Inference on Extreme Quantiles of Unobserved Individual Heterogeneity

Vladislav Morozov

UPF and BSE

Motivating Example

Two panels: firms in denser areas (“cities”) and firms in less dense areas (“countryside”). Firms differ in their productivity θ_i.

Conjecture: competition is stronger in cities (Asplund and Nocke, 2006). Therefore, minimal productivity needed to survive in the city is higher than minimal productivity needed to survive in the countryside.

Null: $F^{-1}_{Cities}(0) - F^{-1}_{Countryside}(0) \geq 0$.

Issue: firm-level TFP θ_i is not observed and has to be estimated from the data

Motivating Example

Two panels: firms in denser areas ("cities") and firms in less dense areas ("countryside"). Firms differ in their productivity θ_i.

Conjecture: competition is stronger in cities (Asplund and Nocke, 2006). Therefore, minimal productivity needed to survive in the city is higher than minimal productivity needed to survive in the countryside.

Null: $F_{Cities}^{-1}(0) - F_{Countryside}^{-1}(0) \geq 0$.

Issue: firm-level TFP θ_i is not observed and has to be estimated from the data

Setting and Objective

Setting: panel data (not necessarily cross-section $\times$ time structure) or meta-analysis. Units i differ in some unobserved heterogeneity θ_i. θ_i may be

1. Heterogeneous parameters of linear or nonlinear models. Example: $y_{it} = m(x_{it}, \theta_i) + u_{it}$ where m is a known function and θ_i is a scalar.
2. Nonparametric objects: if $y_{it} = m_i(x_{it}) + u_{it}$, then θ_i may be value of m_i at a point x_0: $\theta_i = m_i(x_0)$.

Object of interest: we wish to

1. Conduct inference on quantiles of θ_i that are close to zero or one – extreme quantiles
2. Test if support of θ_i lies inside $(-\infty, C]$, $[c, \infty)$, or $[c, C]$ + construct confidence intervals for the support. This a hypothesis about the theoretical maximum/minimum value of θ_i:

Setting and Objective

Setting: panel data (not necessarily cross-section $\times$ time structure) or meta-analysis. Units i differ in some unobserved heterogeneity θ_i. θ_i may be

1. Heterogeneous parameters of linear or nonlinear models. Example: $y_{it} = m(x_{it}, \theta_i) + u_{it}$ where m is a known function and θ_i is a scalar.
2. Nonparametric objects: if $y_{it} = m_i(x_{it}) + u_{it}$, then θ_i may be value of m_i at a point x_0: $\theta_i = m_i(x_0)$.

Object of interest: we wish to

1. Conduct inference on quantiles of θ_i that are close to zero or one – extreme quantiles
2. Test if support of θ_i lies inside $(-\infty, C]$, $[c, \infty)$, or $[c, C]$ + construct confidence intervals for the support. This a hypothesis about the theoretical maximum/minimum value of θ_i:

Setting and Objective

Setting: panel data (not necessarily cross-section $\times$ time structure) or meta-analysis. Units i differ in some unobserved heterogeneity θ_i. θ_i may be

1. Heterogeneous parameters of linear or nonlinear models. Example: $y_{it} = m(x_{it}, \theta_i) + u_{it}$ where m is a known function and θ_i is a scalar.
2. Nonparametric objects: if $y_{it} = m_i(x_{it}) + u_{it}$, then θ_i may be value of m_i at a point x_0: $\theta_i = m_i(x_0)$.

Object of interest: extreme quantiles of θ_i.
Challenge: θ_i have to be estimated with noise from individual data.

Objective

Conduct inference on extreme quantiles of θ_i given only noisy observations.

Economic Examples

Many examples proceed from recent papers considering the distribution of individual heterogeneity:

- Heterogeneous productivity of firms (Combes et al., 2012a)
- Worker productivity (Combes et al., 2012b; Eeckhout et al., 2014; de la Roca and Puga, 2017)
- Relationship between schooling and health (Auld and Sidhu, 2005)
- Role of skill and scale in mutual funds (Barras et al., 2021)

Challenge

The big issue: we do not observe θ_i directly. Coefficients have to be estimated from data, we only have noisy 'observations'. Noise shrinks in panel length T

- If we observed θ_i directly, the solution is known completely from extreme value literature (de Haan and Ferreira, 2006)
- If we were interested in a "central" quantile of θ_i, the problem of noise is solved by Jochmans and Weidner (2022): quantile of noisy distribution is normal + additional bias terms to account for noise

We care about an extreme quantile under a noisy setting: this paper

Challenge

The big issue: we do not observe θ_i directly. Coefficients have to be estimated from data, we only have noisy 'observations'. Noise shrinks in panel length T

- If we observed θ_i directly, the solution is known completely from extreme value literature (de Haan and Ferreira, 2006)
- If we were interested in a "central" quantile of θ_i, the problem of noise is solved by Jochmans and Weidner (2022): quantile of noisy distribution is normal + additional bias terms to account for noise

We care about an extreme quantile under a noisy setting: this paper

Setup

Suppose that θ_i are iid from cdf F. We are interested in conducting inference on $F^{-1}(1-\delta)$.

We do not observe θ_i, but we observe noisy estimates

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\varepsilon_{i,T}, \quad i = 1, \ldots, N \tag{1}$$

where p is a known convergence rate and $\varepsilon_{i,T} = O_p(1)$ and has cdf G_T. Setup similar to one of Jochmans and Weidner (2022).
Example: ϑ_i is the OLS estimator in $y_{it} = \theta_i' x_{it} + u_{it}$. Then

$$\vartheta_i = \theta_i + \underbrace{\left(T^{-1}\sum\nolimits_t x_{it}x_{it}'\right)^{-1} T^{-1}\sum\nolimits_t x_{it}u_{it}}_{\frac{1}{T^{1/2}}\varepsilon_{i,T}} \tag{2}$$

Setup

Suppose that θ_i are iid from cdf F. We are interested in conducting inference on $F^{-1}(1-\delta)$.

We do not observe θ_i, but we observe noisy estimates

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\varepsilon_{i,T}, \quad i = 1, \dots, N \tag{1}$$

where p is a known convergence rate and $\varepsilon_{i,T} = O_p(1)$ and has cdf G_T. Setup similar to one of Jochmans and Weidner (2022).

Example: ϑ_i is the OLS estimator in $y_{it} = \theta_i' x_{it} + u_{it}$. Then

$$\vartheta_i = \theta_i + \underbrace{\left(T^{-1}\sum_t x_{it}x_{it}'\right)^{-1} T^{-1}\sum_t x_{it}u_{it}}_{\frac{1}{T^{1/2}}\varepsilon_{i,T}} \tag{2}$$

Setup

Suppose that θ_i are iid from cdf F. We are interested in conducting inference on $F^{-1}(1-\delta)$.

We do not observe θ_i, but we observe noisy estimates

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\varepsilon_{i,T}, \quad i = 1, \dots, N \tag{1}$$

where p is a known convergence rate and $\varepsilon_{i,T} = O_p(1)$ and has cdf G_T. Setup similar to one of Jochmans and Weidner (2022).

