 0.999 Quantile

cen_p	1%				0.1%			
	Cov		Lgth		Cov		Lgth	
n=1000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.86	0.92	128.9	102.6	0.83	0.95	61.68	75.08
t(2)	0.85	0.93	123.0	103.1	0.83	0.94	60.87	72.62
F(4,4)	0.85	0.91	226.8	178.1	0.83	0.95	106.3	130.8
dPlN	0.85	0.93	93.16	78.21	0.82	0.95	45.16	55.13
n=2000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.89	0.92	79.20	72.25	0.87	0.94	39.71	44.91
t(2)	0.88	0.94	75.60	71.65	0.88	0.95	39.32	45.96
F(4,4)	0.89	0.92	136.6	126.5	0.88	0.93	69.35	80.28
dPlN	0.88	0.92	56.17	51.48	0.87	0.93	28.99	32.38
n=5000	ml	fk	ml	fk	ml	fk	ml	fk
GPD	0.94	0.95	43.06	40.41	0.92	0.94	23.96	24.15
t(2)	0.92	0.95	40.75	39.25	0.92	0.93	23.39	24.15
F(4,4)	0.92	0.95	73.55	70.90	0.91	0.94	41.11	41.62
dPlN	0.92	0.92	30.65	28.32	0.91	0.93	17.31	18.13

Note: Entries are the coverage probabilities (Cov) and the averaged length (Lgth) of the maximum likelihood intervals (ml) and the fixed-k intervals (fk) for the 99.9% quantiles. Data are generated from the Pareto(0.5), the absolute value of Student-t(2), the F(4,4), and the dPlN distributions with the censored probability (cen_p) being 1% or 0.01%. The results are based on 1000 simulations. The level of significance is 5%.

5 Empirical Applications

Many applications in economics and finance involve estimation and inference of tail features with censored data. In this section, we apply the proposed method to the two datasets we discussed earlier in this paper. Our empirical analysis highlights the potential of our approach.

5.1 US Individual Earnings

Our first application is about the tail features of the individual earnings distribution. Following the convention, we use the variable ERN_VAL in the March CPS dataset and drop the individuals that are younger than 18 or older than 70 years old. This yields 115,424 observations in the 2019 sample. The censoring threshold is 310000 USD, which leads to a 0.58% censoring fraction in the full sample and various censoring fractions in different subsamples. The first several columns in Table 6 present the sample sizes (n) and the numbers (cen#) and the fractions (cen%) of the censored observations, respectively. We use the previously introduced method to construct the 95% confidence intervals of the tail index and the 99% and 99.9% quantiles. Specifically, we follow the simulation study to use the maximum likelihood confidence intervals developed in Section 2 when k is larger than 250 and switch to the fixed-k confidence intervals (17) otherwise. The last six columns in Table 6 present the results with $k = [0.05n]$. The results based on other choices are similar and reported in Appendix A.3.

Several interesting findings can be summarized as follows. First, in Panel A, the tail index is around 0.5 in the full sample, as commonly found in the existing literature. But it is substantially different across subsamples. Second, the tail also exhibits substantial heterogeneity across genders. In particular, the male sample has significantly higher quantiles than the female at both the 99% and 99.9% levels. Third, this difference also exists across races. In particular, the 99.9% quantile of all males is at least twice larger than that of the black males. All such heterogeneity provides new evidence for potential racial and gender discrimination. Finally, Panel B depicts the heterogeneity across ages, with substantially heavier tails showing up in the middle-aged groups.

Table 6: Empirical Results in 2019 March CPS Data

Panel A: 95% confidence intervals in race-and-gender-based subsamples									
race-gender	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
full sample	115424	672	0.58	(0.41	0.52)	(24.24	25.32)	(62.63	75.61)
all males	55553	491	0.88	(0.35	0.53)	(28.64	30.68)	(67.71	92.58)
all females	59871	181	0.30	(0.42	0.55)	(18.02	19.03)	(46.34	58.01)
white males	43371	419	0.97	(0.88	1.00)	(29.83	33.56)	(145.2	279.0)
white females	45424	141	0.31	(0.42	0.57)	(18.09	19.28)	(46.56	60.66)
Asian males	3676	50	1.36	(0.00	0.45)	(30.73	37.20)	(48.03	86.76)
Asian females	4099	22	0.54	(0.35	0.94)	(20.77	27.16)	(45.12	145.0)
Hispanic males	44420	445	1.00	(0.71	0.95)	(31.10	34.75)	(119.3	214.0)
Hispanic females	48192	155	0.32	(0.47	0.62)	(18.70	19.90)	(50.17	66.13)
black males	6144	12	0.20	(0.22	0.58)	(15.92	18.22)	(28.73	49.39)
black females	7827	9	0.16	(0.16	0.44)	(13.90	15.64)	(25.25	37.18)
Panel B: 95% confidence intervals in age-based subsamples									
age	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
18-30	27829	35	0.13	(0.33	0.49)	(12.69	13.60)	(28.16	36.16)
30-40	25213	158	0.63	(0.28	0.50)	(24.12	26.36)	(52.24	75.73)
40-50	23419	213	0.91	(0.83	1.00)	(28.89	33.82)	(119.7	297.2)
50-60	21767	196	0.90	(0.52	0.83)	(28.19	32.21)	(78.05	154.2)
60-65	17196	70	0.41	(0.17	0.41)	(20.95	23.13)	(41.86	59.31)

Note: Entries are the sample size (n), the number of censored observations (cen#), the censored fraction in percentage points (cen%), 95% confidence intervals of the tail index and those of the 99% and 99.9% quantiles measured in 10^4 USD. The results are based on the variable ERN_VAL in the CPS dataset and equivalently the variable inclongj from the IPUMS dataset. Data are available at https://cps.ipums.org/cps.

5.2 Macroeconomic Disasters

This section studies the size distribution of macroeconomic disasters, which is an important research topic in macroeconomics. Barro and Ursúa (2008) and Barro and Jin (2011) construct and analyze the dataset that consists of annual GDP (and consumption) growth rates in 36 countries from 1870 to 2005. The authors sort these observations and define a macroeconomic disaster if the GDP declines by more than 10%. This leads to $k = 157$ tail observations. Then the authors fit these data to the (double) Pareto distribution to estimate the Pareto exponent, which is the reciprocal of the tail index, and back out the coefficient of the relative risk aversion by a theoretical model (eq.2 in Barro and Jin (2011)).

However, the largest disasters tend to be missing because some governments collapsed or were fighting wars (p.1581 in Barro and Jin (2011)). Ignoring these missing data in the upper tail could lead to substantial bias, as we show in the Monte Carlo simulations. We revisit this problem by applying our fixed-k method since k is only moderate. Specifically, the most recent data missing happens in four countries, which are Greece, Malaysia, the Philippines, and Singapore during WWII. Therefore, we set $m = 4$ and apply the fixed-k method to construct the 95% confidence intervals for the tail index ξ and those for the coefficient of relative risk aversion by solving eq.2 in Barro and Jin (2011). For comparison, we also construct the intervals based on Hill (1975)'s estimator and the bias-reduced estimator (GI) proposed by Gabaix and Ibragimov (2011). Table 7 presents the result.

