Tilburg University

Unified Extreme Value Estimation for Heterogeneous Data

Einmahl, John; He, Y.

Publication date:
2020

Document Version
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Einmahl, J., & He, Y. (2020). *Unified Extreme Value Estimation for Heterogeneous Data*. (CentER Discussion Paper; Vol. 2020-025). CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2020-025

**UNIFIED EXTREME VALUE ESTIMATION FOR
HETEROGENEOUS DATA**

By

John H.J. Einmahl, Yi He

29 September 2020

ISSN 0924-7815
ISSN 2213-9532

Unified extreme value estimation for heterogeneous data

John H.J. Einmahl*

Yi He

Tilburg University

University of Amsterdam

September 25, 2020

Abstract

We develop a universal econometric formulation of the empirical power laws possibly driven by parameter heterogeneity. Our approach extends classical extreme value theory to specifying the behavior of the empirical distribution of a general data set with possibly heterogeneous marginal distributions and a complex dependence structure. The main assumption is that in the intermediate tail the empirical distribution approaches some heavy-tailed distribution with a positive extreme value index. In this setup the Hill estimator consistently estimates this extreme value index and, on a log-scale, extreme quantiles are consistently estimated. We discuss several model examples that satisfy our conditions and demonstrate in simulations how heterogeneity may generate the dynamics of empirical power laws. We observe a dynamic cross-sectional power law for the new confirmed COVID-19 cases and deaths per million people across countries, and show that this international inequality is largely driven by the heterogeneity of the countries' scale parameters.

Keywords: Power law; Extreme values; Heterogeneous data; COVID-19; Inequality

JEL Codes: C14, C21, I19

1 Introduction

The empirical power law is a fundamental observation for many economic data sets: the tail probability seems to decrease at a polynomial rate rather than an exponential rate. It is also called the heavy tail phenomenon since large values occur more frequently than those from, for example, the (log)normal distribution in such a way that high-order sample

*John Einmahl holds the Arie Kapteyn Chair 2019–2022 and gratefully acknowledges the corresponding research support.

moments may diverge with the sample size. At least since [Pareto \(1896\)](#) and [Zipf \(1949\)](#), the empirical power law has been well documented for income and wealth; see, e.g., [Piketty and Saez \(2003\)](#) for the US and France, [Atkinson \(2005\)](#) for the UK, the excellent book [Piketty and Goldhammer \(2014\)](#) for a general overview, and the recent survey [Zucman \(2019\)](#) for evidence across many countries. The heavy tail phenomenon extends enormously to various contexts, including city and firm size (e.g., [Stanley et al., 1995](#), [Gabaix, 1999](#), [Axtell, 2001](#), [Gabaix and Ioannides, 2004](#)), consumption ([Toda and Walsh, 2015](#)), macroeconomic disasters ([Barro and Jin, 2011](#)), social and economic networking (e.g., [Barabási and Albert, 1999](#), [Jackson, 2009](#), [Newman et al., 2006](#)), CEO pay ([Gabaix and Landier, 2008](#)), finance (e.g., [Mandelbrot, 1963](#), [Fama, 1963](#), [Jansen and de Vries, 1991](#), [Longin, 1996](#), [Gopikrishnan et al., 1999](#), [Gopikrishnan et al., 2000](#), [Gabaix et al., 2006](#), [Plerou and Stanley, 2007](#), [Kyle and Obizhaeva, 2016](#)), international trade ([di Giovanni et al., 2011](#); [di Giovanni and Levchenko, 2012](#)), and many others. For more examples in economics and finance we refer to the excellent surveys [Gabaix \(2009\)](#) and [Gabaix \(2016\)](#). For many examples beyond economics we refer to, e.g., the survey [Clauset et al. \(2009\)](#).

The larger observations are often depicted in a log-log plot: plot on logarithmic axes the data ranks in descending order as a function of the data values. In [Figure 1](#) we observe a linear pattern in these plots that identifies empirical power laws for four data examples:

- (1) Census 2010 resident population in millions for the largest 100 US cities collected from the online database of U.S. Census Bureau, Population Division;
- (2) Number of friends of the most connected 3000 users in a Facebook network collected from the Koblenz Network Collection (KONECT);
- (3) The largest 50 monthly stock returns in March 2020 on the largest 1000 US companies with share codes 10 and 11 collected from the Center for Research in Security Prices (CRSP);
- (4) The wealth of top 100 billionaires among the Forbes 400 richest Americans in 2019.

Denote an entire dataset by $(X_1^{(p)}, \dots, X_p^{(p)})$ where the data dimension p is large, and define its empirical distribution function (df) by

$$F_{emp}(x) = \frac{1}{p} \sum_{i=1}^p \mathbb{1}(X_i^{(p)} \leq x). \quad (1.1)$$

Figure 1: Empirical power laws in log-log plots. The axes are logarithmic.

The fitted straight line in each log-log plot supports the empirical power law, that is,

$$1 - F_{emp}(x) \text{ is approximately } x^{-1/\hat{\gamma}} \text{ for large } x, \quad (1.2)$$

where $-1/\hat{\gamma}$ is the slope of the fitted line and $\hat{\gamma}$ is the estimated extreme value index. The larger the estimated extreme value index $\hat{\gamma}$, the heavier the tail.

In contrast to the increasing complexity of the economic generating mechanisms, the explanation of the power law is remarkably straightforward from a probabilistic point of view. The seminal works by [Fisher and Tippett \(1928\)](#) and [Gnedenko \(1943\)](#) show that the asymptotic distribution of the maximum of a random sample, up to proper standardization, can only take a specific form depending on a single parameter γ , called the extreme value index (EVI). An equivalent formulation of these results by [Balkema and de Haan \(1974\)](#) and [Pickands \(1975\)](#) shows that the conditional excess distribution beyond a sufficiently large threshold can only approach a generalized Pareto law. The power law immediately follows from the limiting generalized Pareto distribution in case the EVI γ is positive. Historically,

these two formulations of the power law led to two different statistical characterizations of empirical power laws, the so-called block-maxima method and peaks-over-threshold method, that model the maxima over subsamples and all large observations in the full sample, respectively. Our log-log plot in Figure 1 uses the latter one. For excellent accounts of extreme value theory we refer to the books Embrechts et al. (1997), de Haan and Ferreira (2006) and Resnick (2007). For similar results under weak temporal dependence see, e.g., Leadbetter et al. (1983).

A caveat of the aforementioned classical extreme value theory is that it works with homogeneous, that is equally distributed data. In fact, Hüsler (1986) shows that the standard formulation does not generalize to arbitrary non-stationary data. This immediately raises questions to econometricians who are dealing with power laws of heterogeneous economic data at a given point in time or dynamic cross-sectional data over time; see, e.g., Allen et al. (2012), Kelly and Jiang (2014), and Gabaix et al. (2016). For all practical purposes, however, economists tend to disregard the statistical complexity by directly working with data from a common distribution. This convenience comes at an obvious cost of neglecting the micro heterogeneity, which is then absorbed in the common hypothetical macro distribution. One possible endorsement for such an approach approach was given recently in Einmahl et al. (2016), which shows that the estimated EVI $\hat{\gamma}$ is still consistent for the common EVI γ for heavy-tailed data if they only differ by scale in the tail. In contrast, as we show later on, in general the heterogeneity effect is not only non-negligible, but also can generate a positive γ , that is, a heavy tail, even for light-tailed Gaussian data.

The motivation of this paper is to formalize the econometric theory for the aforementioned practical approach adopted by economists. In Section 2, we introduce Chang’s condition, see Chang (1964) and also Wellner (1978), that states that in the intermediate tail the empirical df F_{emp} of possibly heterogeneous data approaches some limiting df F . Note that Chang’s condition requires no specific form of the dependence structure nor tail homogeneity of the data. We then show that the estimated EVI converges to the EVI of the limiting distribution under a weak stability condition. Most importantly, now we can interpret the empirical power law as a finite sample approximation if the limiting distribution is heavy-tailed.

To illustrate how the micro heterogeneity influences the limiting df, we derive this limiting df and verify Chang’s condition for various interesting models in Section 3. These models reveal that, as noted above, the heavy tail may be directly generated by the parameter heterogeneity of the data set even if each individual observation comes from a light-tailed

distribution.

We then demonstrate how heterogeneity may generate the dynamics of empirical power laws by simulations in Section 4. Our data generating processes are calibrations based on the US income inequality data and financial data. Our model can replicate the rapid rise in US income inequality, which is driven by the scale heterogeneity of the cross-sectional data. Heterogeneity is a fundamental property of individual incomes (see, e.g., [Fagereng et al., 2020](#)), and our analysis may provide new insights into the dynamics of inequality ([Gabaix et al., 2016](#)). Furthermore, our model also suggests that pooling heterogeneous EGARCH data over time may significantly reduce the variability of the EVI estimator but not so much the estimation bias.

