Inference on Extreme Quantiles of Unobserved Individual Heterogeneity

Vladislav Morozov*

October 16, 2022

Abstract

We develop a methodology for conducting inference on extreme quantiles of unobserved individual heterogeneity (heterogeneous coefficients, heterogeneous treatment effects, and other unobserved heterogeneity) in a panel data or meta-analysis setting. Examples of interest include productivity of most and least productive firms or prediction intervals for study-specific treatment effects in meta-analysis. Inference in such a setting is challenging. Only noisy estimates of unobserved heterogeneity are available, and approximations based on the central limit theorem work poorly for extreme quantiles. For this situation, under weak assumptions we derive an extreme value theorem for noisy estimates and appropriate rate and moment conditions. In addition, we develop a theory for intermediate order statistics. Both extreme and intermediate order theorems are then used to construct confidence intervals for extremal quantiles. The limiting distribution is non-pivotal, and we show consistency of both subsampling and simulating from the limit distribution. Furthermore, we provide a novel self-normalized intermediate order theorem. In a Monte Carlo exercise, we show that the resulting extremal confidence intervals have favorable coverage properties in the tail.

KEYWORDS
unobserved heterogeneity, panel data, heterogeneous coefficients, estimation noise, extreme value theory, feasible inference

*vladislav.morozov@upf.edu; Department of Economics and Business, Universitat Pompeu Fabra and Barcelona School of Economics

†Acknowledgments: I am indebted to Christian Brownlees and Kirill Evdokimov for their support and guidance. I am also grateful to Iván Fernández-Val, Adam Lee, Zoel Martin, Geert Mesters, Katerina Petrova, Barbara Rossi, Alina Shirshikova, Piotr Zwiernik, and the participants at the 2021 European Winter Meeting of the Econometric Society, the 2021 Spanish Economic Association Symposium, the 2022 Royal Economic Society Annual Conference, and the 27th International Panel Data Conference for comments and discussions.

1 Introduction

Extreme quantiles of unobserved individual heterogeneity (UIH) are often of interest in analysis of economic panel data and in random effects meta-analysis; UIH includes heterogeneous coefficients, treatment effects, etc. For example, in the setting of Combes et al. (2012) we might be interested in estimating the lowest level of firm productivity that is compatible with firm survival in a given area. This corresponds to finding the zeroth quantile of the distribution of productivity of surviving firms. In another example, in random effects meta-analysis, the true treatment effects of the same intervention are allowed to differ between studies. A 95% prediction interval is an interval such that the true treatment effect in a new study falls with probability 95% in it. Such intervals provide an important summary of between-study heterogeneity (Higgins et al., 2009; Nagashima et al., 2019). To construct such an interval, we need to estimate and quantify uncertainty about the 2.5th and 97.5th percentiles of the distribution of study-specific treatment effects; these percentiles may be relatively extreme for datasets of results of individual studies when number of studies is not large.

The key challenge to estimation and inference is that only noisy estimates of UIH are available, and it is not a priori clear when such estimates yield useful information about the quantiles of interest. These individual estimates are constructed from individual time series, studies, or cluster-level data, and thus are subject to estimation noise. This estimation noise and the true UIH are often related in a complex and unobservable manner due to the dependence between UIH and covariates used in estimation (e.g. Heckman (2001); Browning and Carro (2007, 2010); Okui and Yanagi (2019b); Breitung and Salish (2021)). Importantly, estimation noise does not "average out" when dealing with extreme quantiles, in contrast to estimation of means.

In this paper we develop a methodology for conducting inference on extreme quantiles of UIH on the basis of noisy estimates, and we establish sharp conditions under which noisy estimates of UIH are informative about the quantiles of interest. The noisy estimates can be used for estimation and inference if the tails of their distribution converge to the tails of the distribution of interest in a certain weak pointwise sense. We construct confidence intervals for extreme quantiles of UIH using self-normalizing ratios of extreme or intermediate order statistics of the estimates. We also provide hypothesis tests about the support of UIH. To obtain the relevant asymptotic distributions, we prove appropriate extreme and intermediate value theorems (EVT and IVT) for noisy estimates.[1]