As shown in the table, the fixed-k intervals contain substantially larger values of the tail index than the other two methods that ignore the tail censoring. This is coherent with our simulation results in Table 2. By taking the reciprocal, the Pareto exponent is estimated to be approximately 7 in Barro and Jin (2011) but less than 1 by the new method. Therefore, taking the tail censoring into account leads to a substantially heavier tail in the disaster size. Accordingly, the coefficient of risk aversion is found to be around 0.75, which is significantly lower than 3 in Barro and Jin (2011). These results undermine their conclusion that "the (Hill) estimate of the upper-tail exponent is likely to have only a small upward bias due to missing extreme observations, which have to be few in number."

Table 7: Empirical Results in Macroeconomic Disasters

Method	Hill		GI		New	
			Tail Index			
95% CIs	(0.12	0.17)	(0.16	0.26)	(0.57	1.00)
			Coefficient of Risk Aversion			
95% CIs	(3.73	5.10)	(2.44	3.88)	(0.58	1.04)

Note: Entries are 95% confidence intervals (CIs) of the tail index of the disaster size distribution and the coefficient of risk aversion, based on the Hill estimator (Hill), the bias-reduced estimator proposed by Gabaix and Ibragimov (2011) (GI), and the fixed-k method by inverting (13). Data are available at https://scholar.harvard.edu/barro/data_sets.

6 Concluding Remarks

This paper develops a new approach to estimate and conduct inference about tail features for censored data. The method can be viewed as a hybrid approach that uses the maximum likelihood estimation when the tail sample size is large and switches to a small sample modification otherwise. As shown in Monte Carlo simulations, the new method has excellent small sample performance.

This new approach is empirically relevant in broad areas studying tail features (e.g., tail index and extreme quantiles). We illustrate this with the March CPS data and the macroeconomic disaster data and find considerably different results from the existing literature.

There are theoretical extensions and empirical applications of our method, which we suppress in the current paper due to space limitations. We list a few here. First, our method naturally applies to the no censoring case by setting $\kappa = \infty$ in the MLE and $m = 0$ in the fixed-k method. Besides, we can follow Müller and Wang (2019) to construct the (quantile) unbiased estimation of the tail features, which could perform better in terms of mean absolute deviation and mean squared error, especially when k is not large.

Second, many other tail features can be learned by our new method as long as they can be expressed as functions of the tail index. For example, the conditional tail expectation is another important risk measure in finance, which is defined as the expectation conditional on being larger than some high quantile, that is, $\mathbb{E}\left[Y_i|Y_i > Q(1-p)\right]$. By reparametrizing $p = h/n$ for some $h > 0$ and using EV theory, we can obtain that that $(\mathbb{E}\left[Y_i|Y_i > Q(1-h/n)\right] - b_n)/a_n \to h^{-\xi}/(\xi(1-\xi)) - 1/\xi$, which again entirely depends on ξ and h (p.1336 in Müller and Wang (2019)). Then we can construct the fixed-k intervals for this quantity in an analogous fashion to (16).

Finally, our method also allows from weak dependent data if some additional regularity condition is satisfied. In particular, EV theory holds under weak dependence, such as α-mixing, as long as the largest order statistics do not show up in a cluster. This is referred to as the non-cluster condition. See, for example, Leadbetter (1983), O'Brien (1987), Mikosch and Stărică (2000), Chernozhukov (2005), and Chernozhukov and Fernández-Val (2011).

Appendix

A.1 Computational Details

The estimators defined in Section 3 require evaluation of some expectations. Define $\Gamma(\cdot)$ as the Gamma function and $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$ as the incomplete Gamma function. Also define $e(\mathbf{x}^*, s) = \exp\left(-(1+1/\xi)\sum_{i=1}^k \log(1+\xi x_i^* s)\right)$. Change of variables and integration by parts yield that

$$\begin{aligned}
&\mathbb{E}_\xi\left[X_{m+1} - X_{m+k}|\mathbf{X}^* = \mathbf{x}^*\right] f_{\mathbf{X}^*|\xi}(\mathbf{x}^*) \\
= &\frac{\Gamma(k+m-\xi)}{m!}\int_0^{+\infty} s^{k-1}\exp\left(-\frac{m}{\xi}\log(1+\xi s) - (1+\frac{1}{\xi})\sum_{i=1}^k \log(1+\xi x_i^* s)\right) ds.
\end{aligned}$$

and

$$\begin{aligned}
&f_{Y^*(\xi),\mathbf{X}^*|\xi}(y, \mathbf{x}^*) \\
= &\frac{1}{m!}\int_0^{+\infty} \mathbf{1}\left[s + \frac{x_i^*}{y}(q(\xi) - s) > 0 \text{ for all } i \text{ and } x_i^* > x_j^* \text{ if } i < j\right]\left|\left(\frac{q(\xi,h) - s}{y}\right)^{k-1}\frac{1}{y}\right| \\
&\times \exp\begin{pmatrix} -\frac{m}{\xi}\log\left(1+\xi x + \frac{1}{y}\xi(q(\xi,h) - x)\right) - (1+\xi s)^{-1/\xi} \\ -(1+\frac{1}{\xi})\sum_{i=1}^k \log\left(1+\xi s + \frac{x_i^*}{y}\xi(q(\xi) - s)\right)\end{pmatrix} ds,
\end{aligned}$$

where $q(\xi,h) = \left(h^{-\xi} - 1\right)/\xi$. We evaluate these by numerical quadrature.

To determine the Lagrange multipliers Λ, we use the algorithm developed by Elliott, Müller, and Watson (2015). In particular, we restrict $\Lambda(\cdot)$ to be point masses with the support on the discretized Ξ, that is, $\Xi_D = \{1/50, 2/50, ..., 1\}$ and determine the 50 point masses by fixed-point iterations based on Monte Carlo estimates of the coverage probabilities. To do this, we simulate the coverage probabilities with 20000 i.i.d. draws from a proposal with ξ uniformly drawn from Ξ_D, and iteratively increase or decrease the point masses on Ξ_D as a function of whether the coverage given that value of ξ is larger or smaller than the nominal level 0.05. Stop this iteration until the differences between the coverages for all values of ξ and 0.05 are lower than a pre-specified tolerance $\varepsilon = 0.001$. This tolerance can be arbitrarily small at the cost of longer computation time and larger numbers of simulation draws.

For any given k, h, and m, the Lagrange multipliers only need to be determined once.

The tables of the Lagrange multipliers and the corresponding MATLAB code are provided on our website: https://sites.google.com/site/yulongwanghome/.

A.2 Proof

To prove Proposition 1, we first establish two intermediate results, Lemmas 1 and 2. Throughout the proof, we suppress the subscript n in T_n and u_n for notational simplicity.