Finally, in Section 5, we study the COVID-19 infection inequality across countries. We document a persistent cross-sectional power law across the new confirmed COVID-19 cases and deaths per million people across countries. We show that the inequality is largely driven by the heterogeneity of the country-specific scale parameters.

The proofs of the results in Sections 2 and 3 are deferred to Section 6.

2 Unified Estimation Theory for Empirical Power Laws

Throughout we denote our data vector by $X = (X_1^{(p)}, \dots, X_p^{(p)})$, where the dimension p is large. In general, X is an arbitrary high-dimensional random vector with unspecified dependence structure and possibly heterogeneous tail distributions. Typically, these data are cross-sectional observations given at a point in time or a pool of possibly imbalanced panel data over a certain time period. Let the empirical df be as defined in (1.1). For any probability df F on $\mathbb{R}$, we define its generalized quantile function by $Q(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$.

Our theory relies on the following assumption, which is established in [Chang \(1964\)](#) for i.i.d. data.

Assumption 1 (Chang’s condition). In the intermediate tail the empirical df F_{emp} approaches some increasing df F in such a way that

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{1 - F_{emp}(x)}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0, \quad p \rightarrow \infty,$$

for some $c > 0$ and all sequences $a = a(p) \rightarrow 0$ such that $pa \rightarrow \infty$, $a \in (0, 1)$.

The universe of limiting dfs F is diversified and F may or may not have a heavy tail. Note that the assumption allows even a deterministic data vector X , such as the growth rate

parameters of individuals, stock volatilities or market betas in a cross section. The lower bound c being non-divergent ensures that the limiting df F is unique at each fixed point $x \geq c$. To relax this condition, if necessary, one may replace c by another intermediate quantile sequence as in [Gardes \(2015\)](#) and our asymptotic theory generalizes.

For tail inference we further assume that the limiting distribution is heavy-tailed.

Assumption 2 (Regular variation). The df F is heavy-tailed, i.e., is in the maximum domain of attraction with EVI $\gamma > 0$. In other words, the survival function $1 - F$ is regularly varying at infinity with index $-\frac{1}{\gamma}$:

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad x > 0.$$

The EVI quantifies the tail heaviness of F : the larger γ , the heavier tail and less finite moments F has.

Combining Assumptions 1 and 2, one may now interpret the empirical power laws (1.2) from an asymptotic perspective: the empirical power law is a finite sample approximation of that of the limiting df F . Our model suggests that, if the practitioners believe that the data exhibit certain tail behavior, the power law is the only possible observation under regular variation. This may explain why the power law is such a widely observed phenomenon in real-life data sets.

As noted in the Introduction, our interest is to estimate the tail of the limiting df F which reflects the asymptotic behavior of the larger observations. The rationale is similar to that behind fitting lines in the empirical power law plots as in the Introduction. Specifically, we order the data to obtain the largest $k + 1$ order statistics given by $X_{p-k:p} \leq X_{p-k+1:p} \leq \dots \leq X_{p-1:p} \leq X_{p:p}$ for estimation, where $k = k(p) \in \{1, \dots, p - 1\}$ is such that:

Assumption 3. $k/\log p \rightarrow \infty$ and $k/p \rightarrow 0$ as $p \rightarrow \infty$.

We use the [Hill \(1975\)](#) estimator of the EVI $\gamma > 0$ given by

$$\hat{\gamma} = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{p-i:p} - \log X_{p-k:p}, \quad (2.1)$$

which is the sample average of the excesses over the intermediate quantile $X_{p-k:p}$, on a log scale. We further estimate the extreme quantile of the limiting distribution $Q(1 - \tau)$ with the [Weissman \(1978\)](#) estimator

$$\hat{Q}(1 - \tau) = \left(\frac{k}{p\tau} \right)^{\hat{\gamma}} X_{p-k:p} \quad (2.2)$$

where $\tau = \tau(p)$ is extremely small in the sense that $\tau \rightarrow 0$ according $p\tau \rightarrow \lambda \in [0, \infty)$.

To exclude the very irregular cases where the sample maximum dominates the estimator, we require:

Assumption 4 (Stability condition). $\log X_{p:p} = O_{\mathbb{P}}(\log p)$, as $p \rightarrow \infty$.

If the data are i.i.d. observations from F , this condition is readily satisfied under the regular variation Assumption 2; see, e.g., [de Haan and Ferreira \(2006\)](#), Chapter 1. We argue that this condition is very mild in practice, as the largest point often does not overwhelm the fitting process in the log-log plots such as those in the Introduction.

Now, our unified consistency theorem for empirical power laws is as follows.

Theorem 1. *Under Assumptions 1 – 4, we have, as $p \rightarrow \infty$,*

$$\hat{\gamma} \xrightarrow{\mathbb{P}} \gamma.$$

An interesting corollary for extreme quantile estimation then follows.

Corollary 1. *Under the assumptions of Theorem 1 we have for τ such that $p\tau \rightarrow \lambda \in [0, \infty)$*

$$\frac{\log \hat{Q}(1 - \tau)}{\log Q(1 - \tau)} \xrightarrow{\mathbb{P}} 1.$$

3 Applications to Heterogeneous Econometric Models

In this section, we present and discuss some interesting examples of heterogeneous data that satisfy our assumptions and hence the Hill estimator is consistent in those settings.

Our first example is a generalization of the heteroscedastic extremes model in [Einhorn et al. \(2016\)](#). Let $X = (X_1^{(p)}, \dots, X_p^{(p)})$ have independent entries with continuous dfs $F_{p,1}, \dots, F_{p,p}$. Suppose there exists a df F and a set of non-negative constants $\{c_{p,i}\}$ such that

$$\lim_{x \rightarrow \infty} \frac{1 - F_{p,i}(x)}{1 - F(x)} = c_{p,i}$$

uniformly for all p and $1 \leq i \leq p$, where $\frac{1}{p} \sum_{i=1}^p c_{p,i} \rightarrow 1$, as $p \rightarrow \infty$. Furthermore, assume that the $\{c_{p,i}\}$ are bounded and that the $\{c_{p,i} : c_{p,i} > 0\}$ are bounded away from zero.

Theorem 2 (Heteroscedastic extremes). *For this generalized heteroscedastic extremes model, Chang’s condition (Assumption 1) holds. Furthermore, if F satisfies the regular variation condition (Assumption 2), then the stability condition (Assumption 4) holds.*

Our second example is a heterogeneous scales model defined by

$$X_i^{(p)} = \mu + \sigma_{p,i} Z_i, \quad i = 1, 2, \dots, p, \quad \mu \in \mathbb{R},$$

where the Z_i are i.i.d. random variables with df G and the $\sigma_{p,1} \leq \sigma_{p,2} \leq \dots \leq \sigma_{p,p}$ are positive scale parameters. We assume that the scale parameters resemble the intermediate quantiles of a continuous df F_σ , with positive left endpoint, in such a way that $\sigma_{p,i}/Q_\sigma(i/p) \rightarrow 1$ uniformly for $1 \leq i \leq (1-a)p$, for all intermediate sequences $a = a(p) \rightarrow 0$ such that $pa \rightarrow \infty$. We require that Q_σ is continuous on $(0, 1]$ and that for some $t^+ \in (0, 1)$,

$$\lim_{\delta \downarrow 0} \sup_{0 < t \leq t^+, |s/t-1| \leq \delta} |Q_\sigma(1-s)/Q_\sigma(1-t) - 1| = 0. \quad (3.1)$$

Theorem 3 (Heterogeneous scales). *For the heterogeneous scales model, if F defined by $F(x) = \int_0^\infty G\left(\frac{x-\mu}{s}\right) dF_\sigma(s)$, $x \in \mathbb{R}$, satisfies Assumption 2, then Chang's condition (Assumption 1) holds. Furthermore, if, for some $\alpha > 0$, $1 - G(x) \leq x^{-\alpha}$, for large x , and $\sigma_{p,p} \leq p^{1/\alpha}$, for large p , then the stability condition holds.*

It easily follows that if Y has df F_σ and is independent of Z_1 , then $\mu + YZ_1$ has df F as in this theorem. If F_σ satisfies Assumption 2, and G has a not-heavier tail, i.e., $\lim_{x \rightarrow \infty} \frac{1-G(x)}{1-F_\sigma(x)} \in [0, \infty)$, then the limiting F has the same EVI as F_σ , see Lemma 4.1 in the survey paper [Jessen and Mikosch \(2006\)](#). When G has a heavier tail than F_σ , i.e., $\lim_{x \rightarrow \infty} \frac{1-G(x)}{1-F_\sigma(x)} = \infty$, and G satisfies Assumption 2, F has the same EVI as G . In other words, either the scale parameter heterogeneity or the tail of the marginal variables may generate a heavy tail.