The asymptotic distributions of the statistics we propose are based on extreme and intermediate asymptotic approximations. Such approximations are suited to inference on extreme quantiles, and complement the central approximations constructed by Jochmans and Weidner (2022) for conducting inference on central quantiles of UIH.[2] The three asymptotic approximations differ in how the quantile of interest is modeled as a function of the cross-sectional size. Extreme order approximations model the quantile of interest as drifting to 0 or 1 at rate proportional to the cross-sectional dimension.[3] The resulting confidence intervals and tests for extreme quantiles of UIH are based on a ratio of extreme order statistics of the estimated UIH. The distribution of such ratios depends on the extreme value (EV) index of the distribution of latent true UIH, and thus is non-pivotal. The corresponding critical values can be obtained by subsampling or by simulation. For simulating the critical values, we propose using the Pickands (1975) estimator for the EV index and prove its consistency. We establish consistency of the resulting critical values for both methods. In intermediate order approximations, the quantile of interest drifts to 0 or 1 at a rate slower than the cross-sectional sample size. In this case we base confidence intervals on a ratio statistic of intermediate order statistics. This ratio exploits a certain knife-edge bias property between

[1]For an introduction to extreme value theory see, among others, Resnick (1987); Leadbetter et al. (1983); de Haan and Ferreira (2006). See chapter 2 of de Haan and Ferreira (2006) for an introduction to the limit theory of intermediate order statistics (also called intermediate extreme order statistics).

[2]Jochmans and Weidner (2022) also study the problem of estimating the distribution of UIH given only noisy estimates; this is the forward problem, compared to the inverse problem of estimating quantiles. The problem of estimating the distribution of UIH in the presence of estimation noise is also tightly linked to the normal means problem (see Efron (2011); Weinstein et al. (2018) for some recent contributions).

[3]This approach is related to the fixed-k tail inference methods of Müller and Wang (2017), which they develop in a noiseless setting.

two perfectly dependent random variables to normalize the statistic.[4] In contrast, central order approximations are based on the central limit theorem for quantiles; the quantile of interest is modeled as fixed.

We show that the noisy estimates of UIH can be used for estimation and valid inference on the quantiles of interest under mild conditions. There are two key restrictions. First, we impose weak assumptions on the marginal distributions of UIH and of the estimation noise. We do not restrict the joint distribution of the true UIH and the estimation noise. This allows estimation noise and the true UIH to be related in a complex manner, as they typically are outside of experimental settings (Heckman, 2001; Browning and Carro, 2007, 2010; Okui and Yanagi, 2019b; Breitung and Salish, 2021). Second, the tail of the noisy distribution of estimates has to converge to the tail of the latent distribution of interest in a certain pointwise sense; a condition that we label pointwise asymptotic tail equivalence.

When tail equivalence holds, noisy estimates inherit extreme value properties of the UIH. This allows us to establish an EVT and supporting distributional results for noisy estimates. These results in turn serve as a foundation for inference.

For a broad class of distributions, we obtain sufficient conditions for the tail equivalence conditions in terms of rate restrictions on the cross-sectional dimensional and individual sample sizes. The sufficient conditions only require typical moment or normality assumptions on the estimation noise and a lower bound for the EV index. The rate restrictions are mild if the distribution has an infinite right endpoint, but may be more stringent for distributions with finite endpoints. There are two main components which make restrictions milder: how light the tails of the noise distribution are, and how heavy the tails of the true distribution of the effects are. Intuitively, lighter-tailed noise and heavier-tailed UIH both increase signal strength.[5]

In simulations we find that if rate conditions hold, extreme-order approximations offer favorable coverage properties in the tails, including when compared with other approximations.

[4]This result may also be of interest in the noiseless case.

[5]In particular, this leads to a curious situation where it is desirable to know that the UIH have only a limited number of moments.

The rate conditions are important, and their failure may lead to distorted inference.

Several recent papers have proposed methods for studying distributional properties of UIH. Examples include papers by Arellano and Bonhomme (2012); Jochmans and Weidner (2022); Okui and Yanagi (2019a,b). In a contribution related to ours, Jochmans and Weidner (2022) study the issue of estimating the distribution function and central quantiles of UIH from noisy estimates. They develop rate conditions and asymptotic normality results in a similar setting to ours. However, their methodology is based on a central limit theorem for quantiles, and cannot be used for inference on extreme quantiles. Other approaches, such as kernel density estimators for the distribution (Okui and Yanagi, 2019a; Barras et al., 2021) are not well-suited for tackling our problem.[6]

Our setting and results should be contrasted with the literature on estimating extreme conditional quantiles or treatment effects in cross-sectional settings (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011; Zhang, 2018). The focus of those papers is on modeling extreme conditional quantiles of observable quantities, whereas our focus lies squarely on unobserved quantities and the issue of estimation noise. A similar point applies to the literature on bounding the distribution of treatment effects (Heckman et al., 1997; Fan and Park, 2010; Firpo and Ridder, 2019).