Lemma 1 *Suppose the conditional CDF F_u is exactly the GPD (2). Denote the log-likelihood as*

$$l_i(\xi,\sigma) = D_i \log\left(1 - G\left(T_u;\xi,\sigma\right)\right) + \left(1 - D_i\right)\log g\left(Y_i;\xi,\sigma\right),$$

where $T_u = T - u$. Then the elements of the Fisher-information matrix are given by

$$\begin{aligned} M_{11} &\equiv \mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\xi^2}\right] = \frac{2}{(1+\xi)(1+2\xi)} + \frac{z^{-2-1/\xi}}{(1+\xi)(1+2\xi)\xi^2}\times \\ &\quad \left\{-1-\xi+z(2+4\xi)-z^2(1+\xi)(1+2\xi)\right\} \\ M_{22} &\equiv \mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\sigma^2}\right] = \frac{1}{(1+2\xi)\sigma^2} - \frac{z^{-2-1/\xi}}{(1+2\xi)\sigma^2} \\ M_{12} &\equiv \mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\sigma\partial\xi}\right] = \frac{1}{(1+\xi)(1+2\xi)\sigma} + \frac{z^{-2-1/\xi}}{(1+\xi)(1+2\xi)\xi^2\sigma}\times \\ &\quad \left\{-(1+\xi)^2+(1-2z)(1+\xi)(1+2\xi)+z(2+\xi)(1+2\xi)\right\} \end{aligned}$$

where $z = 1 + \xi T_u/\sigma$.

The notation $\mathbb{E}_{GPD}$ indicates that the expectation is taken with respect to the exact GPD in this lemma only.

Proof of Lemma 1 Denote

$$\begin{aligned} -l_i(\xi,\sigma) &= -D_i\log\left(1-G\left(T_u;\xi,\sigma\right)\right) - \left(1-D_i\right)\log g\left(Y_i;\xi,\sigma\right) \qquad (18) \\ &= \frac{D_i}{\xi}\log\left(1+\frac{\xi T_u}{\sigma}\right) + \left(1-D_i\right)\left(1+\frac{1}{\xi}\right)\log\left(1+\frac{\xi Y_i}{\sigma}\right) + \left(1-D_i\right)\log\sigma. \end{aligned}$$

Then by substituting $z = 1 + \xi T_u/\sigma$ and some elementary calculation, we have

$$-\frac{\partial l_i(\xi,\sigma)}{\partial\xi} = D_i\left\{-\frac{1}{\xi^2}\log z + \frac{1}{\xi^2}\left(1-z^{-1}\right)\right\}$$

$$+(1-D_i)\left\{-\frac{1}{\xi^2}\log\left(1+\frac{\xi}{\sigma}Y_i\right)+\frac{1}{\xi}\left(1+\frac{1}{\xi}\right)\left(1-\left(1+\frac{\xi}{\sigma}Y_i\right)^{-1}\right)\right\},$$

$$-\frac{\partial l_i(\xi,\sigma)}{\partial\sigma} = -D_i\left\{\frac{1}{\xi\sigma}\left(1-z^{-1}\right)\right\} + (1-D_i)\left\{-\frac{1}{\xi\sigma}+\frac{1}{\sigma}\left(1+\frac{1}{\xi}\right)\left(1+\frac{\xi}{\sigma}Y_i\right)^{-1}\right\},$$

$$-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\xi^2} = D_i\left\{\frac{2}{\xi^3}\log z-\frac{T_u^2}{\sigma^2\xi}z^{-2}-\frac{2T_u}{\sigma\xi^2}z^{-1}\right\} + (1-D_i)\left\{\frac{2}{\xi^3}\log\left(1+\frac{\xi Y_i}{\sigma}\right)-\frac{3+\xi}{\xi^3}+\frac{2(2+\xi)}{\xi^3}\left(1+\frac{\xi}{\sigma}Y_i\right)^{-1}-\frac{1+\xi}{\xi^3}\left(1+\frac{\xi Y_i}{\sigma}\right)^{-2}\right\},$$

$$\begin{aligned}-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\sigma^2} &= D_i\left\{-\frac{T_u^2\xi}{\sigma^4}z^{-2}+\frac{2T_u}{\sigma^3}z^{-1}\right\} + (1-D_i)\left\{\frac{1}{\xi\sigma^2}-\frac{1}{\sigma^2}\left(1+\frac{1}{\xi}\right)\left(1+\frac{\xi Y_i}{\sigma}\right)^{-2}\right\}\\ &= D_i\left\{\frac{1}{\sigma^2\xi}\left(1-z^{-2}\right)\right\}+(1-D_i)\left\{\frac{1}{\xi\sigma^2}-\frac{1}{\sigma^2}\left(1+\frac{1}{\xi}\right)\left(1+\frac{\xi Y_i}{\sigma}\right)^{-2}\right\},\end{aligned}$$

and

$$\begin{aligned}&-\frac{\partial^2 l_i(\xi,\sigma)}{\partial\sigma\partial\xi}\\ &= D_i\left\{\frac{T_u^2}{\sigma^3}\left(1+\frac{T_u\xi}{\sigma}\right)^{-2}\right\}+(1-D_i)\left\{\frac{1}{\sigma\xi^2}-\frac{2+\xi}{\sigma\xi^2}\left(1+\frac{Y_i\xi}{\sigma}\right)^{-1}+\frac{(1+\xi)}{\sigma\xi^2}\left(1+\frac{Y_i\xi}{\sigma}\right)^{-2}\right\}\\ &= D_i\left\{\frac{1}{\sigma\xi^2}\left(1-2z^{-1}+z^{-2}\right)\right\}+(1-D_i)\left\{\frac{1}{\sigma\xi^2}-\frac{2+\xi}{\sigma\xi^2}\left(1+\frac{Y_i\xi}{\sigma}\right)^{-1}+\frac{(1+\xi)}{\sigma\xi^2}\left(1+\frac{Y_i\xi}{\sigma}\right)^{-2}\right\}.\end{aligned}$$

Using the definition of GPD (2), we have that

$$\mathbb{E}_{GPD}\left[D_i|Y_i>u\right] = z^{-1/\xi} \tag{19}$$

$$\mathbb{E}_{GPD}\left[\left(1+\frac{\xi Y_i}{\sigma}\right)^{-r}\mathbf{1}\left[Y_i\le T\right]\middle| Y_i>u\right] = \frac{1-z^{-r-1/\xi}}{1+r\xi}\text{ for any } r>0 \tag{20}$$

$$\mathbb{E}_{GPD}\left[\log\left(1+\frac{\xi Y_i}{\sigma}\right)\mathbf{1}\left[Y_i\le T\right]\middle| Y_i>u\right] = \xi - z^{-1/\xi}\left(\xi+\log z\right). \tag{21}$$