Corollary 2 (High-dimensional Gaussian vector with heterogeneous scales). *Suppose that the heterogeneous scales model holds with*

$$X_i^{(p)} \sim \mathcal{N}(\mu, \sigma_{p,i}^2), \quad i = 1, \dots, p,$$

where F_σ satisfies Assumption 2 with $\gamma > 0$. Then Chang's condition holds with F as in the theorem and G the standard normal df; F is heavy-tailed with the same γ as F_σ .

In this situation, the empirical power law is driven by the heterogeneity of the scales instead of the tail behavior of the marginal variables.

The next corollary to Theorem 3 generalizes it to dependent data.

Corollary 3 (Heterogeneous scales and dependence). *Consider the heterogeneous scales setup in and above Theorem 3, with $\mu = 0$. Let $Y > 0$ be a random variable and define*

$$\tilde{X}_i^{(p)} = Y X_i^{(p)}, \quad i = 1, \dots, p.$$

Let $\tilde{\gamma}$ be the Hill estimator of the $\tilde{X}_i^{(p)}$ and $\hat{\gamma}$ the Hill estimator of the $X_i^{(p)}$. Then $\tilde{\gamma} = \hat{\gamma}$ and hence $\tilde{\gamma} \xrightarrow{\mathbb{P}} \gamma$.

Remark 1 (High-dimensional lognormal vector with dependent components and heterogeneous scales). A very interesting example of the $\tilde{X}_i^{(p)}$ is obtained by taking $Y = \exp(Z)$, with $Z \sim N(0, \rho\sigma^2)$, $\sigma^2 > 0$, $0 < \rho \leq 1$, and $X_i^{(p)} = \sigma_{p,i} \exp(\tilde{Z}_i)$ with $\tilde{Z}_i, i = 1, \dots, p$, independent and $N(0, (1-\rho)\sigma^2)$ -distributed. In this case $\tilde{X}_i^{(p)} = \sigma_{p,i} \exp(Z + \tilde{Z}_i)$ and $(Z + \tilde{Z}_1, \dots, Z + \tilde{Z}_p)$ is multivariate normal with all correlations equal to ρ . Clearly the $\tilde{X}_i^{(p)}$ are dependent and all have a lognormal distribution.

The next example is a heterogeneous locations model defined by

$$X_i^{(p)} = \mu_{p,i} + Z_i, \quad i = 1, 2, \dots, p,$$

where the Z_i are i.i.d. random variables with df G and the $\mu_{p,1} \leq \dots \leq \mu_{p,p}$ are location parameters. Suppose that for some continuous df F_μ and some weight function $q : (0, 1) \rightarrow (0, \infty)$ it holds that

$$\max_{ap \leq i \leq (1-a)p} \frac{|\mu_{p,i} - Q_\mu(i/p)|}{q(i/p)} \rightarrow 0, \quad (3.2)$$

for all sequences $a = a(p) \rightarrow 0$ such that $pa \rightarrow \infty$. Define for $\eta \in \mathbb{R}$,

$$F_{(\eta)}(x) = \int_0^1 G(x - Q_\mu(t) - \eta q(t)) dt, \quad x \in \mathbb{R},$$

and write F for $F_{(0)}$.

Theorem 4 (Heterogeneous locations). *Assume*

- (i) $\lim_{\eta \rightarrow 0} \sup_{x \geq c} \left| \frac{1 - F_{(\eta)}(x)}{1 - F(x)} - 1 \right| = 0$ for some c ,
- (ii) $\lim_{\delta \downarrow 0} \sup_{0 < t \leq 1/2, |s/t - 1| \leq \delta} \left\{ \frac{|Q_\mu(s) - Q_\mu(t)|}{q(t)} + \frac{|Q_\mu(1-s) - Q_\mu(1-t)|}{q(1-t)} \right\} = 0$,
- (iii) $\lim_{\delta \downarrow 0} \sup_{0 < t \leq 1/2, |s/t - 1| \leq \delta} \left\{ \left| \frac{q(s)}{q(t)} - 1 \right| + \left| \frac{q(1-s)}{q(1-t)} - 1 \right| \right\} = 0$.

Then Chang's condition holds. Furthermore, if, for some $\alpha > 0$, $1 - G(x) \leq x^{-\alpha}$ and $\mu_{p,p} \leq p^{1/\alpha}$ for large p , then the stability condition holds.

It readily follows that if Y has df F_μ and is independent of Z_1 , then $Y + Z_1$ has df $F = F_{(0)}$ as defined above the theorem. The limiting df F has the same EVI γ as that of F_μ , if F_μ satisfies Assumption 2, when G has a not-heavier tail such that $\lim_{x \rightarrow \infty} \frac{1 - G(x)}{1 - F_\mu(x)} \in [0, \infty)$ by the proposition on page 278 in Feller (1978) about convolutions of regular varying distributions;

see also Lemma 3.1 in [Jessen and Mikosch \(2006\)](#). When G has a heavier tail than F_μ , i.e., $\lim_{x \rightarrow \infty} \frac{1-G(x)}{1-F_\mu(x)} = \infty$, the limiting df F has the same EVI as G if G satisfies Assumption 2 with a positive EVI γ . In other words, both the location parameter heterogeneity or the tail of the marginal variables may generate a heavy tail.

The examples in Theorems 2-4 and many other examples can be generalized to the setup where many small possibly heterogeneous sets of possibly dependent observations are pooled. This situation can also be seen as a (general, unbalanced) panel data setup. Let T be a fixed, positive integer. Consider data $X_{i,t}$, with $i = 1, \dots, p^*$, and $t \in \mathcal{T}_i$, where $\mathcal{T}_i$ is a non-empty subset of $\{1, \dots, T\}$, for all $1 \leq i \leq p^*$. Now for the asymptotic theory, we assume that for each fixed $t \in \{1, \dots, T\}$, the number of data points $X_{i,t}$ either tends to infinity or stays bounded, as $p^* \rightarrow \infty$. Let p_t be the number of $X_{i,t}$ and denote with $\mathcal{T}_\infty$ those t for which $p_t \rightarrow \infty$ as $p^* \rightarrow \infty$; naturally we require $\mathcal{T}_\infty \neq \emptyset$. Note that the data $X_{i,t}$ may be dependent here.

Theorem 5. *Assume that for each fixed $t \in \mathcal{T}_\infty$, the data $X_{i,t}$ satisfy Chang's condition with $F = F_t$ and $p = p_t$. Let $p = \sum_{t=1}^T p_t$ and assume, for $t \in \mathcal{T}_\infty$, that $\lim_{p^* \rightarrow \infty} p_t/p = w_t > 0$. Consider the empirical distribution function F_{emp} of all p random variables $X_{i,t}$. Define $F = \sum_{t \in \mathcal{T}_\infty} w_t F_t$ and assume F is continuous. Then, as $p \rightarrow \infty$, Chang's condition holds. Denote with $H_{i,t}$ the distribution function of $X_{i,t}$ and assume for all i and t with $t \notin \mathcal{T}_\infty$ that for large x , $H_{i,t}(x) \geq H(x)$ for some fixed distribution function H . Then, if for each fixed $t \in \mathcal{T}_\infty$, the stability condition holds with $p = p_t$, then the stability condition holds for the maximum $X_{p,p}$ of all p data.*

This theorem states that, quite generally, Chang's condition and the stability condition remain valid under pooling. Nothing is required for the data subset $\{X_{i,t}\}$, for fixed i . As proved in Section 2, these two conditions lead to the consistency of the Hill estimator if k is chosen according to Assumption 3 and if F satisfies Assumption 2. The latter assumption follows easily, if the individual $F_t, t \in \mathcal{T}_\infty$, satisfy it for some $\gamma_t > 0$ (but this can be weakened). The pooled or panel data setup provides a very general setting where consistency of the Hill estimator is established.

4 Simulation Study

In this section, we present a simulation study to demonstrate how heterogeneous data generate heavy tails in high dimensions and to evaluate the finite sample performance of the Hill

estimator. We consider three data generating processes:

- (I) Heterogeneous lognormal variables $X_i^{(p)} = \sigma_{p,i} \exp(Z_i)$ with $(Z_1, \dots, Z_p)$ centered multivariate normal with all variances equal to $\sigma = 0.3$ and a common correlation $\rho \in [0, 1]$. The scale parameters $\sigma_{p,i}$ are the quantiles at probability level $\frac{i+p-1}{2p}$ of the log-logistic distribution (Fisk, 1961) with unit median and $\gamma = 0.39$.
- (II) Heterogeneous lognormal variables $\{X_{i,t}^{(p)} : i = 1, \dots, p\}$ at fixed time t generated from the autoregressive model

$$\log X_{i,s}^{(p)} = \lambda \log \sigma_{p,i} + \phi \log X_{i,s-1}^{(p)} + \varepsilon_{i,s}, \quad s = 1, \dots, t,$$

with $\lambda = 0.1$, $\phi = 0.94$ and initial values $X_{i,0}^{(p)} = \sigma_{p,i} \exp(Z_i)$ generated from DGP I with independent $Z_i \sim N(0, \sigma^2)$ with $\sigma = 0.3$. We generate the i.i.d. errors $\varepsilon_{i,s} \sim N(0, \sigma_\varepsilon^2)$, independent of the initial values $\{X_{i,0}^{(p)}\}$, with $\sigma_\varepsilon = \frac{\lambda\sqrt{1-\phi^2}}{1-\phi}\sigma = 0.17$.