Example: ϑ_i is the OLS estimator in $y_{it} = \theta_i' x_{it} + u_{it}$. Then

$$\vartheta_i = \theta_i + \underbrace{\left(T^{-1}\sum\nolimits_t x_{it}x_{it}'\right)^{-1} T^{-1}\sum\nolimits_t x_{it}u_{it}}_{\frac{1}{T^{1/2}}\varepsilon_{i,T}} \tag{2}$$

Assumptions

θ_i and $\varepsilon_{i,T}$ can have complicated dependence structure $\Rightarrow$ we impose no assumptions on joint distribution.

Global maintained assumptions are minimal and on marginals only

- F satisfies an extreme value theorem – an assumption that F is regularly varying with extreme value index γ.
    1. $\gamma < 0$: finite endpoint $F^{-1}(1) < \infty$
    2. $\gamma > 0$: infinite endpoint $F^{-1}(1) = \infty$, distribution behaves like a power law
    3. $\gamma = 0$: $F^{-1}(1)$ may be finite or infinite, tails are "light"
- G_T (law of $\varepsilon_{i,T}$) form a tight family indexed by T

EVT assumption: makes inference possible if we saw the original data.

Related Literature

- In treatment effect literature endpoints of impact are of interest (Heckman et al., 1997; Manski, 1990). Fan and Park (2010) show how to conduct pointwise inference on quantiles of the bounds for binary treatment. Zhang (2018) examines extreme treatment effects in a cross-sectional setting
- Extremal quantile regression (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011) looks at extreme conditional quantiles of data. We consider a mean regression and look at quantiles of individual effect. In general there is no direct relation between the two
- Not restricting the joint distribution of θ_i and $\varepsilon_{i,T}$ means deconvolution is not an option for obtaining the distribution of θ_i (like in Evdokimov (2010); Arellano and Bonhomme (2012))

Road Map and Results

When are the noisy observables informative about the quantiles of interest?

- We give sharp conditions under which an extreme value theorem holds for noisy observables $\vartheta_{i,T}$
- These conditions are implicit rate restrictions on panel size, and we provide sufficient conditions for them to hold

Steps to construct the tests and CIs of interest:

- We provide a 'feasible' extreme value theorem suitable for inference and show how to estimate quantiles of the limit distribution with subsampling and by simulation
- We propose CIs and tests about extreme quantiles using extreme order approximations.
- Application to firm productivity in areas of above- and below-median density (Combes et al., 2012a)
- Extra: feasible intermediate extreme value theorem and intermediate extreme inference. Estimation of extreme value parameter γ

Distribution of Noisy Maximum I

Define the noisy and the noiseless maxima

$$\vartheta_{N,N,T} = \max\{\vartheta_{1,T}, \dots, \vartheta_{N,T}\}, \quad \theta_{N,N} = \max\{\theta_1, \dots, \theta_N\} \tag{3}$$

Under our assumptions for some sequence a_N, b_N the noiseless maximum satisfies an extreme value theorem (EVT)

$$a_N^{-1}(\theta_{N,N} - b_N) \Rightarrow Q, \quad N \to \infty \tag{4}$$

where Q is a GEV(γ) random variable.

We seek conditions for

$$a_N^{-1}(\vartheta_{N,N,T} - b_N) \Rightarrow Q, \quad , N, T \to \infty \tag{5}$$

When can we apply noiseless inference theory to the noisy case?

Proposition (Extreme value theorem for maximum of noisy observations)

Let $\varepsilon_{i,T} \sim G_T$, $\theta_i \sim F$, and $\vartheta_{i,T} = \theta_i + T^{-p}\varepsilon_{i,T}$. If a_N, b_N are such that as $N \to \infty$

$$a_N^{-1}(\theta_{N,N} - b_N) \Rightarrow Q \tag{6}$$

and for each $\tau \in (0,\infty)$ let the tail equivalence conditions hold:

$$\sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1-\frac{1}{N\tau}-u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right) \to 0$$

$$\inf_{u\in\left[1-\frac{1}{N\tau},1\right]} \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(2-\frac{1}{N\tau}-u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right) \to 0$$

then as $N, T \to \infty$

$$a_N^{-1}(\vartheta_{N,N,T} - b_N) \Rightarrow Q \tag{7}$$

Distribution of Noisy Maximum: Discussion

The inf and sup conditions of the previous theorem impose tail equivalence between the noiseless distribution of θ and the noisy distribution ϑ_i:

- Equivalence required to hold in a weak pointwise sense (for each τ separately)
- Intuitive explanation: the $F^{-1} + T^{-p}G_T^{-1}$ terms are approximately the quantiles of the noisy estimates; the inf and sup adjusting for the unknown joint distribution of θ_i and $\varepsilon_{i,T}$. The $-F^{-1}$ term corresponds to the noiseless quantiles. If the two are the same in the limit, noisy estimates $\vartheta_{i,T}$ are informative about the true quantiles of interest.

The sup and inf conditions are sharp: if they fail, we can construct a joint distribution of $(\theta_i, \varepsilon_{i,T})$ such that $\vartheta_{N,N,T}$ has a different limit distribution or does not converge at all.

Distribution of Noisy Maximum: Discussion

The inf and sup conditions of the previous theorem impose tail equivalence between the noiseless distribution of θ and the noisy distribution ϑ_i:

- Equivalence required to hold in a weak pointwise sense (for each τ separately)
- Intuitive explanation: the $F^{-1} + T^{-p}G_T^{-1}$ terms are approximately the quantiles of the noisy estimates; the inf and sup adjusting for the unknown joint distribution of θ_i and $\varepsilon_{i,T}$. The $-F^{-1}$ term corresponds to the noiseless quantiles. If the two are the same in the limit, noisy estimates $\vartheta_{i,T}$ are informative about the true quantiles of interest.

The sup and inf conditions are sharp: if they fail, we can construct a joint distribution of $(\theta_i, \varepsilon_{i,T})$ such that $\vartheta_{N,N,T}$ has a different limit distribution or does not converge at all.

When do Inf and Sup Conditions Hold? Rate Restriction on (N, T)

The sup and inf conditions are an implicit rate restriction, which depends on F and G_T.

Suppose that F has EV index γ and $\mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$. For the noisy EVT to hold it's sufficient that

$$\frac{N^{1/\beta-\gamma}(\log T)^{1/\beta}}{T^p} \to 0. \tag{8}$$

Examples: let $p = 1/2$ (parametric rate)

1. If $\gamma > 1/\beta$ (F is heavy-tailed relative to G_T), then there is no restriction: the heavy tail of θ dominates the lighter tail of noise
2. If θ is normal and $\beta = 8$, then $\gamma = 0$, then condition is $\frac{N \log^4 T}{T^2} \to 0$
3. If $\theta \sim \text{Uniform}[0, \theta_F]$ and $\beta = 8$ (eight moments), then $\gamma = -1$, and the condition is $\frac{N^9(\log T)}{T^4} \to 0$

When do Inf and Sup Conditions Hold? Rate Restriction on (N, T)

The sup and inf conditions are an implicit rate restriction, which depends on F and G_T.