Then using (19)-(21) to obtain that

$$\begin{aligned}
&\mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i\left(\xi,\sigma\right)}{\partial\xi^2}\middle| Y_i>u\right]\\
=\;& z^{-1/\xi}\left\{\frac{2}{\xi^3}\log z-\frac{3}{\xi^3}+\frac{4}{\xi^3}z^{-1}-\frac{1}{\xi^3}z^{-2}\right\}\\
&+\frac{2}{\xi^3}\left(\xi-z^{-1/\xi}\left(\xi+\log z\right)\right)-\frac{3+\xi}{\xi^3}(1-z^{-1/\xi})\\
&+\frac{2\left(2+\xi\right)}{\xi^3}\frac{1-z^{-1-1/\xi}}{1+\xi}-\frac{1+\xi}{\xi^3}\frac{1-z^{-2-1/\xi}}{1+2\xi}\\
=\;& \frac{2}{\left(1+\xi\right)\left(1+2\xi\right)}+\frac{z^{-2-1/\xi}}{(1+\xi)(1+2\xi)\xi^2}\times\\
&\left\{-1-\xi+z(2+4\xi)-z^2(1+\xi)(1+2\xi)\right\},
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i\left(\xi,\sigma\right)}{\partial\sigma^2}\middle| Y_i>u\right]\\
=\;& z^{-1/\xi}\left\{\frac{1}{\sigma^2\xi}\left(1-z^{-2}\right)\right\}+\mathbb{E}_{GPD}\left[(1-D_i)\left\{\frac{1}{\xi\sigma^2}-\frac{1}{\sigma^2}\left(1+\frac{1}{\xi}\right)\left(1+\frac{\xi Y_i}{\sigma}\right)^{-2}\right\}\right]\\
=\;& \frac{1}{\left(1+2\xi\right)\sigma^2}-\frac{z^{-2-1/\xi}}{\left(1+2\xi\right)\sigma^2},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{GPD}\left[-\frac{\partial^2 l_i\left(\xi,\sigma\right)}{\partial\sigma\partial\xi}\middle|\right] =\;& z^{-1/\xi}\left\{\frac{1}{\sigma\xi^2}\left(1-2z^{-1}+z^{-2}\right)\right\}\\
&+\left(1-z^{-1/\xi}\right)\frac{1}{\sigma\xi^2}-\frac{2+\xi}{\sigma\xi^2}\frac{1-z^{-1-1/\xi}}{1+\xi}+\frac{(1+\xi)}{\sigma\xi^2}\frac{1-z^{-2-1/\xi}}{1+2\xi}\\
=\;& \frac{1}{\left(1+\xi\right)\left(1+2\xi\right)\sigma}+\frac{z^{-2-1/\xi}}{\left(1+\xi\right)\left(1+2\xi\right)\xi^2\sigma}\times\\
&\left\{-\left(1+\xi\right)^2+\left(1-2z\right)\left(1+\xi\right)\left(1+2\xi\right)+z\left(2+\xi\right)\left(1+2\xi\right)\right\}.
\end{aligned}$$

This completes the proof. ■

Lemma 2 *Suppose Condition 2 holds. Then for any positive-valued integrable function $h(\cdot)$ on $(1,\infty)$ and any $c \in (1,\infty)$, there exists some function $\phi(u) \to 0$ as $u \to \infty$ such that*

$$\int_1^c h(t)\frac{L(tu)}{L(u)}dt = \int_1^c h(t)\,dt + \phi(u)\int_1^c h(t)\,k(t)\,dt + o(\phi(u)).$$

Proof of Lemma 2 The assumption on $L(\cdot)$ is sufficient for the SR2 assumption in Goldie and Smith (1987) and Smith (1987). Therefore, we have

$$L(tu)/L(u) = 1 + k(t)\phi(u) + o(\phi(u)),$$

where $k(t) = C\int_1^t u^{-\beta-1}du$ for some constant C and $\phi(u) = u^{-\beta}$ (pp.1179-1181 in Smith (1987)). Then it suffices to show that

$$\frac{h(t)\{L(tu)/L(u)-1\}}{\phi(u)}$$

is uniformly dominated by some integrable function of t as $u \to \infty$. This is done by noting that

$$\begin{aligned}\left|\frac{h(t)\{L(tu)/L(u)-1\}}{\phi(u)}\right| &\leq C\left|h(t)\left(1-t^{-\beta}\right)\right| \\ &\leq C\left|h(t)\right|\end{aligned}$$

for $t > 1$. ■

Proof of Proposition 1 Denote $S_k(\xi,\sigma)$ as the 2×1 vector with components $-(\partial/\partial\xi)\sum_{i=1}^k l_i(\xi,\sigma)$ and $-\sigma(\partial/\partial\sigma)\sum_{i=1}^k l_i(\xi,\sigma)$, where

$$l_i(\xi,\sigma) = D_i\log(1-G(T_u;\xi,\sigma)) + (1-D_i)\log g(Y_i;\xi,\sigma)$$

with $T_u = T-u$. Also denote

$$M_k(\xi,\sigma) = k^{-1}\begin{bmatrix} -\sum_{i=1}^k \frac{\partial l_i^2(\xi,\sigma)}{\partial\xi^2} & -\sigma\sum_{i=1}^k \frac{\partial l_i^2(\xi,\sigma)}{\partial\xi\partial\sigma} \\ -\sigma\sum_{i=1}^k \frac{\partial l_i^2(\xi,\sigma)}{\partial\xi\partial\sigma} & -\sigma^2\sum_{i=1}^k \frac{\partial^2 l_i(\xi,\sigma)}{\partial\sigma^2}\end{bmatrix}.$$

Then the intermediate value theorem yields that

$$k^{1/2}\begin{pmatrix}\hat{\xi}-\xi\\ \hat{\sigma}/\sigma-1\end{pmatrix}=M_k\left(\dot{\xi},\dot{\sigma}\right)^{-1}\left(k^{-1/2}S_k\left(\xi,\sigma\right)+o_p\left(1\right)\right)$$

for some intermediate values $\dot{\xi}$ and $\dot{\sigma}$. We next show that (i) $k^{-1/2}S_k(\xi,\sigma)$ converges to the normal random variable by using Lyapunov Central Limit Theorem (CLT) and (ii) $M_k(\tilde{\xi},\tilde{\sigma})$ uniformly converges to M over $(\tilde{\xi},\tilde{\sigma})$ in a shrinking neighborhood centered at (ξ,σ).

To show (i), we use Lemma 2 and integration by parts to obtain that, for any $r>0$

$$\begin{aligned}
&\mathbb{E}\left[\left(1+\frac{Y_i}{u}\right)^{-r}\mathbf{1}\left[Y_i\leq T\right]\middle|Y_i>u\right] \qquad (22)\\
=&\ -\int_0^{T-u}\left(1+\frac{y}{u}\right)^{-r}d\left(1-F_u\left(y\right)\right)\\
=&\ -\left(1+\frac{y}{u}\right)^{-r}\left(1-F_u\left(y\right)\right)\Big|_0^{T-u}-r\int_0^{T-u}\left(1-F_u\left(y\right)\right)\left(1+\frac{y}{u}\right)^{-r-1}u^{-1}dy\\
=&\ 1-\left(1+\frac{T-u}{u}\right)^{-r}\left(1-F_u\left(T-u\right)\right)\\
&-r\int_1^{\frac{T-u}{u}+1}t^{-r-1-\alpha}\frac{L\left(ut\right)}{L\left(u\right)}dt\\
=&\ 1-\left(T/u\right)^{-r-\alpha}+\frac{r}{r+\alpha}\left(\left(T/u\right)^{-r-\alpha}-1\right)+O\left(\phi\left(u\right)\right)\\
=&\ \frac{\alpha}{r+\alpha}\left(1-\left(T/u\right)^{-r-\alpha}\right)+O\left(\phi\left(u\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\log\left(1+\frac{Y_i}{u}\right)\mathbf{1}\left[Y_i\leq T\right]\middle|Y_i>u\right] \qquad (23)\\
=&\ -\int_0^{T-u}\log\left(1+\frac{y}{u}\right)d\left(1-F_u\left(y\right)\right)\\
=&\ -\log\left(1+\frac{y}{u}\right)\left(1-F_u\left(y\right)\right)\Big|_0^{T-u}+\int_0^{T-u}\left(1-F_u\left(y\right)\right)\left(1+\frac{y}{u}\right)^{-1}u^{-1}dy\\
=&\ -\log\left(1+\frac{T-u}{u}\right)\left(1-F_u\left(T-u\right)\right)+\int_1^{\frac{T-u}{u}+1}\frac{L\left(ut\right)}{L\left(u\right)}t^{-1-\alpha}dt\\
=&\ -\left(\log(T/u)\right)\left(T/u\right)^{-\alpha}+\frac{1}{\alpha}\left(1-\left(T/u\right)^{-\alpha}\right)+O\left(\phi\left(u\right)\right),
\end{aligned}$$