- (III) A pooled data set of heterogeneous EGARCH(1,1) data $\{X_{i,t}^{(p)} : i = 1, \dots, p, t = 1, \dots, T\}$ such that

$$\begin{cases} X_{i,t}^{(p)} = \sqrt{h_{i,t}^{(p)}} Z_{i,t} \\ \log h_{i,t}^{(p)} = 0.2 \log \sigma_{p,i}^2 + 0.1 Z_{i,t-1} + 0.8 \log h_{i,t-1}^{(p)} \end{cases}$$

where the $Z_{i,t} \sim N(0, 1)$ are independent errors and the scale parameters $\{\sigma_{p,i}\}$ are the same as in DGP I. We use the initial values $h_{i,1}^{(p)} = \sigma_{p,i}^2$.

All three DGPs satisfy the (pooled) heterogeneous scales model in Section 3. The extreme value indices of DGP I and II are calibrated by using the empirical extreme value indices for United States income distributions from 1973 to 2018, as shown in Figure 2. The empirical extreme value indices (blue line) are calculated as $1 + \log_{10} \frac{S(0.1)}{S(1)}$, where $S(p)$ denotes the share of the top p th percentile of the income data taken from Table A.1 in Piketty and Saez (2003) and the updated data set downloaded at:

<https://eml.berkeley.edu/saez/TabFig2018.xls>.

For DGP I and the initial values in DGP II, we have set the cross-sectional EVI $\gamma = 0.39$ equal to the empirical EVI of 1973. For DGP II, the cross-sectional extreme value indices change over time and obey the first-order difference equation

$$\gamma_t = 0.1\gamma + 0.94\gamma_{t-1} = 0.039 + 0.94\gamma_{t-1},$$

Figure 2: Empirical extreme value indices and calibrated extreme value indices for US income (excluding capital gains) distributions between 1973 and 2018

where γ_t denotes the cross-sectional EVI at time t (or year $1973 + t$ in Figure 2), and the intercept 0.039 and the autoregressive coefficient 0.94 are the least-squares estimates for the time series of empirical extreme value indices. Our model can replicate the rapid rise in US income inequality and implies a long-run steady state $\lim_{t \rightarrow \infty} \gamma_t = 0.65$. For each time period t , the cross-sectional data set in DGP II can be rewritten as a heterogeneous scales model:

$$X_{i,t}^{(p)} = \left(\sigma_{p,i} \exp \left(\tilde{Z}_{i,t} \right) \right)^{\gamma_t / \gamma} = \sigma_{p,i}^{\gamma_t / \gamma} \exp \left(Z_{i,t} \right), \quad \tilde{Z}_{i,t} = \frac{\gamma_t}{\gamma} Z_{i,t} \sim N \left(0, \sigma_t^2 \right),$$

where

$$\sigma_t^2 = (1 - \rho_t) \sigma^2, \quad \rho_t = 1 - \left(\frac{\gamma}{\gamma_t} \right)^2 \left(\phi^{2t} + \frac{1 - \phi^{2t}}{1 - \phi^2} \frac{\sigma_\varepsilon^2}{\sigma^2} \right).$$

Although the cross-sectional data are independent at time t , the distribution of the Hill estimator is equivalent to that for a power transformation of DGP I with correlation $\rho = \rho_t$ after cancelling the common random scale component; see Remark 1. We have chosen $\sigma_\varepsilon = \frac{\lambda \sqrt{1 - \phi^2}}{1 - \phi} \sigma$ such that ρ_t dies out in the long run, that is,

$$\lim_{t \rightarrow \infty} \rho_t = 1 - \left(\frac{1 - \phi}{\lambda} \right)^2 \cdot \frac{1}{1 - \phi^2} \frac{\sigma_\varepsilon^2}{\sigma^2} = 0.$$

We argue that $\sigma_\varepsilon = 0.17$ is a reasonably calibrated volatility for (high) income growth; see, e.g., Gabaix et al. (2016). Finally, DGP III is motivated by the financial applications in Kelly and Jiang (2014), where all the daily (idiosyncratic) stock losses in the same month are pooled for the monthly estimation. We have chosen the initial values of $h_{i,t}$ such that

the cross-sectional extreme value indices stabilize over time t for simplicity; otherwise the cross-sectional extreme value indices follow a similar dynamics as DGP II.

Figure 3: Empirical power laws in log-log plots for one sample with $p = 1000$.

Figure 3 shows the log-log plots for the largest 5% values from one sample of size $p = 1000$ from DGP I–III. From the first row, for DGP I, we observe that the Hill estimate is closer

Table 1: Average and mean absolute deviation of the relative estimation errors $\hat{\gamma}/\gamma - 1$ based on 5000 replications for dimension $p = 200, 1000$ and 5000 .

DGP I: $\gamma = 0.39$					
p	$\rho = 0.00$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$	$\rho = 1.00$
200	-0.031 (0.117)	-0.036 (0.110)	-0.040 (0.101)	-0.044 (0.084)	-0.045 (0.000)
1000	-0.017 (0.075)	-0.020 (0.069)	-0.021 (0.062)	-0.022 (0.053)	-0.025 (0.000)
5000	-0.008 (0.046)	-0.011 (0.043)	-0.011 (0.039)	-0.012 (0.033)	-0.012 (0.000)
DGP II: $\gamma_t = 0.039 + 0.94\gamma_{t-1}$, $\gamma_0 = 0.39$					
p	$t = 1$	$t = 26$	$t = 51$	$t = 75$	$t = 100$
200	-0.004 (0.095)	-0.018 (0.124)	-0.026 (0.120)	-0.024 (0.120)	-0.027 (0.119)
1000	0.016 (0.061)	-0.008 (0.076)	-0.015 (0.074)	-0.017 (0.074)	-0.017 (0.074)
5000	0.028 (0.037)	-0.001 (0.048)	-0.006 (0.047)	-0.008 (0.046)	-0.008 (0.046)
DGP III: $\gamma = 0.39$					
p	$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$
200	-0.008 (0.132)	-0.011 (0.096)	-0.013 (0.077)	-0.016 (0.068)	-0.015 (0.062)
1000	-0.006 (0.086)	-0.005 (0.060)	-0.009 (0.050)	-0.010 (0.042)	-0.010 (0.038)
5000	-0.004 (0.054)	-0.004 (0.038)	-0.005 (0.031)	-0.004 (0.027)	-0.005 (0.024)

to the true value $\gamma = 0.39$ for the larger ρ . In the second row, for DGP II, we report the estimates for $t = 1$ (year 1974) and $t = 45$ (year 2018). The Hill estimates are both close to the true empirical extreme value indices, which are different over time as discussed above. In the last row, for DGP III, we observe that the Hill estimate for the pooled data with $T = 4$ is slightly closer to the true value than that for $T = 1$.

To demonstrate the convergence of the Hill estimator as the dimension p grows, we report in Table 1 the average and mean absolute deviation of the relative estimation error of the Hill estimates against the true values based on 5000 replications for dimension $p = 200, 1000$ and 5000 with k is equal to 10%, 5%, and 2.5% of the dimension p for DGP I and II and of the pooled sample size pT for DGP III, respectively. In general, the absolute bias and the variability decrease as the dimension p grows for all DGPs, in line with the consistency of the Hill estimator. For DGP I, the mean absolute deviation decreases as the correlation ρ grows. This is because the Hill estimator is invariant with respect to the factor Y in Remark 1, while the idiosyncratic component $\tilde{Z}_i$ there has a decreasing variance as ρ grows. Indeed,

in the extreme case with $\rho = 1$, the Hill estimator is fully driven by the parameters and becomes deterministic. For DGP II, for each p , the bias and variability of the Hill estimator stabilize as t grows and approaches that for DGP I with $\rho = 0$ by construction. For DGP III, pooling the cross-sectional data over multiple time periods yields a larger sample size and hence smaller variability.

5 Analysis of International COVID-19 Data

In this section, we study the empirical power laws of international COVID-19 data. Our sample period is from December 31, 2019 to September 1, 2020. The data are obtained from the database:

<https://ourworldindata.org/coronavirus-source-data>,

which is regularly updated through “Our World in Data”. We report results for the weekly new confirmed cases and for the weekly new deaths for the countries with a population of at least 1 million people. The raw data are 7-days rolling sums and reported on a daily basis. We have repeated the analysis in this section for biweekly data and the findings are qualitatively the same. As the outbreaks did not occur simultaneously over the world, we align the timeline over countries, beginning with the day the total number of confirmed cases per million people reached 1.