The rest of the paper is organized as follows. Section 2 formalizes our setup and assumptions, and provides examples. In section 3 we lay down the probabilistic foundations of our inference theory by proving extreme and intermediate extreme value theorems for noisy estimates. Building on these results, in section 4 we discuss three approaches to inferences: extreme, intermediate, and central order asymptotic approximations. Using those approximations we construct confidence intervals and tests about the support of the distribution of unobserved UIH. In section 5 we explore performance of those CIs in a Monte Carlo setting. All proofs are collected in the Appendix.

[6]For example, if we use a kernel with bounded support, the resulting kernel density estimator will have a finite right endpoint, even if the true distribution has an infinite one, and vice versa.

2 Setting and Assumptions

2.1 Problem statement

Let i index cross-sectional units or individual studies; in what follows we will refer to them as "units". Let θ_i be the scalar individual heterogeneity of interest: θ_i may be a treatment effect, effect size, a coordinate of a vector of individual-specific coefficients, value of a function at a point, etc.. Suppose that θ_i are sampled in an iid fashion from some distribution F.

Our goal is to estimate quantiles $F^{-1}(q) = \inf\{x : F(x) \geq q\}$ where q is close to 0 or 1, and to conduct inference on them. In particular, we are interested in constructing confidence intervals and hypothesis tests for hypotheses like $H_0 : \theta_i \geq 0$, which is equivalent to $H_0 : F^{-1}(0) \geq 0$. Without loss of generality we focus on the right tail of F.

2.2 Data generating process

We do not observe θ_i directly; instead we only see noisy observations $\vartheta_{i,T}$ generated as

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

where T_i is the sample size available for unit i, ε_{i,T_i} is the scaled estimation error, $\varepsilon_{i,T} = O_p(1)$ for all T, and $p > 0$ is the rate of convergence of the estimator. For clarity of exposition, in what follows we assume that for all units $T_i = T$ for some common T; in Appendix B we show that our results naturally extend to the case where sample sizes are unequal between units. p is determined by the estimation method used in a given case. For estimators convergent at the parametric rate $T^{-1/2}$, p is equal to $1/2$, but we allow other rates. $\mathbb{E}(\varepsilon_{i,T})$ may be nonzero and need not converge to 0 as $T \to \infty$; however, this bias may not diverge as $T \to \infty$. See example 2 below for $p \neq 1/2$ and presence of asymptotic bias.

Representation (1) is general. It is compatible with any estimation method for which a suitable rate of convergence is known; $\varepsilon_{i,T}$ can always be defined as $\varepsilon_{i,T} = T^p(\vartheta_{i,T} - \theta_i)$.

Before stating our assumptions on θ_i and $\varepsilon_{i,T}$, we provide several examples of how $\vartheta_{i,T}$ may be constructed. Intuitively, $\vartheta_{i,T}$ are estimators of θ_i that have growing precision. $\vartheta_{i,T}$

can potentially be biased. Our setting nests the setup considered by Jochmans and Weidner (2022) and the typical setup of random effects meta-analysis (see e.g. Higgins et al. (2009)).

Example 1 (Unit-wise OLS)**.** Let the outcome y be determined by the covariate x as $y_{it} = \theta_i x_{it} + u_{it}$, $\mathbb{E}(x_{it}u_{it}) = 0$, $i = 1, \dots, N, t = 1, \dots, T$. We estimate heterogeneous coefficients θ_i by unit-by-unit application of OLS: $\vartheta_{i,T} = \left(T^{-1}\sum_{t=1}^T x_{it}^2\right)^{-1} T^{-1}\sum_{t=1}^T x_{it}y_{it}$. Define $\varepsilon_{i,T} = \left(T^{-1}\sum_{t=1}^T x_{it}^2\right)^{-1} T^{-1/2}\sum_{t=1}^T x_{it}u_{it}$ to write $\vartheta_{i,T}$ as in eq. (1) for $p = 1/2$. More generally, let x_{it} be a vector of covariates in the model: $y_{it} = \beta_i' x_{it} + u_{it}$. In this case the parameter of interest θ_i may be the first coordinate of β_i or a more general smooth transformation of β_i, such as a long-run effect in a dynamic model (Pesaran and Smith, 1995).