where recall $\phi(u) = u^{-\beta}$. Use the change of variable $\sigma = u/\alpha$ and recall $\alpha = 1/\xi$. Then by Lemma 1, we have $z = 1 + \frac{T-u}{\sigma}\xi = T/u$ and

$$\begin{aligned}
& \mathbb{E}\left[-\frac{\partial l_i(\xi,\sigma)}{\partial \xi} \middle| Y_i > u \right] \\
= \; & \mathbb{E}\left[D_i | Y_i > u\right] \left\{ -\frac{1}{\xi^2}\log z + \frac{1}{\xi^2}\left(1 - z^{-1}\right) \right\} \\
& + \mathbb{E}\left[\left\{ -\frac{1}{\xi^2}\log\left(1 + \frac{\xi}{\sigma}Y_i\right) + \frac{1}{\xi}\left(1 + \frac{1}{\xi}\right)\left(1 - \left(1 + \frac{\xi}{\sigma}Y_i\right)^{-1}\right) \right\} \mathbf{1}\left[Y_i \le T\right] \middle| Y_i > u \right] \\
= \; & \left(1 - F_u(T-u)\right) \left\{ -\frac{1}{\xi^2}\log(T/u) + \frac{1}{\xi^2}\left(1 - (T/u)^{-1}\right) \right\} \\
& - \frac{1}{\xi^2}\mathbb{E}\left[\log\left(1 + \frac{Y_i}{u}\right) \mathbf{1}\left[Y_i \le T\right] \middle| Y_i > u \right] \\
& + \mathbb{E}\left[\frac{1}{\xi}\left(1 + \frac{1}{\xi}\right)\left(1 - \left(1 + \frac{Y_i}{u}\right)^{-1}\right) \mathbf{1}\left[Y_i \le T\right] \middle| Y_i > u \right] \\
= \; & (T/u)^{-\alpha}\left\{ -\alpha^2 \log(T/u) + \alpha^2\left(1 - (T/u)^{-1}\right) \right\} - \alpha^2\left(-\left(\log(T/u)\right)(T/u)^{-\alpha} + \frac{1}{\alpha}\left(1 - (T/u)^{-\alpha}\right) \right) \\
& + \alpha(1+\alpha)\left(1 - (T/u)^{-\alpha}\right) - \alpha^2\left(1 - (T/u)^{-1-\alpha}\right) + O(\phi(u)) \\
= \; & O(\phi(u)).
\end{aligned}$$

Similarly by Condition 3 and repetitively using (22) and (23), we have that

$$\begin{aligned}
\mathbb{E}\left[-\sigma\frac{\partial l_i(\xi,\sigma)}{\partial \sigma} \middle| Y_i > u \right] &= O(\phi(u)) \\
\mathbb{E}\left[\left(\frac{\partial l_i(\xi,\sigma)}{\partial \xi}\right)^2 \middle| Y_i > u \right] &= M_{11} + O(\phi(u)) + o(1) \\
\mathbb{E}\left[\sigma^2\left(\frac{\partial l_i(\xi,\sigma)}{\partial \sigma^2}\right)^2 \middle| Y_i > u \right] &= M_{22} + O(\phi(u)) + o(1) \\
\mathbb{E}\left[\sigma\frac{\partial l_i(\xi,\sigma)}{\partial \sigma}\frac{\partial l_i(\xi,\sigma)}{\partial \xi} \middle| Y_i > u \right] &= M_{12} + O(\phi(u)) + o(1),
\end{aligned}$$

and the third moments conditional on $Y_i > u$ of $\partial l_i(\xi,\sigma)/\partial\xi$ and $\sigma\partial l_i(\xi,\sigma)/\partial\sigma$ are also bounded as $u \to \infty$. Then by Lyapunov CLT, we have $k^{-1/2}S_k(\xi,\sigma) \xrightarrow{d} \mathcal{N}(0, M)$.

Now it remains to show that $M_k(\tilde{\xi},\tilde{\sigma})$ uniformly converges to M in the neighborhood that $\{(\tilde{\sigma},\tilde{\xi}) : |\tilde{\sigma}/\sigma - 1| \le \varepsilon_k \text{ and } |\tilde{\xi} - \xi| \le \varepsilon_k\}$ for some $\varepsilon_k = o(1)$ satisfying $k^{1/2}\varepsilon_k \to \infty$.

To this end, it suffices to show that $\mathbb{E}[|\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\xi^3||Y_i>u]$, $\mathbb{E}[|\sigma\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\xi^2\partial\sigma||Y_i>u]$, $\mathbb{E}[|\sigma^3\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\sigma^3||Y_i>u]$, and $\mathbb{E}[\sigma^2|\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\sigma^2\partial\xi||Y_i>u]$ are all uniformly bounded over this neighborhood. This is done by straightforward calculations as we show in Lemma 3. For brevity, we present the proof for $\mathbb{E}[|\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\xi^3||Y_i>u]$ only since the argument for the other terms are similar (cf. pp.1178-1180 in Smith (1987)). ■

Lemma 3 $\mathbb{E}[|\partial^3 l_i(\tilde{\xi},\tilde{\sigma})/\partial\xi^3||Y_i>u]$ *is uniformly bounded over* $\{(\tilde{\sigma},\tilde{\xi}):|\tilde{\sigma}/\sigma-1|\leq\varepsilon_k$ *and* $|\tilde{\xi}-\xi|\leq\varepsilon_k\}$ *for any* $\varepsilon_k\to 0$ *as* $k\to\infty$.

Proof of Lemma 3 Substituting (18) to obtain that

$$\begin{aligned}
&\frac{\partial^3 l_i\left(\tilde{\xi},\tilde{\sigma}\right)}{\partial\xi^3}\\
=\;& D_i\left\{\begin{array}{l}\frac{2T_u^3}{\tilde{\sigma}^3\tilde{\xi}}\left(1+\frac{T_u\tilde{\xi}}{\tilde{\sigma}}\right)^{-3}+\frac{3T_u^2}{\tilde{\sigma}^2\tilde{\xi}^2}\left(1+\frac{T_u\tilde{\xi}}{\tilde{\sigma}}\right)^{-2}\\ +\frac{6T_u}{\tilde{\sigma}\tilde{\xi}^3}\left(1+\frac{T_u\tilde{\xi}}{\tilde{\sigma}}\right)^{-1}-\frac{6}{\tilde{\xi}^4}\log\left(1+\frac{T_u\tilde{\xi}}{\tilde{\sigma}}\right)\end{array}\right\}\\
&+(1-D_i)\left\{\begin{array}{l}\frac{11}{\tilde{\xi}^4}+\frac{2}{\tilde{\xi}^3}-\frac{2}{\tilde{\xi}^4}\left(1+\frac{\tilde{\xi}}{\tilde{\sigma}}Y_i\right)^{-3}+\left(\frac{9}{\tilde{\xi}^4}+\frac{6}{\tilde{\xi}^3}\right)\left(1+\frac{\tilde{\xi}}{\tilde{\sigma}}Y_i\right)^{-2}\\ -\left(\frac{18}{\tilde{\xi}^4}+\frac{6}{\tilde{\xi}^3}\right)\left(1+\frac{\tilde{\xi}}{\tilde{\sigma}}Y_i\right)^{-1}-\frac{6}{\tilde{\xi}^4}\log\left(1+\frac{\tilde{\xi}Y_i}{\tilde{\sigma}}\right)\end{array}\right\}\\
\equiv\;& B_{1n}(\tilde{\xi},\tilde{\sigma})+B_{2n}(\tilde{\xi},\tilde{\sigma}).
\end{aligned}$$