Figure 4 depicts the distribution of the outbreak dates, and the cross-sectional average over countries of the 7-day rolling sum of new confirmed cases and of new deaths (both per million people) as a function of the number of days since the outbreak. We consider data until 165 days after the outbreak to maintain a minimum of 100 countries in our cross-sectional estimations. The average number of new confirmed cases grows rapidly in the first month, stabilizes over the second month, and grows approximately in a linear pattern afterwards. The number of new deaths peaks around day 40, then decreases dramatically until about day 80 and continues to grow slowly to around day 120 and then tends to stabilize at the end.

We observe a persistent cross-sectional power law across countries, indicating a high degree of international inequality of new COVID-19 infections and deaths. To the best of our knowledge, this is the first paper exhibiting cross-sectional empirical power laws of international COVID-19 data. To demonstrate the empirical power laws, we report log-log plots for the cross-sectional numbers of new confirmed cases and new deaths per million people 7, 30 and 90 days since the outbreak in Figure 5. For each plot, we use the largest 15% values

Figure 4: The time distribution of outbreak dates (left), and the cross-sectional averages of 7-day rolling sums of new confirmed cases per million people and of new deaths per million people (right).

available on the day (i.e. $k \approx 23$). The estimated extreme value indices range between 0.66 and 1.08, suggesting a heavy tail with infinite variance for the limiting distributions.

Next, observe that the scale parameters of new confirmed cases and new deaths show a high degree of heterogeneity across countries. Figure 6 depicts the log-log plots for the 15% largest mean absolute deviations for weekly new confirmed cases and weekly new deaths and shows again empirical power law behavior with somewhat similar estimates for the EVI as in Figure 5.

Therefore, in Figure 7 we plot the time series of estimated cross-sectional extreme value indices for the raw numbers and the standardized numbers normalized per country by the mean absolute deviations. Our normalization is based on the data over the entire period and not yet useful in the beginning of the outbreak. After this burning period, we observe a dramatic reduction in the estimated extreme value indices after standardization. It suggests that the cross-sectional power laws are largely driven by the scale parameter heterogeneity across countries throughout the COVID-19 pandemic.

(a) 7 days since outbreak

(b) 30 days since outbreak

(c) 90 days since outbreak

Figure 5: Empirical power laws of cross-sectional observations.

Figure 6: Empirical power laws of mean absolute deviations across countries.

Figure 7: Time series of estimated cross-sectional extreme value indices for raw and standardized numbers.

6 Proofs

Proof of Theorem 1. Define $b_p = \frac{\sqrt{k}}{p\sqrt{\log p}}$ and consider

$$\begin{aligned}
 \frac{1}{k} \sum_{i=0}^{\lfloor pb_p \rfloor} \log X_{p-i:p} - \log X_{p-k:p} &\leq \frac{\lfloor pb_p \rfloor + 1}{k} (\log X_{p:p} - \log X_{p-k:p}) \\
 &= O_{\mathbb{P}} \left(\frac{1}{k} \frac{\sqrt{k}}{\sqrt{\log p}} \log p \right) = o_{\mathbb{P}}(1).
 \end{aligned}$$

Write $Y_i^{(p)} = \frac{1}{1-F(X_i^{(p)})}$. Then $X_i^{(p)} = U(Y_i^{(p)})$, with the function $U = \left(\frac{1}{1-F}\right)^{-1}$. Let $Y_{j:p}$, $j = 1, \dots, p$ be the order statistics of the $Y_i^{(p)}$. By Assumptions 1 and 2 we have that $Y_{p-k:p} \xrightarrow{\mathbb{P}} \infty$. Now, as in the proof of Theorem 3.2.2 (the consistency of the Hill estimator in the i.i.d. case) in [de Haan and Ferreira \(2006\)](#), it suffices to prove that, as $p \rightarrow \infty$,

$$\frac{1}{k} \sum_{i=\lfloor pb_p \rfloor + 1}^{k-1} \log \frac{Y_{p-i:p}}{Y_{p-k:p}} \xrightarrow{\mathbb{P}} 1. \quad (6.1)$$

Let G_p be the empirical df of the $Y_i^{(p)}$, $i = 1, \dots, p$. Then Assumption 1 implies for $\delta > 0$

$$\sup_{F(c)+\delta \leq u \leq 1-2a_p} \left| (1-u)G_p^{-1}(u) - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.2)$$

Let $\varepsilon > 0$. Then with probability tending to 1, we have from (6.2) with $a_p \leq \frac{1}{2}b_p$ (and $pa_p \rightarrow \infty$)

$$-\varepsilon + \frac{1}{k} \sum_{i=\lfloor pb_p \rfloor + 1}^{k-1} \log \frac{k}{i} \leq \frac{1}{k} \sum_{i=\lfloor pb_p \rfloor + 1}^{k-1} \log \frac{Y_{p-i:p}}{Y_{p-k:p}} \leq \varepsilon + \frac{1}{k} \sum_{i=\lfloor pb_p \rfloor + 1}^{k-1} \log \frac{k}{i}.$$

It is elementary to show that $\frac{1}{k} \sum_{i=\lfloor pb_p \rfloor + 1}^{k-1} \log \frac{k}{i} \rightarrow 1$. This yields (6.1). $\square$

Proof of Corollary 1. Write

$$\frac{\log \widehat{Q}(1-\tau)}{\log Q(1-\tau)} = \frac{\log X_{p-k:p} + \widehat{\gamma} \log \frac{k}{p\tau}}{\log Q\left(1 - \frac{k}{p}\right) + \gamma \log \frac{k}{p\tau}} \cdot \frac{\log \left(Q\left(1 - \frac{k}{p}\right) \left(\frac{k}{p\tau}\right)^\gamma\right)}{\log Q(1-\tau)}.$$

We have from the [Potter \(1942\)](#) inequality that for $\varepsilon > 0$, $x \geq 1$ and $t \geq t_0$

$$(1-\varepsilon)x^{\gamma(1-\varepsilon)} \leq \frac{Q(1-(tx)^{-1})}{Q(1-t^{-1})} \leq (1+\varepsilon)x^{\gamma(1+\varepsilon)} \quad (6.3)$$

Now using (6.2), twice (6.3), and the consistency of $\widehat{\gamma}$ yields the result. $\square$

Proof of Theorem 2. Reshuffle the indices i such that the positive $c_{p,i}$ have a lower index than the $c_{p,i}$ equal to 0 and let κ_p be the number of positive $c_{p,i}$, $i = 1, \dots, p$. Let $U_i = 1 - F_{p,i}(X_i^{(p)})$. Then the U_i are independent uniform-(0,1) random variables, and

$$\begin{aligned} 1 - F_{emp}(x) &= \frac{1}{p} \sum_{i=1}^p \mathbb{1}(X_i^{(p)} > x) = \frac{1}{p} \sum_{i=1}^p \mathbb{1}(U_i < 1 - F_{p,i}(x)) \\ &= \frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < 1 - F_{p,i}(x)) + \frac{1}{p} \sum_{i=\kappa_p+1}^p \mathbb{1}(U_i < 1 - F_{p,i}(x)). \end{aligned}$$

Let $\eta > 0$. For $i > \kappa_p$ and $x \geq c$ (for some large c) we have

$$1 - F_{p,i}(x) \leq \eta(1 - F(x)),$$

and hence

$$\frac{1}{p} \sum_{i=\kappa_p+1}^p \mathbb{1}(U_i < 1 - F_{p,i}(x)) \leq \frac{1}{p} \sum_{i=\kappa_p+1}^p \mathbb{1}(U_i < \eta(1 - F(x))).$$

Define $\Gamma_p(y) = \frac{1}{p} \sum_{i=1}^p \mathbb{1}(U_i \leq y)$, the uniform empirical distribution function. Then

$$\frac{\frac{1}{p} \sum_{i=\kappa_p+1}^p \mathbb{1}(U_i < 1 - F_{p,i}(x))}{1 - F(x)} \leq \frac{\Gamma_p(\eta(1 - F(x)))}{1 - F(x)}, \quad \text{for } x \geq c.$$

Using Theorem 0 in [Wellner \(1978\)](#), we obtain that with probability tending to 1, as $p \rightarrow \infty$,

$$\sup_{c \leq x \leq Q(1-a_p)} \frac{\Gamma_p(\eta(1 - F(x)))}{\eta(1 - F(x))} \cdot \eta \leq 2\eta.$$

Since $\eta > 0$ can be made arbitrarily small, it remains to show that

$$\sup_{c \leq x \leq Q(1-a_p)} \left| \frac{\frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < 1 - F_{p,i}(x))}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.4)$$

Let $\varepsilon \in (0, 1/(2d))$, where $d = \sup_p \max_{1 \leq i \leq p} c_{p,i} \in (0, \infty)$.