Suppose that F has EV index γ and $\mathbb{E}|\varepsilon_{i,T}|^{\beta} < \infty$. For the noisy EVT to hold it's sufficient that

$$\frac{N^{1/\beta-\gamma}(\log T)^{1/\beta}}{T^p} \to 0. \tag{8}$$

Examples: let $p = 1/2$ (parametric rate)

1. If $\gamma > 1/\beta$ (F is heavy-tailed relative to G_T), then there is no restriction: the heavy tail of θ dominates the lighter tail of noise
2. If θ is normal and $\beta = 8$, then $\gamma = 0$, then condition is $\frac{N\log^4 T}{T^2} \to 0$
3. If $\theta \sim \text{Uniform}[0, \theta_F]$ and $\beta = 8$ (eight moments), then $\gamma = -1$, and the condition is $\frac{N^9(\log T)}{T^4} \to 0$

Sufficient Condition For Rate Conditions

No global moment condition is possible for noisy EVT because the normalizing rate depends on F and is unknown!

We can provide a uniform guarantee if there's a bound on γ:

Proposition

Let one of the following conditions hold:

1. Let $\sup_T \mathbb{E}|\varepsilon_{i,T}|^{\beta} < \infty$ for some $\beta > 0$, and let $N^{1/\beta-\gamma'}(\log(T))^{1/\beta}/T^p \to 0$ for some γ'.
2. For all T, let $\varepsilon_{i,T}$ be normal and let $N^{-\gamma'}\sqrt{\log(N)}/T^p \to 0$ for some γ'.

Let F have EV index $\gamma > \gamma'$. Then the tail equivalence conditions hold for F and G_T.

If F behaves likes a power law (Gabaix, 2009, 2016), take $\gamma' = 0$.

Towards Inference: Choosing Normalizing Constants

We now know when $a_N^{-1}(\vartheta_{N,N,T} - b_N)$ converges in distribution to the same limit as $a_N^{-1}(\theta_{N,N} - b_N)$.

What are (a_N, b_N)? limits with specific choices

Choice of b_N is crucial and reflects the parameter of interest. Let $l \geq 0$ be a parameter chosen by us. Set

$$b_N = F^{-1}\left(1 - \frac{l}{N}\right) \tag{9}$$

We model the quantile of interest as drifting to the right at a rate N^{-1}.

Practical interpretation: suppose that $N = 100$ and we set $l = 5$. Then $b_N = F^{-1}(0.95)$. In a sample of 100 units this choice of l means doing inference on the 95th percentile.

Towards Inference: Scaling Constants a_N and Joint EVT

The constants a_N are unknown and generally depend on the extreme quantiles of F and cannot be estimated. example forms of a_N Instead we use will get rid of a_N by using a vector of top k statistics.

Lemma

Suppose that tail equivalence holds. Let $\vartheta_{1,N,T} \leq \cdots \leq \vartheta_{N,N,T}$ be order statistics. Let $k < \infty$ be fixed. There exist constants a_N, c_N such that as $N, T \to \infty$

$$\left(\frac{\vartheta_{N,N,T} - c_N}{a_N}, \frac{\vartheta_{N-1,N,T} - c_N}{a_N}, \ldots, \frac{\vartheta_{N-k,N,T} - c_N}{a_N}\right) \Rightarrow \operatorname{sgn}(\gamma)\left((E_1^*)^{-\gamma}, (E_1^* + E_2^*)^{-\gamma}, \ldots, (E_1^* + E_2^* + \cdots + E_k^*)^{-\gamma}\right) \quad (10)$$

where $E_1^*, \ldots, E_k^*$ are iid standard exponential.

Towards Inference: Scaling Constants a_N and Joint EVT

The constants a_N are unknown and generally depend on the extreme quantiles of F and cannot be estimated. example forms of a_N Instead we use will get rid of a_N by using a vector of top k statistics.

Lemma

Suppose that tail equivalence holds. Let $\vartheta_{1,N,T} \leq \cdots \leq \vartheta_{N,N,T}$ be order statistics. Let $k < \infty$ be fixed. There exist constants a_N, c_N such that as $N, T \to \infty$

$$\left(\frac{\vartheta_{N,N,T} - c_N}{a_N}, \frac{\vartheta_{N-1,N,T} - c_N}{a_N}, \dots, \frac{\vartheta_{N-k,N,T} - c_N}{a_N}\right) \Rightarrow \operatorname{sgn}(\gamma)\left((E_1^*)^{-\gamma}, (E_1^* + E_2^*)^{-\gamma}, \dots, (E_1^* + E_2^* + \cdots + E_k^*)^{-\gamma}\right) \quad (10)$$

where $E_1^*, \dots, E_k^*$ are iid standard exponential.

Feasible Noisy EVT

We get a result suitable for inference by taking the ratio of order statistics and using our choice of b_N for centering.

Theorem

Let tail equivalence hold. let $\vartheta_{N-k,N,T}$ be the $(k+1)$st largest ϑ_i and $l \geq 0$ be fixed, let $r \geq 0, q \geq 1$ be natural numbers. Then

$$\frac{\vartheta_{N-r,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^* + \cdots + E_{r+1}^*)^{-\gamma} - l^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}, \tag{11}$$

where $E_1^*, E_2^*, \ldots$ are independent standard exponential RVs.

Feasible Noisy EVT

Theorem

Let tail equivalence hold. let $\vartheta_{N-k,N,T}$ be the $(k+1)$st largest ϑ_i and $l \geq 0$ be fixed, let $r \geq 0, q \geq 1$ be natural numbers. Then

$$\frac{\vartheta_{N-r,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^* + \cdots + E_{r+1}^*)^{-\gamma} - l^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}, \tag{11}$$

where $E_1^*, E_2^*, \ldots$ are independent standard exponential RVs.

The statistic on the left does not depend on γ and the only unknown is the object of interest! Limit distribution depends on γ, but critical values can be obtained by subsampling, and we prove consistency. (We also provide consistent estimators of γ).

Subsampling γ estimators

Confidence Intervals

Statistic: $\frac{\vartheta_{N-r,N,T} - F^{-1}(1-l/N)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^* + \cdots + E_{r+1}^*)^{-\gamma} - l^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}$. Let $\hat{c}_\alpha$ be the estimated αth quantile of the limit.

Confidence intervals for $F^{-1}(1-l/N)$:

$$CI_\alpha = \left[\vartheta_{N-r,N,T} - \hat{c}_{1-\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right), \vartheta_{N-r,N,T} - \hat{c}_{\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)\right]$$

1. Interpretation: asymptotic $(1-\alpha)$ CI for $F^{-1}(1-l/N)$ – object of interest shifting with N.
2. What $F^{-1}(1-l/N)$ is depends on N and l. Example: if $N = 200$ and $l = 10$, then CI_α is an interval for $F^{-1}(0.95)$.
3. Picking r and q: may be any, but: q in the range 2-10 seems to work well with little difference. Leading choices for r: 0 (centered at the maximum) or l (centered at the corresponding $(1-l/N)$th sample quantile)

Confidence Intervals

Statistic: $\frac{\vartheta_{N-r,N,T} - F^{-1}(1-I/N)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^* + \cdots + E_{r+1}^*)^{-\gamma} - I^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}$. Let $\hat{c}_\alpha$ be the estimated αth quantile of the limit.