Since $T_u=T-u=O(\sigma)$ and $\mathbb{E}\left[D_i|Y_i>u\right]<1$, the expectation of B_{1n} is then uniformly bounded over $\{(\tilde{\sigma},\tilde{\xi}):|\tilde{\sigma}/\sigma-1|\leq\varepsilon_k$ and $|\tilde{\xi}-\xi|\leq\varepsilon_k\}$. To bound B_{2n}, it suffices to show that for any $\tilde{\xi}$ and $\tilde{\sigma}$ in the set $\{(\tilde{\sigma},\tilde{\xi}):|\tilde{\sigma}/\sigma-1|\leq\varepsilon_k$ and $|\tilde{\xi}-\xi|\leq\varepsilon_k\}$ and for any $r>0$,

$$\mathbb{E}\left[\left(1+\frac{\tilde{\xi}}{\tilde{\sigma}}Y_i\right)^{-r}\mathbf{1}\left[Y_i\leq T\right]\middle|Y_i>u\right]<\infty \text{ and} \tag{24}$$

$$\mathbb{E}\left[\log\left(1+\frac{\tilde{\xi}Y_i}{\tilde{\sigma}}\right)\mathbf{1}\left[Y_i\leq T\right]\middle|Y_i>u\right]<\infty, \tag{25}$$

which are similar to (22) and (23), respectively. In particular, denote $v=(\xi\tilde{\sigma})/(\tilde{\xi}\sigma)$ so that $|v-1|\leq\varepsilon$ for some constant $\varepsilon\to 0$. Then using the change of variable $\sigma=u/\alpha=u\xi$, we have

$$
\begin{aligned}
& \mathbb{E}\left[\left(1+\frac{\tilde{\xi}}{\tilde{\sigma}}Y_i\right)^{-r}\mathbf{1}\left[Y_i \leq T\right]\middle| Y_i > u\right] \\
= & \mathbb{E}\left[\left(1+\frac{Y_i}{vu}\right)^{-r}\mathbf{1}\left[Y_i \leq T\right]\middle| Y_i > u\right] \\
= & -\int_0^{T-u}\left(1+\frac{y}{vu}\right)^{-r} d\left(1-F_u\left(y\right)\right) \\
= & -\left(1+\frac{y}{vu}\right)^{-r}\left(1-F_u\left(y\right)\right)\Big|_0^{T-u} - vr\int_0^{T-u}\left(1-F_u\left(y\right)\right)\left(1+\frac{y}{vu}\right)^{-r-1}u^{-1}dy \\
= & 1-\left(1+\frac{T-u}{vu}\right)^{-r}\left(1-F_u\left(T-u\right)\right) \\
& -vr\int_1^{\frac{T-u}{u}+1}\left(1+\frac{t-1}{v}\right)^{-r-1}t^{-\alpha}\frac{L\left(ut\right)}{L\left(u\right)}dt.
\end{aligned}
$$

Condition 3 and Lemma 2 yield that the above item is bounded. A very similar argument applies to (25), which completes the proof. ■

Proof of Proposition 2 The proof follows analogously from Theorem 4.3.1 and Remark 4.3.7 in de Haan and Ferreira (2007). We now provide the details. Recall $d_n = (m+k)/(np_n)$ and write $u = \hat{Q}\left(1-p_n d_n\right)$, which is $Y_{(m+k)}$. Then we decompose $\hat{Q}\left(1-p_n\right) - Q(1-p_n)$ as

$$
\sqrt{k}\frac{\hat{Q}\left(1-p_n\right)-Q(1-p_n)}{\sigma q_{\xi}(d_n)} = C_{1n}+C_{2n}+C_{3n}-C_{4n},
$$

where

$$
\begin{aligned}
C_{1n} &= \sqrt{k}\frac{\hat{Q}\left(1-p_n d_n\right)-Q(1-p_n d_n)}{\sigma}\frac{1}{q_{\xi}(d_n)} \\
C_{2n} &= \frac{\hat{\sigma}}{\sigma}\left\{\frac{\sqrt{k}}{q_{\xi}(d_n)}\left(\frac{d_n^{-\hat{\xi}}-1}{\hat{\xi}}-\frac{d_n^{-\xi}-1}{\xi}\right)\right\} \\
C_{3n} &= \sqrt{k}\left(\frac{\hat{\sigma}}{\sigma}-1\right)\frac{d_n^{-\xi}-1}{\xi q_{\xi}(d_n)} \\
C_{4n} &= \frac{\sqrt{k}}{q_{\xi}(d_n)}\left(\frac{Q(1-p_n)-Q(1-p_n d_n)}{\sigma}-\frac{d_n^{-\xi}-1}{\xi}\right).
\end{aligned}
$$

We next derive the limits of C_{jn} for $j = 1, 2, 3, 4$. To this end, we define $U(t) = Q(1 - 1/t)$ and denote $U'(t) = \partial U(t)/\partial t$. We introduce the second-order tail approximation that

$$\lim_{t\to\infty} \frac{\frac{U(ty)-U(t)}{a(t)} - \frac{y^{\xi}-1}{\xi}}{A(t)} = H(y) \tag{26}$$

as in Theorem 2.3.12 in de Haan and Ferreira (2007). Condition 2 implies that $a(t) = tU'(t) = 1/\left(tf(Q(1-1/t))\right)$, $A(t) \propto t^{-\beta/\alpha}$, and $H(y) = -y^{\xi}(y^{-\beta} - 1)/\beta$.

For C_{1n}, substitute $t = 1/\left(p_n d_n\right)$ and use Theorem 2.4.1 in de Haan and Ferreira (2007) to obtain that

$$\sqrt{k}\frac{\left(\hat{Q}\left(1 - p_n d_n\right) - Q(1 - p_n d_n)\right)}{a_n(1/(p_n d_n))} \xrightarrow{d} \mathcal{N}\left(0, 1\right).$$

Then by Theorem 1.1.6 in de Haan and Ferreira (2007), we have that σ is asymptotically equivalent to $a(1/(p_n d_n))$ as $n \to \infty$, which further implies that $C_{1n} \xrightarrow{d} \mathcal{N}\left(0, q_{\xi}(d_0)^{-2}\right)$.