Let $B_i, i = 1, \dots, \kappa_p$, be independent Bernoulli random variables with $\mathbb{P}(B_i = 1) = c_{p,i}\varepsilon$. Write $N_p = \sum_{i=1}^{\kappa_p} B_i$. Let $V_j, j = 1, \dots, \kappa_p$, be independent uniform variables on $(0, 1)$, independent of the B_i 's. Pair the B_i and V_j in the following way: assign the first N_p, V_j to the B_i which are equal to 1 (and the other V_j to the B_i which are 0). Define

$$U_i = B_i c_{p,i}\varepsilon V_j + (1 - B_i)[c_{p,i}\varepsilon + (1 - c_{p,i}\varepsilon)V_j].$$

Then $U_1, \dots, U_{\kappa_p}$ are independent uniform variables on $(0, 1)$. (These U_i are different from those in (6.4), but since we consider only convergence in probability that is allowed.) Let $\delta > 0$. For x large, since $\inf_p \min_{1 \leq i \leq \kappa_p} c_{p,i} > 0$,

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < (1 - F(x))c_{p,i}(1 - \delta)) &\leq \frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < (1 - F_{p,i}(x))) \\ &\leq \frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < (1 - F(x))c_{p,i}(1 + \delta)). \end{aligned} \quad (6.5)$$

Write $\tilde{\Gamma}_m(y) = \frac{1}{m} \sum_{j=1}^m \mathbb{1}(V_j \leq y)$, $m = 1, \dots, \kappa_p$. Now we have, for either choice of sign, for large x ,

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^{\kappa_p} \mathbb{1}(U_i < (1 - F(x))c_{p,i}(1 \pm \delta)) &= \frac{1}{p} \sum_{j=1}^{N_p} \mathbb{1}[V_j < (1 - F(x))(1 \pm \delta)/\varepsilon] \\ &= \frac{N_p}{p} \tilde{\Gamma}_{N_p}((1 - F(x))(1 \pm \delta)/\varepsilon). \end{aligned}$$

Hence, for large x , the ratio in (6.4) can be bounded from below/above by

$$\frac{\tilde{\Gamma}_{N_p}((1 - F(x))(1 \pm \delta)/\varepsilon)}{\frac{p}{N_p}(1 - F(x))} =: R_{p,\pm}(x).$$

Clearly $\frac{EN_p}{p} = \frac{1}{p} \sum_{i=1}^{\kappa_p} c_{p,i} \varepsilon \rightarrow \varepsilon$. Now by the law of large numbers and again Theorem 0 in Wellner (1978), we have that

$$\sup_{c \leq x \leq Q(1-a_p)} |R_{p,\pm} - 1| \xrightarrow{\mathbb{P}} \delta.$$

Since $\delta > 0$ is arbitrary, this in combination with (6.5) yields (6.4).

The stability condition follows from

$$Q_{p,i}(t) \leq Q(1 - (1 - t)/(2d)), \quad t_0 \leq t < 1, \text{ for some } t_0,$$

and the asymptotic behavior of the minimum of p independent uniform-(0,1) random variables. $\square$

Proof of Theorem 3. Without loss of generality, we may and will assume $\mu = 0$.

Let $U_1, \dots, U_p$ be i.i.d. uniform-(0,1) random variables and denote their order statistics by $U_{1:p} \leq \dots \leq U_{p:p}$. Let $Z_1, \dots, Z_p$ be i.i.d. random variables from the df G and independent of the U_i . Consider the empirical survival functions

$$S_p(x) := \frac{1}{p} \sum_{i=1}^p \mathbb{1}(Q_\sigma(U_i)Z_i > x) \text{ and } \tilde{S}_p(x) := \frac{1}{p} \sum_{i=1}^p \mathbb{1}(Q_\sigma(U_{i:p})Z_i > x), x \in \mathbb{R}.$$

Note that the $Q_\sigma(U_i)Z_i$ are i.i.d. with df F as in the theorem. Also observe that $\tilde{S}_p \stackrel{d}{=} S_p$ by the exchangeability of the Z_i and the independence of the U_i and the Z_i .

By Theorem 0 in Wellner (1978), for any intermediate sequence $a = a(p) \rightarrow 0$ such that $pa \rightarrow \infty$,

$$\sup_{x \leq Q(1-a)} \left| \frac{\tilde{S}_p(x)}{1 - F(x)} - 1 \right| \stackrel{d}{=} \sup_{x \leq Q(1-a)} \left| \frac{S_p(x)}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.6)$$

We need to show a similar result, namely

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{1 - F_{emp}(x)}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.7)$$

We first show

$$\sup_{1 \leq i \leq (1-b)p} \left| \frac{Q_\sigma(U_{i:p})}{Q_\sigma(i/p)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.8)$$

where $b = b(p) \rightarrow 0$ is such that $b/a \rightarrow 0$ and $pb \rightarrow \infty$. The Glivenko-Cantelli theorem yields

$$\sup_{1 \leq i \leq p} |U_{i:p} - i/p| \xrightarrow{\mathbb{P}} 0.$$

From this result, we obtain the convergence in (6.8) for the supremum over $1 \leq i \leq (1-t^+)p$, by the (uniform) continuity of Q_σ on $(0, 1-t^+]$ and $Q_\sigma(0+) > 0$. Theorem 0 in Wellner (1978) also states that

$$\sup_{1 \leq i \leq (1-b)p} \left| \frac{1 - U_{i:p}}{1 - i/p} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.9)$$

Using this in combination with (3.1), we obtain the convergence in (6.8) for the supremum over $p(1-t^+) \leq i \leq (1-b)p$. Clearly by the first assumption of the model, (6.8) implies

$$\sup_{1 \leq i \leq (1-b)p} \left| \frac{Q_\sigma(U_{i:p})}{\sigma_{p,i}} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.10)$$

Now decompose

$$1 - F_{emp}(x) = \frac{1}{p} \sum_{i=1}^{p-\lfloor pb \rfloor} \mathbb{1}(\sigma_{p,i} Z_i > x) + \frac{1}{p} \sum_{i=p-\lfloor pb \rfloor+1}^p \mathbb{1}(\sigma_{p,i} Z_i > x) =: \widehat{S}_p^{(1)}(x) + \widehat{S}_p^{(2)}(x).$$

Note that

$$\sup_{c \leq x \leq Q(1-a)} \frac{\widehat{S}_p^{(2)}(x)}{1 - F(x)} \leq \frac{b}{a} \rightarrow 0.$$

Hence for (6.7) it remains to show that

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{\widehat{S}_p^{(1)}(x)}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.11)$$

Similarly, we can decompose

$$\widetilde{S}_p(x) = \frac{1}{p} \sum_{i=1}^{p-\lfloor pb \rfloor} \mathbb{1}(Q_\sigma(U_{i:p}) Z_i > x) + \frac{1}{p} \sum_{i=p-\lfloor pb \rfloor+1}^p \mathbb{1}(Q_\sigma(U_{i:p}) Z_i > x) =: \widetilde{S}_p^{(1)}(x) + \widetilde{S}_p^{(2)}(x).$$

Let $\varepsilon \in (0, 1)$. Recalling (6.10), with probability tending to 1, for all $x > 0$

$$\widetilde{S}_p^{(1)}((1+\varepsilon)x) \leq \widehat{S}_p^{(1)}(x) \leq \widetilde{S}_p^{(1)}((1-\varepsilon)x).$$

For $\widetilde{S}_p^{(2)}$ we use again $b/a \rightarrow 0$. Hence, for (6.11) it suffices to show that, for either choice of sign, with probability tending to 1, for some $N > 0$ and small $\varepsilon > 0$

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{\widetilde{S}_p((1 \pm \varepsilon)x)}{1 - F(x)} - 1 \right| \leq N\varepsilon.$$

For either choice of sign,

$$\left| \frac{\tilde{S}_p((1 \pm \varepsilon)x)}{1 - F(x)} - 1 \right| \leq \left| \frac{\tilde{S}_p((1 \pm \varepsilon)x)}{1 - F((1 \pm \varepsilon)x)} - 1 \right| \frac{1 - F((1 \pm \varepsilon)x)}{1 - F(x)} + \left| \frac{1 - F((1 \pm \varepsilon)x)}{1 - F(x)} - 1 \right|.$$

Recall from (6.6) that, with probability tending to 1,

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{\tilde{S}_p((1 \pm \varepsilon)x)}{1 - F((1 \pm \varepsilon)x)} - 1 \right| \leq \sup_{x \leq Q(1-a')} \left| \frac{\tilde{S}_p(x)}{1 - F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0,$$

where $a' = 1 - F((1 + \varepsilon)Q(1 - a))$ is again an intermediate sequence as the regular variation of F implies that $a'/a \rightarrow (1 + \varepsilon)^{-\frac{1}{\gamma}} \in (0, \infty)$. Furthermore, by Potter (1942), it readily follows that there exists an $\tilde{N} > 0$, not depending on ε , such that

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{1 - F((1 \pm \varepsilon)x)}{1 - F(x)} - 1 \right| \leq \tilde{N}\varepsilon$$

for some $c > 0$. As $\varepsilon > 0$ can be made arbitrarily small, we obtain (6.11) and hence (6.7).