Confidence intervals for $F^{-1}(1 - I/N)$:

$$CI_\alpha = \left[\vartheta_{N-r,N,T} - \hat{c}_{1-\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right), \vartheta_{N-r,N,T} - \hat{c}_{\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)\right]$$

1. Interpretation: asymptotic $(1-\alpha)$ CI for $F^{-1}(1 - I/N)$ – object of interest shifting with N.
2. What $F^{-1}(1 - I/N)$ is depends on N and I. Example: if $N = 200$ and $I = 10$, then CI_α is an interval for $F^{-1}(0.95)$.
3. Picking r and q: may be any, but: q in the range 2-10 seems to work well with little difference. Leading choices for r: 0 (centered at the maximum) or I (centered at the corresponding $(1 - I/N)$th sample quantile)

Confidence Intervals

Statistic: $\frac{\vartheta_{N-r,N,T}-F^{-1}(1-l/N)}{\vartheta_{N-q,N,T}-\vartheta_{N,N,T}} \Rightarrow \frac{(E_1^*+\cdots+E_{r+1}^*)^{-\gamma}-l^{-\gamma}}{(E_1^*+\cdots+E_{q+1}^*)^{-\gamma}-(E_1^*)^{-\gamma}}$. Let $\hat{c}_\alpha$ be the estimated αth quantile of the limit.

Confidence intervals for $F^{-1}(1-l/N)$:

$$CI_\alpha = \left[\vartheta_{N-r,N,T} - \hat{c}_{1-\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right), \vartheta_{N-r,N,T} - \hat{c}_{\alpha/2}\left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)\right]$$

1. Interpretation: asymptotic $(1-\alpha)$ CI for $F^{-1}(1-l/N)$ – object of interest shifting with N.
2. What $F^{-1}(1-l/N)$ is depends on N and l. Example: if $N=200$ and $l=10$, then CI_α is an interval for $F^{-1}(0.95)$.
3. Picking r and q: may be any, but: q in the range 2-10 seems to work well with little difference. Leading choices for r: 0 (centered at the maximum) or l (centered at the corresponding $(1-l/N)$th sample quantile)

Estimators for Extreme Quantiles

We can also construct an estimator for extreme quantiles that improves on the simple sample quantiles.

By the feasible EVT

$$P\left(\frac{\vartheta_{N-r,N,T} - F^{-1}(1 - l/N)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \leq \hat{c}_{1/2}\right) \to 1/2 \tag{12}$$

Rearranging, we obtain the median-unbiased estimator for $F^{-1}(1 - l/N)$

$$\mathcal{M}_{N,T} = \vartheta_{N-r,N,T} - \hat{c}_{1/2}(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}) \tag{13}$$

Tests About Support

We can also test hypotheses about support of θ and quantiles.
For example: $H_0 : F^{-1}(1) \leq C$ vs. $H_1 : F^{-1}(1) > C$ (if $\gamma < 0$ so that $F^{-1}(1) < \infty$ under both null and alternative) example setting

$$W_C = \frac{\vartheta_{N,N,T} - C}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \xRightarrow{H_0} \frac{(E_1^*)^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}.$$

Decision rule: reject if $W_C < \hat{c}_\alpha$. Consistent against fixed alternatives