For C_{2n}, the same argument as part II on pp.136-137 in de Haan and Ferreira (2007) yields that $C_{2n} = k^{1/2}\left(\hat{\xi} - \xi\right) + o_p(1)$. For C_{3n}, by Proposition 1, we have that

$$C_{3n} = k^{1/2}\left(\frac{\hat{\sigma}}{\sigma} - 1\right)\left(\frac{d_0^{-\xi} - 1}{\xi q_{\xi}(d_0)}\right) + o_p(1),$$

where recall $d_0 = \lim_{n\to\infty} d_n > 0$. Note that given $u = Y_{(m+k)}$, the excesses $\{Y_{(m+i)} - Y_{(m+k)}\}_{i=1}^{k-1}$ are asymptotically independent from $Y_{(m+k)}$ (cf. p.1185 in Drees, Ferreira, and de Haan (2004)), and therefore C_{1n} is asymptotically independent from C_{2n} and C_{3n} (see also pp.1180-1181 in Smith (1987)).

Finally, (26) and Condition 4 yield that $\sqrt{k}A(n/(k+m)) = o(1)$ and hence

$$\begin{aligned} C_{4n} &= \sqrt{k}A(n/(k+m))\frac{d_n^{-\xi} - 1}{\xi q_{\xi}(d_n)} \frac{\left(\frac{U(1/p_n) - U(1/(p_n d_n))}{a(n/(m+k))}\frac{\xi}{d_n^{-\xi}-1} - 1\right)}{A(n/(k+m))} \\ &= o(1). \end{aligned}$$

The proof is complete by combining C_{jn} for $j = 1, 2, 3, 4$. ■

Proof of Proposition 3 We prove this by induction. By standard EVT, for any fixed positive integer I,

$$f_{X_1,\ldots,X_I|\xi}(x_1,\ldots,x_I) = V_\xi\left(x_I\right)\prod_{i=1}^{I} v_\xi\left(x_i\right)/V_\xi\left(x_i\right). \tag{27}$$

Consider $m=1$ first. For any fixed positive integer k, (27) with $I=k+1$ implies that

$$\begin{aligned} f_{X_{m+1},\ldots,X_{m+k}|\xi}(x_{m+1},\ldots,x_{m+k}) &= f_{X_2,\ldots,X_{k+1}|\xi}(x_2,\ldots,x_{k+1}) \\ &= \left(\int_{x_2}^{\infty}\frac{v_\xi\left(x_1\right)}{V_\xi\left(x_1\right)}dx_1\right)V_\xi\left(x_{k+1}\right)\prod_{i=2}^{k+1} v_\xi\left(x_i\right)/V_\xi\left(x_i\right) \\ &= -\log V_\xi\left(x_2\right)V_\xi\left(x_{k+1}\right)\prod_{i=2}^{k+1} v_\xi\left(x_i\right)/V_\xi\left(x_i\right), \end{aligned}$$

which satisfies (9).

Now assume (9) holds for some fixed positive integer $m \geq 1$. This implies that for any k,

$$\begin{aligned} & f_{X_{m+2},\ldots,X_{m+1+k}|\xi}(x_{m+2},\ldots,x_{m+k+1}) \\ = & \int_{x_{m+2}}^{\infty} f_{X_{m+1},\ldots,X_{m+k+1}}(x_{m+1},\ldots,x_{m+k+1})dx_{m+1} \\ = & \left(\int_{x_{m+2}}^{\infty}\frac{1}{m!}\left(-\log V_\xi\left(x_{m+1}\right)\right)^m\frac{v_\xi(x_{m+1})}{V_\xi(x_{m+1})}dx_{m+1}\right)V_\xi\left(x_{m+k+1}\right)\prod_{i=m+2}^{m+k+1} v_\xi\left(x_i\right)/V_\xi\left(x_i\right) \\ = & \left(\int_{\log G_\xi(x_{m+2})}^{0}\left(-v\right)^m dv\right)\frac{1}{m!}V_\xi\left(x_{m+k+1}\right)\prod_{i=m+2}^{m+k+1} v_\xi\left(x_i\right)/V_\xi\left(x_i\right) \\ = & \frac{1}{(m+1)!}\left(-\log V_\xi\left(x_{m+2}\right)\right)^{m+1}V_\xi\left(x_{m+k+1}\right)\prod_{i=m+2}^{m+k+1} v_\xi\left(x_i\right)/V_\xi\left(x_i\right), \end{aligned}$$

which means that (9) holds for $m+1$. This completes the proof. ■

A.3 Additional Empirical Results in CPS Data

Tables 8 and 9 depict the results based on $k=[0.04n]$ and $[0.06n]$, respectively.

Table 8: Empirical Results Using 2019 March CPS Data

Panel A: 95% confidence intervals with race-based subsample									
	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
Full Sample	115424	672	0.58	(0.37	0.49)	(24.26	25.32)	(59.88	73.00)
Male	55553	491	0.88	(0.39	0.62)	(28.76	30.97)	(71.86	106.3)
Female	59871	181	0.30	(0.30	0.44)	(18.31	19.33)	(42.46	52.03)
Male White	43371	419	0.97	(0.09	0.34)	(29.57	31.84)	(54.54	75.61)
Female White	45424	141	0.31	(0.29	0.45)	(18.40	19.59)	(42.36	53.68)
Male Asian	3676	50	1.36	(0.00	0.55)	(30.73	37.77)	(48.30	109.6)
Female Asian	4099	22	0.54	(0.13	0.76)	(21.27	26.66)	(42.89	104.5)
Male Hispanic	44420	445	1.00	(0.11	0.36)	(29.98	32.27)	(55.56	77.78)
Female Hispanic	48192	155	0.32	(0.38	0.55)	(18.83	19.99)	(45.77	59.36)
Male Black	6144	12	0.20	(0.16	0.53)	(16.06	18.55)	(29.70	47.84)
Female Black	7827	9	0.16	(0.12	0.44)	(13.94	15.72)	(25.08	36.81)
Panel B: 95% confidence intervals with age-based subsample									
Age	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
18-30	27829	35	0.13	(0.34	0.53)	(12.63	13.54)	(28.42	37.19)
30-40	25213	158	0.63	(0.24	0.51)	(24.15	26.40)	(50.57	75.56)
40-50	23419	213	0.91	(0.00	0.32)	(28.67	31.48)	(48.07	71.41)
50-60	21767	196	0.90	(0.63	1.00)	(28.30	32.89)	(86.35	218.9)
60-65	17196	70	0.41	(0.43	0.76)	(20.29	22.63)	(49.71	88.34)

Note: Entries are the sample size (n), the number of censored observations (cen#), the censored fraction in percentage points (cen%), 95% confidence intervals of the tail index and those of the 99% and 99.9% quantiles measured in 10^4 USD. The results are based on $k = [0.04n]$ and the variable ERN_VAL in the CPS dataset (and equivalently the variable inclongj from the IPUMS dataset). Data are available at https://cps.ipums.org/cps.