For the stability condition: we have for large enough M ,

$$\begin{aligned} \mathbb{P}(X_{p:p} > p^M) &\leq \mathbb{P}(\max_{1 \leq i \leq p} Z_i > p^M / \sigma_{p,p}) \leq \mathbb{P}(\max_{1 \leq i \leq p} Z_i > p^{M-1/\alpha}) \\ &= 1 - G^p(p^{M-1/\alpha}) \leq 1 - (1 - p^{1-M\alpha})^p \leq p^{2-M\alpha} \rightarrow 0, \end{aligned}$$

as $p \rightarrow \infty$. □

Proof of Theorem 4. This proof is somewhat similar to that of Theorem 3, except we need to control the perturbations using the weight function q . Again, let $U_1, \dots, U_p$ be i.i.d. uniform-(0,1) random variables and denote their order statistics by $U_{1:p} \leq \dots \leq U_{p:p}$; also, let $Z_1, \dots, Z_p$ be i.i.d. random variables from the df G and independent of the U_i . Consider the empirical survival functions

$$\begin{aligned} S_{p,\eta}(x) &:= \frac{1}{p} \sum_{i=1}^p \mathbb{1}(Q_\mu(U_i) + \eta q(U_i) + Z_i > x), \\ \tilde{S}_{p,\eta}(x) &= \frac{1}{p} \sum_{i=1}^p \mathbb{1}(Q_\mu(U_{i:p}) + \eta q(U_{i:p}) + Z_i > x), \quad x \in \mathbb{R}. \end{aligned}$$

Note that the $Q_\mu(U_i) + \eta q(U_i) + Z_i$ are i.i.d. with df $F_{(\eta)}$ (as defined just above the theorem). Also observe that $\tilde{S}_{p,\eta} \stackrel{d}{=} S_{p,\eta}$ by the exchangeability of the Z_i and the independence of the U_i and Z_i .

Choose another sequence $b = b(p)$ with $b/a \rightarrow 0$ and $pb \rightarrow \infty$ and write

$$\begin{aligned} 1 - F_{emp}(x) &= \frac{1}{p} \sum_{i=1}^{\lfloor pb \rfloor} \mathbb{1}(\mu_{p,i} + Z_i > x) + \frac{1}{p} \sum_{i=\lfloor pb \rfloor+1}^{p-\lfloor pb \rfloor} \mathbb{1}(\mu_{p,i} + Z_i > x) \\ &\quad + \frac{1}{p} \sum_{i=p-\lfloor pb \rfloor+1}^p \mathbb{1}(\mu_{p,i} + Z_i > x) =: \widehat{S}_p^{(1)}(x) + \widehat{S}_p^{(2)}(x) + \widehat{S}_p^{(3)}(x). \end{aligned}$$

Observe that $\sup_{x \leq Q(1-a)} \frac{\widehat{S}_p^{(1)}(x)}{1-F(x)} \leq \frac{b}{a} \rightarrow 0$ and $\sup_{x \leq Q(1-a)} \frac{\widehat{S}_p^{(3)}(x)}{1-F(x)} \leq \frac{b}{a} \rightarrow 0$. Hence for Chang's condition it suffices to show that

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{\widehat{S}_p^{(2)}(x)}{1-F(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \quad (6.12)$$

Now recall (6.9) and note that $(1 - U_{1:p}, \dots, 1 - U_{p:p}) \stackrel{d}{=} (U_{p:p}, \dots, U_{1:p})$. Using conditions (ii) and (iii),

$$\max_{pb \leq i \leq p(1-b)} \frac{|Q_\mu(U_{i:p}) - Q_\mu(i/p)|}{q(i/p)} \xrightarrow{\mathbb{P}} 0 \text{ and } \max_{pb \leq i \leq p(1-b)} \frac{|q(U_{i:p}) - q(i/p)|}{q(i/p)} \xrightarrow{\mathbb{P}} 0.$$

It follows that

$$\frac{|\mu_{p,i} - Q_\mu(U_{i:p})|}{q(U_{i:p})} \leq \frac{|\mu_{p,i} - Q_\mu(i/p)|}{q(i/p)} \frac{q(i/p)}{q(U_{i:p})} + \frac{|Q_\mu(U_{i:p}) - Q_\mu(i/p)|}{q(i/p)} \frac{q(i/p)}{q(U_{i:p})} = o_{\mathbb{P}}(1),$$

uniformly for $pb \leq i \leq p(1-b)$. Hence, for any $\delta > 0$, with probability tending to one,

$$\frac{\widetilde{S}_{p,-\delta}(x)}{1-F(x)} - \frac{2b}{a} \leq \frac{\widehat{S}_p^{(2)}(x)}{1-F(x)} \leq \frac{\widetilde{S}_{p,\delta}(x)}{1-F(x)}, \quad (6.13)$$

uniformly for $c \leq x \leq Q(1-a)$, where for the first inequality we decompose $\widetilde{S}_{p,-\delta}$ like $1 - F_{emp}$ above. For either choice of sign,

$$\left| \frac{\widetilde{S}_{p,\pm\delta}(x)}{1-F(x)} - 1 \right| \leq \left| \frac{\widetilde{S}_{p,\pm\delta}(x)}{1-F_{(\pm\delta)}(x)} - 1 \right| \left| \frac{1-F_{(\pm\delta)}(x)}{1-F(x)} \right| + \left| \frac{1-F_{(\pm\delta)}(x)}{1-F(x)} - 1 \right|.$$

Let $\varepsilon \in (0, \frac{1}{2})$. Note that by condition (i), we have that, there exists a $\delta_1 > 0$ such that for $0 < \delta \leq \delta_1$ and either choice of sign,

$$\sup_{c \leq x} \left| \frac{1-F_{(\pm\delta)}(x)}{1-F(x)} - 1 \right| \leq \varepsilon. \quad (6.14)$$

Furthermore, by Theorem 0 in Wellner (1978), for either choice of sign,

$$\sup_{x \leq Q(1-a)} \left| \frac{\widetilde{S}_{p,\pm\delta}(x)}{1-F_{(\pm\delta)}(x)} - 1 \right| \stackrel{d}{=} \sup_{x \leq Q(1-a)} \left| \frac{S_{p,\pm\delta}(x)}{1-F_{(\pm\delta)}(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0, \quad (6.15)$$

as $1 - F_{(\pm\delta)}(Q(1-a))$ is again an intermediate sequence, by (6.14). Hence, with probability tending to 1,

$$\sup_{c \leq x \leq Q(1-a)} \left| \frac{\tilde{S}_{p,\pm\delta}(x)}{1 - F(x)} - 1 \right| \leq 2\varepsilon.$$

Combining this with (6.13) and recalling that $b/a \rightarrow 0$, we obtain (6.12).

The proof of the stability condition is straightforward using $X_p^{(p)} \leq \mu_{p,p} + \max_{1 \leq i \leq p} Z_i$. $\square$

Proof of Theorem 5. Let $F_{p,t}$ denote the empirical df of the subsample corresponding to a fixed $t \in \{1, \dots, T\}$. Observe that F_{emp} is a convex combination of the $F_{p,t}$ with weights p_t/p , that is,

$$F_p(x) = \sum_{t=1}^T \frac{p_t}{p} F_{p,t}(x), \quad x \in \mathbb{R}.$$

Recall that $F(x) = \sum_{t \in \mathcal{T}_\infty} w_t F_t(x)$. Denote the generalized quantile function of F as Q and that of F_t as Q_t . Then

$$\begin{aligned} \left| \frac{1 - F_{emp}(x)}{1 - F(x)} - 1 \right| &= \left| \sum_{t=1}^T \frac{p_t}{p} \frac{1 - F_{p,t}(x)}{1 - F(x)} - \sum_{t \in \mathcal{T}_\infty} w_t \frac{1 - F_t(x)}{1 - F(x)} \right| \\ &\leq \sum_{t \in \mathcal{T}_\infty} w_t \left| \frac{1 - F_{p,t}(x)}{1 - F(x)} - \frac{1 - F_t(x)}{1 - F(x)} \right| + \sum_{t \notin \mathcal{T}_\infty} \frac{p_t}{p} \frac{1 - F_{p,t}(x)}{1 - F(x)} \\ &\quad + \sum_{t \in \mathcal{T}_\infty} \left| \frac{p_t}{p} - w_t \right| \frac{1 - F_{p,t}(x)}{1 - F(x)} \\ &=: J_1(x) + J_2(x) + J_3(x). \end{aligned}$$

For any intermediate sequence $a = a(p)$,

$$\sup_{x \leq Q(1-a)} J_2(x) \leq \sum_{t \notin \mathcal{T}_\infty} \frac{p_t}{pa} \rightarrow 0.$$