Monte Carlo: Setup

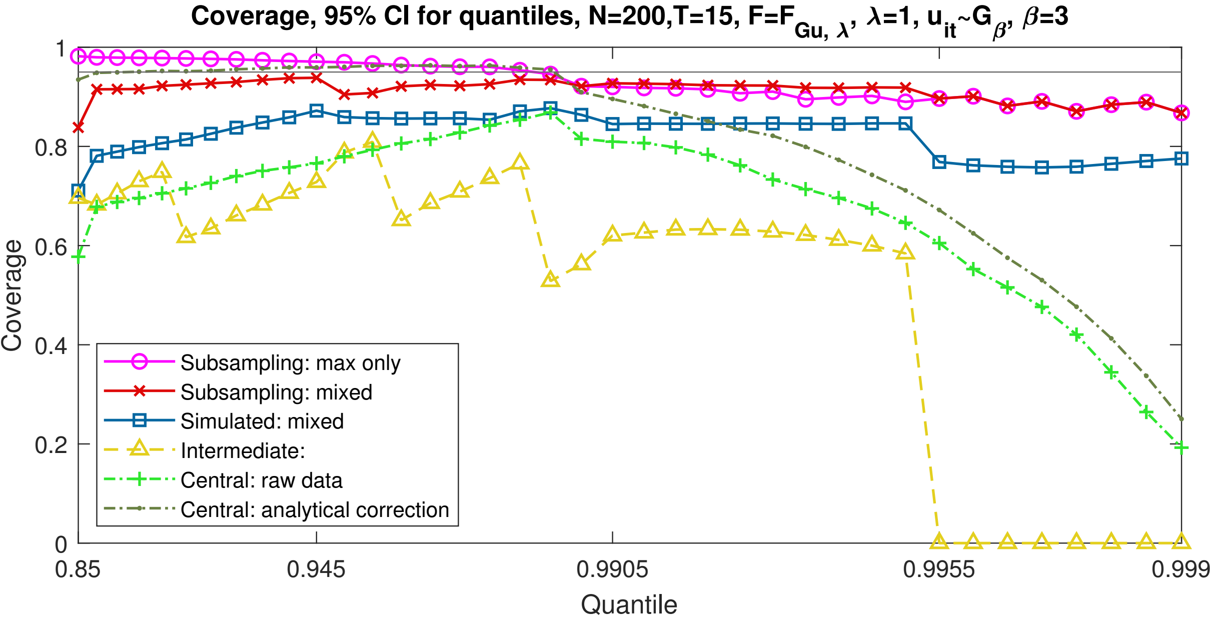
Setup: linear regression

$$y_{it} = \alpha_i + \beta_i x_{it} + \theta_i z_{it} + u_{it} \tag{14}$$

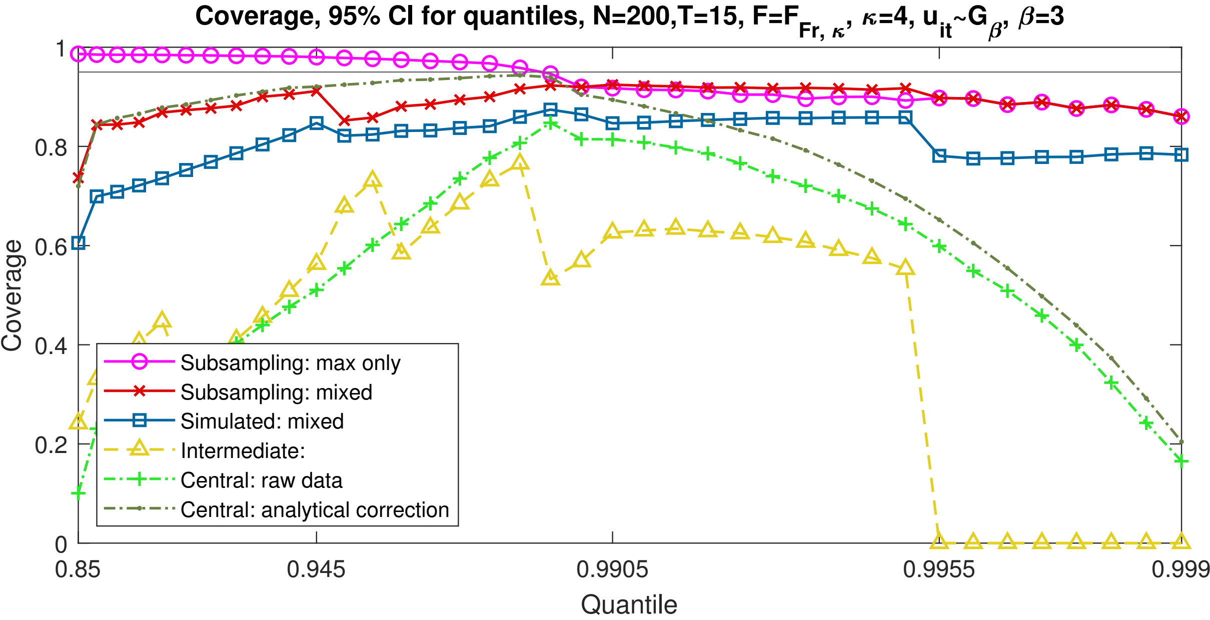
1. Canonical examples of distributions for θ_i: exponential, power law, finite endpoint.
2. u_{it} distributed as heavy-tailed power law with 3 moments (also normally distributed in the paper).

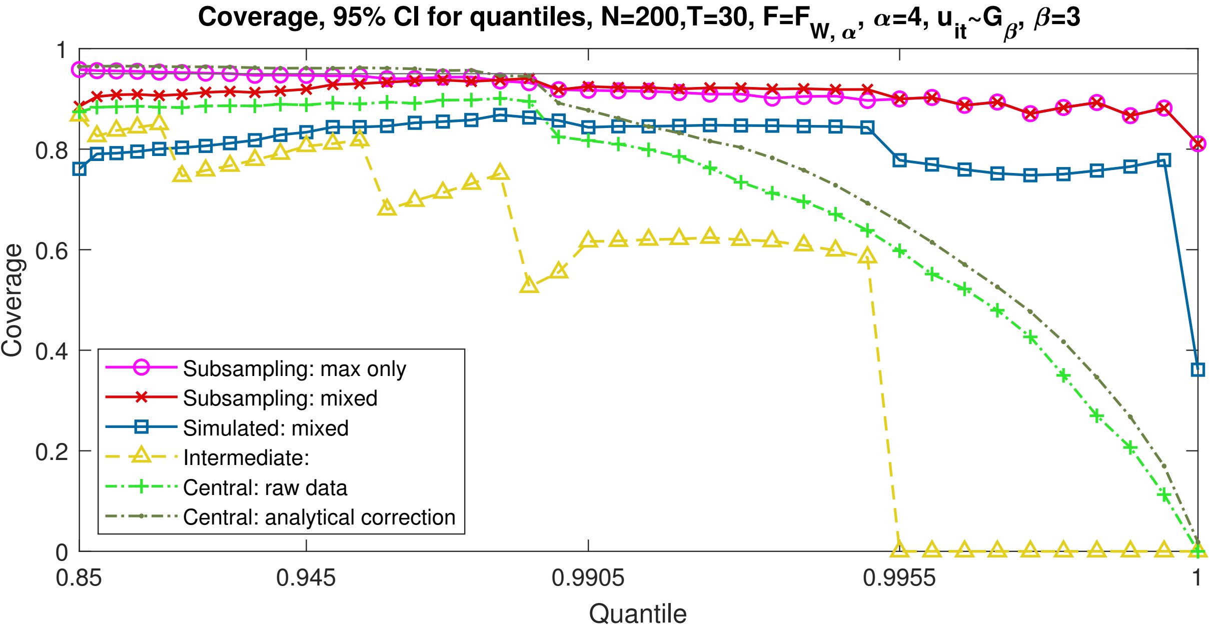
We compare:

1. Extreme approximations: based on the feasible EVT. Different choices of centering order statistics (at corresponding sample quantile or at sample maximum) and different methods of estimating critical values.
2. Intermediate extreme: using feasible intermediate extreme theorem (in the paper)
3. Central: using raw data and applying analytical bias correction (Jochmans and Weidner, 2022)



Coverage, 95% CI for quantiles, N=200,T=15, F=F_{Gu, λ}, λ=1, u_{it}~G_β, β=3
Coverage
Quantile
0
0.2
0.4
0.6
0.8
1
0.85
0.945
0.9905
0.9955
0.999
Subsampling: max only
Subsampling: mixed
Simulated: mixed
Intermediate:
Central: raw data
Central: analytical correction


Coverage, 95% CI for quantiles, N=200,T=15, F=F_{Fr, κ}, κ=4, u_{it}~G_β, β=3
Coverage
Quantile
1
0.8
0.6
0.4
0.2
0
0.85
0.945
0.9905
0.9955
0.999
Subsampling: max only
Subsampling: mixed
Simulated: mixed
Intermediate:
Central: raw data
Central: analytical correction



Coverage, 95% CI for quantiles, N=200,T=30, F=F$_{W, \alpha}$, α=4, u$_{it}$~G$_{\beta}$, β=3
Coverage
Quantile
1
0.8
0.6
0.4
0.2
0
0.85
0.945
0.9905
0.9955
1
Subsampling: max only
Subsampling: mixed
Simulated: mixed
Intermediate:
Central: raw data
Central: analytical correction

Application to Firm Productivity: Background

We illustrate our methods by revisiting the setting of Combes et al. (2012a).

Combes et al. (2012a): firms are more productive in more densely populated areas. They differentiate between two key effects driving productivity difference:

1. Agglomeration economies (sharing suppliers, deeper labor market, learning from others).
2. Firm selection: larger markets have tougher competition – minimal productivity to survive is higher

Key assumption of Combes et al. (2012a): the productivity distribution is the same between areas of above- and below-median employment density, except for three parameters: mean and variance (due to agglomeration), proportion of the left tail truncated (due to selection).

Application: Goals

We nonparametrically examine:

1. Evidence for firm selection:
   - Whether truncation is present in the data – sign and magnitude of γ. Positive values would imply $F^{-1}(0) = -\infty$ – no sharp truncation in data. Negative estimated values of γ would imply finite $F^{-1}(0)$ – sharp threshold required for firm survival.
   - Differences between left tails of distributions between areas – least productive firms
2. Whether the three parameters of Combes et al. (2012a) (mean, variance, truncation) are sufficient to explain the differences between tails of the distributions.