Table 9: Empirical Results Using 2019 March CPS Data

Panel A: 95% confidence intervals in race-based subsamples									
	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
Full Sample	115424	672	0.58	(0.31	0.40)	(24.33	25.34)	(56.01	65.09)
Male	55553	491	0.88	(0.30	0.45)	(28.59	30.52)	(63.88	82.90)
Female	59871	181	0.30	(0.43	0.54)	(18.03	19.05)	(46.57	57.68)
Male White	43371	419	0.97	(0.32	0.50)	(29.55	31.96)	(67.36	93.12)
Female White	45424	141	0.31	(0.43	0.57)	(18.09	19.30)	(47.07	60.62)
Male Asian	3676	50	1.36	(0.00	0.51)	(30.59	38.09)	(48.25	102.7)
Female Asian	4099	22	0.54	(0.16	0.61)	(21.27	26.54)	(41.85	87.70)
Male Hispanic	44420	445	1.00	(0.43	0.61)	(30.25	32.94)	(78.16	113.6)
Female Hispanic	48192	155	0.32	(0.38	0.51)	(18.85	20.01)	(45.79	57.46)
Male Black	6144	12	0.20	(0.19	0.50)	(16.05	18.31)	(28.61	46.11)
Female Black	7827	9	0.16	(0.20	0.47)	(13.83	15.56)	(25.73	38.47)
Panel B: 95% confidence intervals in age-based subsamples									
Age	n	cen#	cen%	tail index		Q(0.99)		Q(0.999)	
18-30	27829	35	0.13	(0.30	0.44)	(12.78	13.68)	(27.65	34.78)
30-40	25213	158	0.63	(0.57	0.79)	(23.95	26.70)	(71.84	118.8)
40-50	23419	213	0.91	(0.27	0.50)	(28.59	31.63)	(59.61	90.30)
50-60	21767	196	0.90	(0.35	0.59)	(28.12	31.56)	(65.11	105.7)
60-65	17196	70	0.41	(0.40	0.64)	(20.54	22.99)	(50.78	80.72)

Note: Entries are the sample size (n), the number of censored observations (cen#), the censored fraction in percentage points (cen%), 95% confidence intervals of the tail index and those of the 99% and 99.9% quantiles measured in 10^4 USD. The results are based on $k = [0.06n]$ and the variable ERN_VAL in the CPS dataset (and equivalently the variable inclongj from the IPUMS dataset). Data are available at https://cps.ipums.org/cps.

References

Aban, I. B., M. M. Meerschaert, and A. K. Panorska (2006): "Parameter Estimation for the Truncated Pareto Distribution," *Journal of the American Statistical Association*, 101(473), 270–277.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Armour, P., R. V. Burkhauser, and J. Larrimore (2013): "Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income," *American Economic Review*, 103(3), 173–177.

Balkema, A. A., and L. de Haan (1974): "Residual Life Time at Great Age," *The Annals of Probability*, 2, 792–804.

Barro, R. J., and T. Jin (2011): "On the Size Distribution of Macroeconomic Disasters," *Econometrica*, 79(5), 1567–1589.

Barro, R. J., and J. F. Ursúa (2008): "Macroeconomic Crisis Since 1870," *Brookings Papers on Economic Activity*, (255-335).

Beirlant, J., I. F. Alves, and M. I. Gomes (2016): "Tail Fitting for Truncated and non-truncated Pareto-type Distributions," *Extremes*, 19(3), 429–462.

Beirlant, J., I. F. Alves, and T. Reynkens (2017): "Fitting tails affected by Truncation," *Electronic Journal of Statistics*, 11(1), 2026–2065.

Brzezinski, M. (2013): "Asymptotic and Bootstrap Inference for Top Income Shares," *Economics Letters*, 120(1), 10–13.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2010): "Estimating Trends in US Income Inequality Using the Current Population Survey: the Importance of ContControl for Censoring," *The Journal of Economic Inequality*, 9(3), 393–415.

Chernozhukov, V. (2005): "Extremal Quantile Autoregression," *The Annals of Statistics*, 33(2), 806–839.

Chernozhukov, V., and I. Fernández-Val (2011): "Inference for extremal conditional quantile models, with an application to market and birthweight Risks," *The Review of Economic Studies*, 78, 559–589.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009): "Power-law Distributions in Empirical Data," *SIAM Review*, 51(4), 661–703.

de Haan, L., and A. Ferreira (2007): *Extreme Value Theory: An Introduction.* Springer Science and Business Media, New York.

Drees, H. (2001): "Minimax Risk Bounds in Extreme Value Theory," *Annuals of Statistics*, 29(1), 266–294.

Drees, H., A. Ferreira, and L. de Haan (2004): "On Maximum Likelihood Estimation of the Extreme Value Index," *The Annals of Applied Probability*, 14(3), 1179–1201.

Eika, L., M. Mogstad, and B. Zafar (2019): "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 127(6), 2795–2835.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present under the Null Hypothesis," *Econometrica*, 83, 771–811.

Gabaix, X., and R. Ibragimov (2011): "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, 29(1), 24–39.

Giesen, K., A. Zimmermann, and J. Suedekum (2010): "The Size Distribution Across All Cities-Double Pareto Lognormal Strikes," *Journal of Urban Economics*, 68(2), 129–137.

Goldie, C. M., and R. L. Smith (1987): "Slow Variation with Remainder: Theory and Applications," *the Quarterly Journal of Mathematics*, 38, 45–71.

Hall, P. (1982): "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of Royal Statistic Society, Series B*, 44(1), 37–42.

Hill, B. M. (1975): "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–1174.

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Ecomometrica*, 71(3), 905–932.

Jenkins, S. P., R. V. Burkhauser, S. Feng, and J. Larrimore (2010): "Measuring Inequality Using Censored Data: A Multiple-imputation Approach to Estimation and Inference," *Journal of the Royal Statistical Society Series A*, 174(1), 63–81.

Leadbetter, M. R. (1983): "Extremes and local dependence in stationary sequences," *Probability Theory and Related Fields*, 65(2), 291–306.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypothesis*. Springer, New York.

Mikosch, T., and C. Stărică (2000): "Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process," *Annals of Statistics*, 24, 1427–1451.

Müller, U. K., and Y. Wang (2017): "Fixed-k Asymptotic Inference about Tail Properties," *the Journal of the American Statistical Association*, 112, 1134–1143.

——— (2019): "Nearly Weighted Risk Minimal Unbiased Estimation," *Journal of Econometrics*, 209, 18–34.

O'Brien, G. (1987): "Extreme Values for Stationary and Markov Sequences," *Annals of Probability*, 15(1), 281–291.

Pickands, III, J. (1975): "Statistical inference using extreme order statistics," *Annals of Statistics*, 3(1), 119–131.

Portnoy, S. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98(464), 1001–1012.

Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32(1), 143–155.

Reed, W. (2003): "The Pareto Law of Incomes-an Explanation and an Extension," *Physica A*, 319(1), 469–486.

Smith, R. L. (1987): "Estimating Tails of Probability Distributions," *Annals of Statistics*, 15, 1174–1207.

Toda, A. A. (2012): "The Double Power Law in Income Distribution: Explanations and Evidence," *Journal of Economic Behavior and Organization*, 84(1), 364–381.

——— (2017): "A Note on the Size Distribution of Consumption: More Double Pareto than Lognormal," *Macroeconomic Dynamics*, 21(6), 1508–1518.

Toda, A. A., and Y. Wang (2019): "Efficient Minimum Distance Estimation of Pareto Exponent from Top Income Shares," *Working Paper*.