Now take another intermediate sequence $b = b(p)$ such that $b/a \rightarrow 0$. We have that

$$\begin{aligned} \sup_{c \leq x \leq Q(1-a)} J_1(x) &\leq \sum_{t \in \mathcal{T}_\infty} w_t \sup_{c \leq x \leq \min\{Q_t(1-b), Q(1-a)\}} \left| \frac{1 - F_{p,t}(x)}{1 - F(x)} - \frac{1 - F_t(x)}{1 - F(x)} \right| \\ &\quad + \sum_{t \in \mathcal{T}_\infty} w_t \sup_{\min\{Q_t(1-b), Q(1-a)\} \leq x \leq Q(1-a)} \left| \frac{1 - F_{p,t}(x)}{1 - F(x)} - \frac{1 - F_t(x)}{1 - F(x)} \right| \\ &=: A_1 + A_2. \end{aligned}$$

By the assumed Chang's conditions for each $t \in \mathcal{T}_\infty$, we have for sufficiently large c ,

$$\begin{aligned} A_1 &= \sum_{t \in \mathcal{T}_\infty} \sup_{c \leq x \leq \min\{Q_t(1-b), Q(1-a)\}} \left| \frac{1 - F_{p,t}(x)}{1 - F_t(x)} - 1 \right| \cdot w_t \frac{1 - F_t(x)}{1 - F(x)} \\ &\leq \max_{t \in \mathcal{T}_\infty} \sup_{c \leq x \leq \min\{Q_t(1-b), Q(1-a)\}} \left| \frac{1 - F_{p,t}(x)}{1 - F_t(x)} - 1 \right| \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

On the other hand, again by the Chang's conditions,

$$\begin{aligned} A_2 &\leq \sum_{t \in \mathcal{T}_\infty} w_t \cdot \frac{1}{a} \sup_{\min\{Q_t(1-b), Q(1-a)\} \leq x \leq Q(1-a)} \{(1 - F_{p,t}(x)) + (1 - F_t(x))\} \\ &\leq \sum_{t \in \mathcal{T}_\infty} \frac{w_t}{a} \{(1 - F_{p,t}(Q_t(1-b))) + (1 - F_t(Q_t(1-b)))\} \\ &\leq \sum_{t \in \mathcal{T}_\infty} \frac{w_t b}{a} \left\{ \frac{1 - F_{p,t}(Q_t(1-b))}{1 - F_t(Q_t(1-b))} + 1 \right\} \\ &\leq \max_{t \in \mathcal{T}_\infty} \left| \frac{1 - F_{p,t}(Q_t(1-b))}{1 - F_t(Q_t(1-b))} + 1 \right| \sum_{t \in \mathcal{T}_\infty} \frac{w_t b}{a} \\ &\leq \max_{t \in \mathcal{T}_\infty} \left| \frac{1 - F_{p,t}(Q_t(1-b))}{1 - F_t(Q_t(1-b))} + 1 \right| \cdot \frac{b}{a} = O_{\mathbb{P}}(1) \cdot o(1) \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Finally,

$$\begin{aligned} J_3(x) &\leq \sum_{t \in \mathcal{T}_\infty} \left| \frac{p_t}{p} - w_t \right| \left| \frac{1 - F_{p,t}(x)}{1 - F(x)} - \frac{1 - F_t(x)}{1 - F(x)} \right| + \sum_{t \in \mathcal{T}_\infty} \left| \frac{p_t}{p} - w_t \right| \frac{1 - F_t(x)}{1 - F(x)} \\ &\leq \max_{t \in \mathcal{T}_\infty} \left| \frac{p_t}{p w_t} - 1 \right| \cdot J_1(x) + \max_{t \in \mathcal{T}_\infty} \left| \frac{p_t}{p w_t} - 1 \right| \xrightarrow{\mathbb{P}} 0, \end{aligned}$$

uniformly for $c \leq x \leq Q(1-a)$.

The stability condition follows readily. $\square$

References

- Allen, L., T. G. Bali, and Y. Tang (2012). Does systemic risk in the financial sector predict future economic downturns? *The Review of Financial Studies* 25(10), 3000–3036.
- Atkinson, A. B. (2005). Top incomes in the UK over the 20th century. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 168(2), 325–343.
- Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science* 293(5536), 1818–1820.
- Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *The Annals of Probability* 2(5), 792–804.

- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science* 286(5439), 509–512.
- Barro, R. J. and T. Jin (2011). On the size distribution of macroeconomic disasters. *Econometrica* 79(5), 1567–1589.
- Chang, L.-C. (1964). On the ratio of the empirical distribution function to the theoretical distribution function. *Selected Transl. in Math. Statist. and Probab. Amer. Math. Soc.* 4, 17–38.
- Clauset, A., C. Shalizi, and M. Newman (2009). Power-law distributions in empirical data. *SIAM Review* 51(4), 661–703.
- de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. New York: Springer.
- di Giovanni, J. and A. A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy* 120(6), 1083–1132.
- di Giovanni, J., A. A. Levchenko, and R. Rancière (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics* 85(1), 42 – 52.
- Einmahl, J. H. J., L. de Haan, and C. Zhou (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78(1), 31–51.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance*. Berlin: Springer.
- Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020). Heterogeneity and persistence in returns to wealth. *Econometrica* 88(1), 115–170.
- Fama, E. F. (1963). Mandelbrot and the stable Paretian hypothesis. *The Journal of Business* 36(4), 420–429.
- Feller, W. (1978). *An Introduction to Probability Theory and Its Applications*, Volume 2. New York: Wiley.
- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 24, pp. 180–190. Cambridge University Press.

- Fisk, P. R. (1961). The graduation of income distributions. *Econometrica* 29(2), 171–185.
- Gabaix, X. (1999). Zipf’s law for cities: an explanation. *The Quarterly Journal of Economics* 114(3), 739–767.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics* 1(1), 255–294.
- Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives* 30(1), 185–206.
- Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2006). Institutional investors and stock market volatility. *The Quarterly Journal of Economics* 121(2), 461–504.
- Gabaix, X. and Y. M. Ioannides (2004). The evolution of city size distributions. In J. V. Henderson and J.-F. Thisse (Eds.), *Cities and Geography*, Volume 4 of *Handbook of Regional and Urban Economics*, pp. 2341 – 2378. Elsevier.
- Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *The Quarterly Journal of Economics* 123(1), 49–100.
- Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The dynamics of inequality. *Econometrica* 84(6), 2071–2111.
- Gardes, L. (2015). A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes. *Extremes* 18(3), 479–510.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics* 44(3), 423–453.
- Gopikrishnan, P., V. Plerou, X. Gabaix, and H. E. Stanley (2000). Statistical properties of share volume traded in financial markets. *Physical Review E* 62, R4493–R4496.
- Gopikrishnan, P., V. Plerou, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley (1999). Scaling of the distribution of fluctuations of financial market indices. *Physical Review E* 60, 5305–5316.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3(5), 1163–1174.

- Hüsler, J. (1986). Extreme values of non-stationary random sequences. *Journal of Applied Probability* 23(4), 937–950.
- Jackson, M. O. (2009). Networks and economic behavior. *Annual Review of Economics* 1(1), 489–511.
- Jansen, D. W. and C. G. de Vries (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics* 73(1), 18–24.
- Jessen, A. H. and T. Mikosch (2006). Regularly varying functions. *Publications de l’Institut Mathématique* 80, 171–192.
- Kelly, B. and H. Jiang (2014). Tail risk and asset prices. *The Review of Financial Studies* 27(10), 2841–2871.
- Kyle, A. S. and A. A. Obizhaeva (2016). Market microstructure invariance: Empirical hypotheses. *Econometrica* 84(4), 1345–1404.
- Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and related properties of random sequences and processes*. New York: Springer.
- Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *The Journal of Business* 69(3), 383–408.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36(4), 394–419.
- Newman, M. E., A.-L. E. Barabási, and D. J. Watts (2006). *The Structure and Dynamics of Networks*. Princeton University Press.
- Pareto, V. (1896). *Cours d’Économie Politique*. F. Rouge, Lausanne.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1), 119–131.
- Piketty, T. and A. Goldhammer (2014). *Capital in the Twenty-First Century*. Harvard University Press.
- Piketty, T. and E. Saez (2003). Income inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118(1), 1–41.

- Plerou, V. and H. E. Stanley (2007). Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets. *Physical Review E* 76, 046109.
- Potter, H. (1942). The mean values of certain Dirichlet series, ii. *Proceedings of the London Mathematical Society* 2(1), 1–19.
- Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.
- Stanley, M. H., S. V. Buldyrev, S. Havlin, R. N. Mantegna, M. A. Salinger, and H. Eugene Stanley (1995). Zipf plots and the size distribution of firms. *Economics Letters* 49(4), 453 – 457.
- Toda, A. A. and K. Walsh (2015). The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy* 123(5), 1177–1200.
- Weissman, I. (1978). Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Association* 73(364), 812–815.
- Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 45(1), 73–88.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison Wesley Press.
- Zucman, G. (2019). Global wealth inequality. *Annual Review of Economics* 11(1), 109–138.