Data: first-stage estimates of TFP θ_i in value-added Cobb-Double production functions of the form $V_{i\,t} = \exp(\theta_i + u_{i\,t})K_{i\,t}^{\beta_K} L_{i\,t}^{\beta_L}$ $(T = 9,\ N \approx 70000)$

Application: Estimation of Productivity and Data

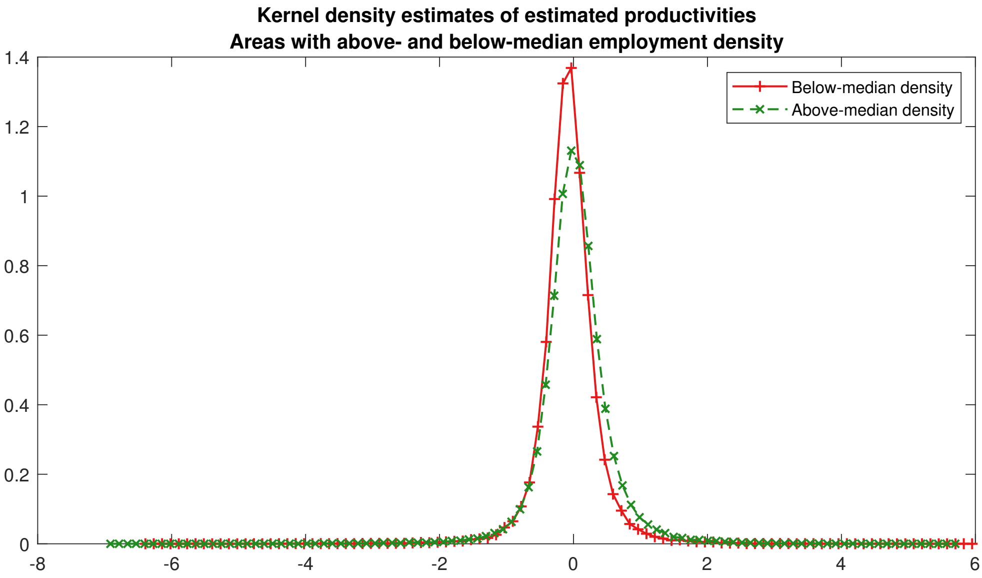
Value-added production function of the form:

$$V_{i\,t} = \exp(\theta_i) F(K_{i\,t}, L_{i\,t}, \text{Sector}) \tag{15}$$

$F(\cdot, \cdot, \text{Sector})$ is a sector-specific Cobb-Douglas function. $V_{i\,t}$ is value added by firm i in period t. θ_i is log TFP – key measure of productivity.
θ_i is estimated as an average residual from the production function:

$$\vartheta_{i,T} = \frac{1}{T} \sum_{t=1}^{T} \left[V_{i\,t} - \hat{F}(K_{i\,t}, L_{i\,t}, \text{Sector}) \right] \tag{16}$$

We use the estimated productivities provided by Combes et al. (2012a). $T = 9$, $N_1 \approx 78000$ (above-median employment density), $N_2 \approx 58000$ (below-median employment density).


Kernel density estimates of estimated productivities
Areas with above- and below-median employment density
Below-median density
Above-median density
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-8
-6
-4
-2
0
2
4
6

Application: Left Tail and Evidence on Selection

We look at the left tails of the raw distributions of estimated probabilities.

1. Estimate of γ is ≈ 0.25 for both denser and less dense areas. Consistent with very heavy-tailed distribution (only around 3-4 finite moments):
   - Serves as indirect confirmation that rate conditions hold in our example (recall: heavy-tailed θ + lighter-tailed estimation noise $\Rightarrow$ no rate condition). Of course, unless it is the noise that is very heavy-tailed.
   - Positive value of γ implies $F^{-1}(0) = -\infty$. No minimal threshold in data.
   - Complements the results of Combes et al. (2012a): they find no difference in truncation between below- and above-median employment density areas
   - $F^{-1}(0) = -\infty$ implies no sharp minimum. Instead it makes sense to look at a collection of quantiles $F^{-1}(\delta)$ for δ small.
2. We find statistically significant difference between left tails of two distributions.
   - For $\delta \leq 0.075$ quantiles of less dense areas uniformly higher – "worst firms of less-dense areas are better"
   - Possibly due to difference in sectoral composition

Application: Left Tail and Evidence on Selection

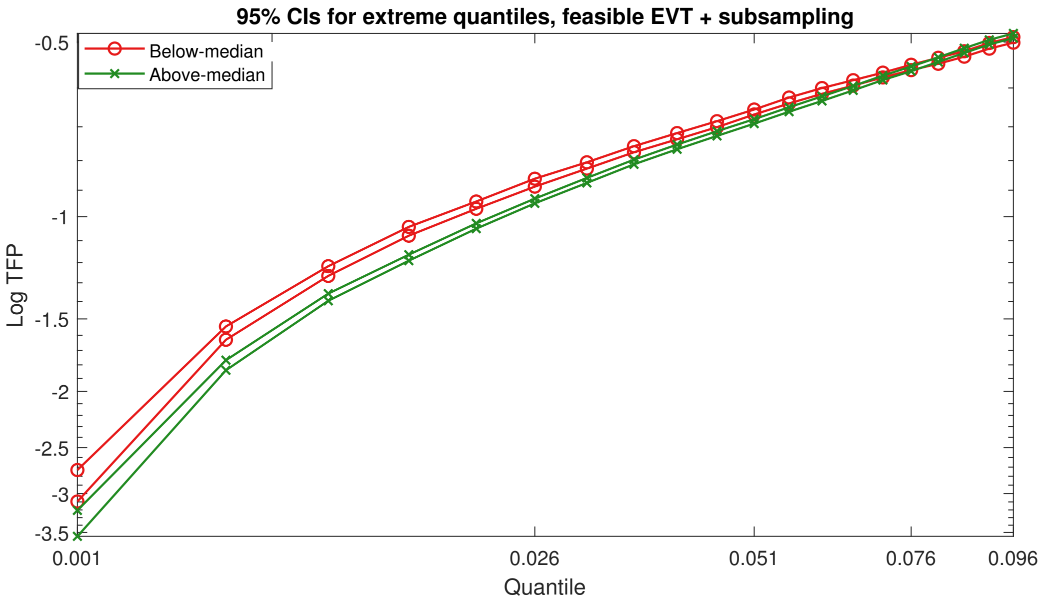
We look at the left tails of the raw distributions of estimated probabilities.

1. Estimate of γ is ≈ 0.25 for both denser and less dense areas. Consistent with very heavy-tailed distribution (only around 3-4 finite moments):
   - Positive value of γ implies $F^{-1}(0) = -\infty$. No minimal threshold in data.
   - Complements the results of Combes et al. (2012a): they find no difference in truncation between below- and above-median employment density areas
   - $F^{-1}(0) = -\infty$ implies no sharp minimum. Instead it makes sense to look at a collection of quantiles $F^{-1}(\delta)$ for δ small.
2. We find statistically significant difference between left tails of two distributions.
   - For $\delta \leq 0.075$ quantiles of less dense areas uniformly higher – "worst firms of less-dense areas are better"
   - Possibly due to difference in sectoral composition

Application: Left Tail and Evidence on Selection

We look at the left tails of the raw distributions of estimated probabilities.

1. Estimate of γ is ≈ 0.25 for both denser and less dense areas. Consistent with very heavy-tailed distribution (only around 3-4 finite moments):
   - Positive value of γ implies $F^{-1}(0) = -\infty$. No minimal threshold in data.
   - Complements the results of Combes et al. (2012a): they find no difference in truncation between below- and above-median employment density areas
   - $F^{-1}(0) = -\infty$ implies no sharp minimum. Instead it makes sense to look at a collection of quantiles $F^{-1}(\delta)$ for δ small.
2. We find statistically significant difference between left tails of two distributions.
   - For $\delta \leq 0.075$ quantiles of less dense areas uniformly higher – "worst firms of less-dense areas are better"
   - Possibly due to difference in sectoral composition

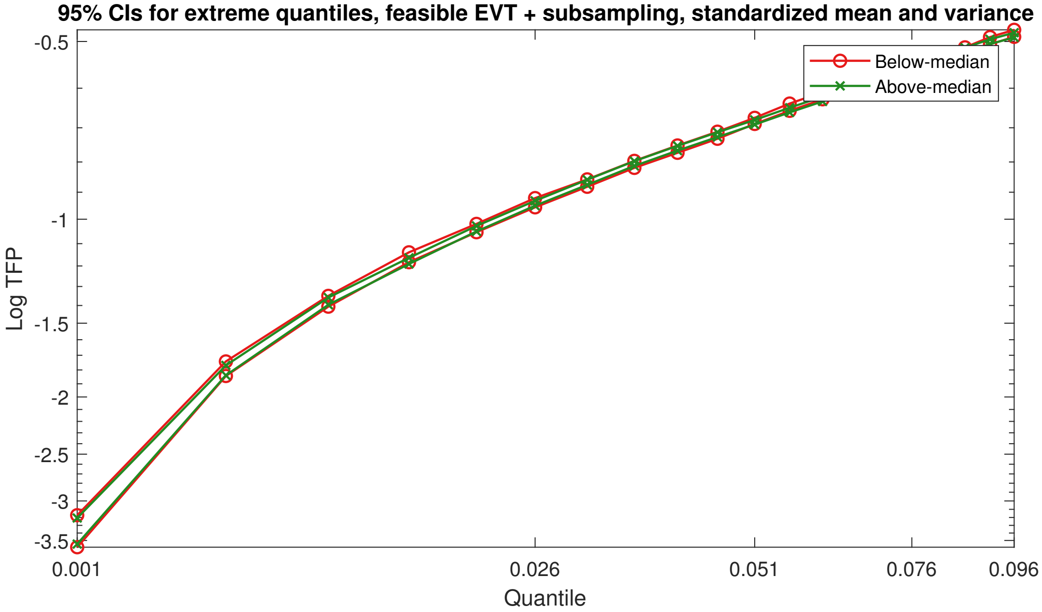


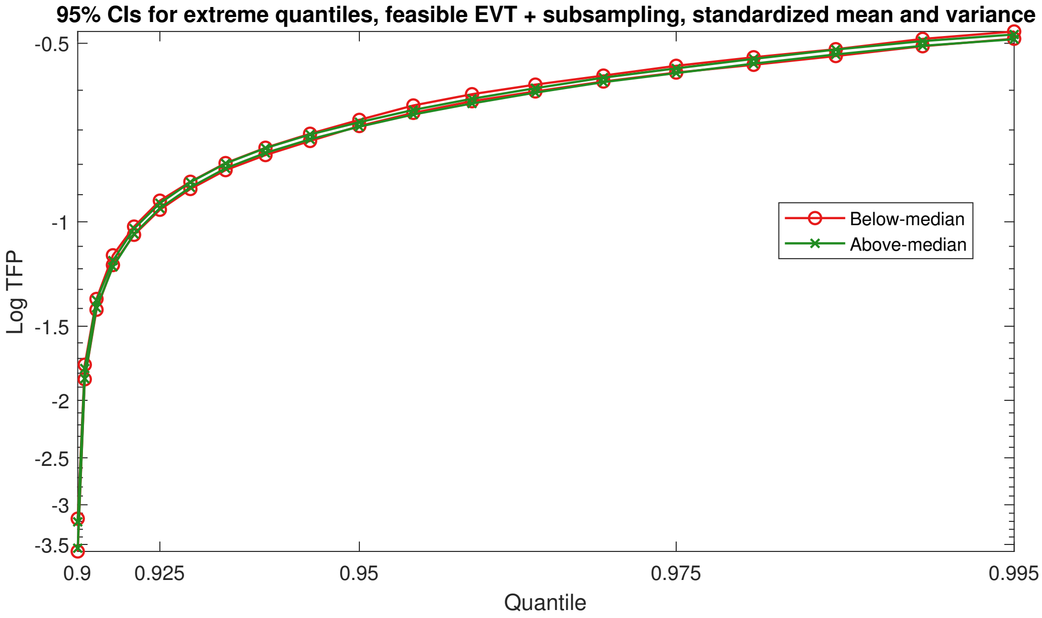
95% CIs for extreme quantiles, feasible EVT + subsampling
Below-median
Above-median
Log TFP
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
0.001
0.026
0.051
0.076
0.096
Quantile

Application: Adjusting Mean and Variance

We examine the right and left tails of the productivity distribution after adjusting for mean and variance (positive estimates for γ incompatible with truncation)

Overall we confirm nonparametrically the assumption of Combes et al. (2012a). There is no significant difference between the two tails after adjusting means and variances


95% CIs for extreme quantiles, feasible EVT + subsampling, standardized mean and variance
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
Log TFP
0.001
0.026
0.051
0.076
0.096
Quantile
Below-median
Above-median



95% CIs for extreme quantiles, feasible EVT + subsampling, standardized mean and variance
Log TFP
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
0.9
0.925
0.95
0.975
0.995
Quantile
Below-median
Above-median

Conclusion

- We address the issue of inference on extreme quantiles of individual heterogeneity based only on noisy estimates
- We obtain sharp conditions under which extreme value theory applies to noisy observables. Under these conditions, noisy observables can be used to form CIs for the extreme quantiles and test hypotheses
- We provide an inference-friendly EVT and show how to obtain the corresponding critical value by subsampling. Further, we show consistency of EV index estimators in our noisy setting
- We compute example rate restrictions and provide sufficient conditions that can be used when there are bounds on the EV parameter
- Extreme approximations perform well under rate conditions derived, but failure of rate conditions may lead to distortions
- Application: extreme quantiles of firm productivity in denser and less dense areas

Theorem (Extreme Value Theorem For Noisy Maximum return)

Let the tail equivalence hold for corresponding scaling factors, the

1. $\gamma > 0$ for Fr a Frechet RV

$$\frac{1}{F^{-1}\left(1-\frac{1}{N}\right)}\left[\vartheta_{N,N,T} - F^{-1}\left(1-\frac{l}{N}\right)\right] \Rightarrow Fr - \left(\frac{1}{l}\right)^{\gamma}, \quad l > 0. \tag{17}$$

2. $\gamma = 0$ for Gu a Gumbel RV

$$\frac{1}{\hat{f}\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}\left(\vartheta_{N,N,T} - F^{-1}\left(1-\frac{l}{N}\right)\right) \Rightarrow Gu - \log(l), \quad l > 0. \tag{18}$$

3. $\gamma < 0$ for W a reverse Weibull RV

$$\frac{1}{F^{-1}(1) - F^{-1}\left(1-\frac{1}{N}\right)}\left(\vartheta_{N,N,T} - F^{-1}\left(1-\frac{l}{N}\right)\right) \Rightarrow W + \left(\frac{1}{l}\right)^{\gamma}, \quad l \geq 0. \tag{19}$$

Intermediate Order Theorem

$\vartheta_{N-k(N),N}$ is intermediate if $k(N) \to \infty$, $k(N) = o(N)$.

Theorem

Let F satisfy a von Mises condition. Let U_i be iid Uniform[0, 1] and suppose additionally that

$$\sup_{u \in [0,1-U_{k,N}]} \frac{\sqrt{k}}{c_N} \left(F^{-1}(u) + \frac{1}{\sqrt{T}} G^{-1}\left(1 - U_{k,N} - u\right) - F^{-1}\left(1 - U_{k,N}\right)\right) \xrightarrow{p} 0 \quad (20)$$

and similar infimum condition. Then

$$\sqrt{k}\frac{\vartheta_{N-k,N} - F^{-1}(1 - k/N)}{c_N} \Rightarrow N(0,1) \quad (21)$$

Subsampling Details

Quantiles of the limit distribution of the feasible EVT can be estimated by subsampling (Politis and Romano, 1994; Politis et al., 1999). Feasible EVT

Split the set of units $\{1, \dots, N\}$ into all possible subsamples of size b and index the subsamples by s, $s = 1, \dots, \binom{N}{b}$. Let $\vartheta^{(s)}_{b-k,b,T}$ be the $(b-k)$th order statistic in subsample s. Define the subsampling estimator $L_{b,N,T}$ for J as

$$L_{b,N,T}(x) = \frac{1}{\binom{N}{b}} \sum_{s=1}^{\binom{N}{b}} \mathbb{I}\{W_{s,b,N,T} \leq x\}, \quad W_{s,b,N,T} = \frac{\vartheta^{(s)}_{b-r,b} - \vartheta_{N-Nl/b,N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}}, \quad \frac{Nl}{b} \leq N.$$

$\hat{c}_\alpha$ is the αth quantile of $L_{b,N,T}(x)$. We show that $\hat{c}_\alpha$ is consistent for true critical value of interest under tail equivalence.

Estimating the EV Index

If tail equivalence holds, the EV index γ can be consistently estimated. Let $k = k(N) \to \infty$ and $k = o(N)$. We prove consistency of the following estimators

1. Hill (1975) estimator: for $\gamma > 0$

$$\hat{\gamma}_H = \frac{1}{k}\sum_{i=0}^{k-1} \log(\vartheta_{N-i,N,T}) - \log(\vartheta_{N-k,N,T}). \tag{22}$$

2. Generalized Pickands estimator of Segers (2005)

$$\hat{\gamma}_k(c,\lambda) = \sum_{j=1}^{k}\left(\lambda\left(\frac{j}{k}\right) - \lambda\left(\frac{j-1}{k}\right)\right)\log\left(\vartheta_{N-\lfloor cj \rfloor,N,T} - \vartheta_{N-j,N,T}\right) \tag{23}$$

where λ is a right-continuous function on $[0,1]$ such that $\lambda(0) = \lambda(1)$ and $\int_0^1 \lambda(t)t^{-1}dt = 1$, $c \in (0,1)$.

References I

Manuel Arellano and Stéphane Bonhomme. Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79(3):987–1020, 2012. ISSN 00346527. doi: 10.1093/restud/rdr045.

Marcus Asplund and Volker Nocke. Firm Turnover in Imperfectly Competitive Markets. *Review of Economic Studies*, 73(2):295–327, 2006. ISSN 00346527. doi: 10.1111/j.1467-937X.2006.00377.x.

M. Christopher Auld and Nirmal Sidhu. Schooling, cognitive ability and health. *Health Economics*, 14(10):1019–1034, 2005. ISSN 10579230. doi: 10.1002/hec.1050.

Laurent Barras, Patrick Gagliardini, and Olivier Scaillet. Skill, Scale and Value Creation in the Mutual Fund Industry. 2021. doi: 10.2139/ssrn.3269995.

References II

Victor Chernozhukov. Extremal Quantile Regression. *Annals of Statistics*, 33(2):806–839, 2005. ISSN 00905364. doi: 10.1214/009053604000001165.

Victor Chernozhukov and Iván Fernández-Val. Inference for Extremal Conditional Quantile Models, With An Application to Market and Birthweight Risks. *Review of Economic Studies*, 78(2):559–589, 2011. ISSN 00346527. doi: 10.1093/restud/rdq020.

Pierre Philippe Combes, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594, 2012a. ISSN 0012-9682. doi: 10.3982/ecta8442.

References III

Pierre Philippe Combes, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. Sorting And Local Wage and Skill Distributions in France. *Regional Science and Urban Economics*, 42(6):913–930, 2012b. ISSN 01660462. doi: 10.1016/j.regsciurbeco.2012.11.003. URL `http://dx.doi.org/10.1016/j.regsciurbeco.2012.11.003`.

Laurens de Haan and Ana Ferreira. *Extreme Value Theory*. Springer, 2006. ISBN 978-0-387-23946-0. doi: 10.1007/0-387-34471-3.

Jorge de la Roca and Diego Puga. Learning By Working in Big Cities. *Review of Economic Studies*, 84(1):106–142, 2017. ISSN 1467937X. doi: 10.1093/restud/rdw031.

Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny. Spatial Sorting. *Journal of Political Economy*, 122(3), 2014. doi: 10.1086/676141.

References IV

Kirill Evdokimov. Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. 2010.

Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010. ISSN 02664666. doi: 10.1017/S0266466609990168.

Xavier Gabaix. Power Laws in Economics and Finance. *Annual Review of Economics*, 1: 255–293, 2009.

Xavier Gabaix. Power Laws in Economics: An Introduction. *Journal of Economic Perspectives*, 30(1):185–205, 2016.

References V

James J. Heckman, Jeffrey Smith, and Nancy Clements. Making the Most out of Programme Evaluations and Social Experiments : Accounting for Heterogeneity in Programme Impacts. *Review of Economic Studies*, 64(4):487–535, 1997. ISSN 00346527. doi: 10.2307/2971729.

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

Koen Jochmans and Martin Weidner. Inference on a Distribution From Noisy Draws. *Econometric Theory*, pages 1–38, 2022. doi: 10.1017/S0266466622000378.

Charles Manski. The Selection Problem. Technical report, 1990. URL `https://econpapers.repec.org/RePEc:att:wimass:90-12`.

References VI

Dimitris N. Politis and Joseph P. Romano. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994. doi: 10.1214/aos/1176325770.

Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer Series in Statistics, 1999. ISBN 978-0-387-98854-2. doi: 10.1007/978-1-4612-1554-7.

Johan Segers. Generalized Pickands Estimators for the Extreme Value Index. *Journal of Statistical Planning and Inference*, 128(2):381–396, 2005. ISSN 03783758. doi: 10.1016/j.jspi.2003.11.004.

Yichong Zhang. Extremal Quantile Treatment Effects. *The Annals of Statistics*, 46(6): 3707–3740, 2018. doi: 10.1214/17-AOS1673.