Example 2 (Nonparametric regression)**.** Let the conditional expectation of y_{it} given the covariate x_{it} be $g_i(x) = \mathbb{E}(y_{it}|x_{it} = x)$. Suppose the effect of interest is the value $g_i(x_0)$ for some fixed value x_0. The Nadaraya-Watson estimator for the regression function g_i is given by $\hat{g}_{i,T}(x) = \left(\sum_{t=1}^T K\left((x_{it} - x)/h\right)\right)^{-1}\sum_{t=1}^T y_{it}K\left((x_{it} - x)/h\right)$ where K is a kernel function and h is a bandwidth parameter. In this case $\vartheta_{i,T} = \hat{g}_i(x_0)$. Define $u_{it}(x) = y_{it} - g_i(x)$, and set $\varepsilon_{i,T} = \left(\sum_{t=1}^T K\left(x_{it} - x_0/h\right)\right)^{-1}\sqrt{Th}\sum_{t=1}^T u_{it}(x_0)K\left((x_{it} - x_0)/h\right)$. Let $h = T^{-s}$, $s \in (0, 1)$. It holds that $\varepsilon_{i,T} = O_p(1)$ and $\vartheta_{i,T} = \theta_i + T^{-(1-s)/2}\varepsilon_{i,T}$ under suitable conditions on h. If h is picked to minimize the MSE, $\varepsilon_{i,T}$ is asymptotically normal, but has a non-zero mean in the limit.

Example 3 (Unit-wise IV)**.** Consider the setup of example 1: let $y_{it} = \theta_i x_{it} + u_{it}$. However, let x_{it} be endogenous in the sense that $\mathbb{E}(u_{it}x_{it}) \neq 0$. Suppose a valid instrument z_{it} is available: $\mathbb{E}(z_{it}u_{it}) = 0$, $\mathbb{E}(z_{it}x_{it}) \neq 0$. In this case θ_i can be estimated consistently with the instrumental variable estimator $\vartheta_{i,T} = \left(T^{-1}\sum_{t=1}^T z_{it}x_{it}\right)^{-1} T^{-1}\sum_{t=1}^T z_{it}u_{it}$. Defining $\varepsilon_{i,T} = \left(T^{-1}\sum_{t=1}^T z_{it}x_{it}\right)^{-1} T^{-1/2}\sum_{t=1}^T z_{it}u_{it}$, we can write $\vartheta_{i,T}$ in form of eq. (1) for $p = 1/2$.

Example 4 (Nonlinear estimator)**.** Let $y_{it} \in \{0, 1\}$ be a binary variable with $P(y_{it} = 1|x_{it}) = \Lambda(\theta_i x_{it})$ where x_{it} is some covariate and Λ is the logistic function. The parameter of interest is θ_i. Let $\vartheta_{i,T}$ be the maximum likelihood estimator based on the data for unit i and define $\varepsilon_{i,T} = \sqrt{T}(\vartheta_{i,T} - \theta_i)$.

Notation Let $\vartheta_{1,N,T} \leq \cdots \leq \vartheta_{N,N,T}$ be the N order statistics of $\{\vartheta_{1,T}, \ldots, \vartheta_{N,T}\}$, and similarly let $\theta_{1,N} \leq \cdots \leq \theta_{N,N}$ be the order statistics of the latent noiseless $\{\theta_1, \ldots, \theta_N\}$. 'Noiseless maximum' always stands for $\theta_{N,N}$; 'noisy maximum' stands for $\vartheta_{N,N,T}$; same applies to all other order statistics. $\Rightarrow$ denotes weak convergence, both of random variables and functions.

2.3 Assumptions

Assumption 1. For each T, $\{(\theta_i, \varepsilon_{i,T})\}_{i=1,\ldots,N}$ are independent and identically distributed random vectors indexed by i.

Observations $\vartheta_{i,T}$ are sampled in an iid fashion. Note that $\vartheta_{i,T}$ can be conditionally heteroskedastic: $\mathrm{Var}(\vartheta_{i,T}|\theta_i) = \mathrm{Var}(\varepsilon_{i,T}|\theta_i)$ can depend on θ_i, provided this variance exists.

Let F be the marginal distribution of θ and let θ_F be the right endpoint of F: $\theta_F = F^{-1}(1)$.

Assumption 2. F is in the weak domain of attraction of an extreme value distribution with extreme value index γ, that is, one of the three following conditions holds.

(1) *Weibull domain of attraction* ($\gamma < 0$): θ_F is finite and and for some $\gamma < 0$ it holds that $\lim_{t \downarrow 0} (1 - F(\theta_F - tx))/(1 - F(\theta_F - t)) = x^{-1/\gamma}$, for all $x > 0$.

(2) *Gumbel domain of attraction* ($\gamma = 0$): $\theta_F \leq +\infty$ and for some positive function $\hat{f}$ $\lim_{t \uparrow \theta_F} (1 - F(t + x\hat{f}(t)))/(1 - F(t)) = e^{-x}$, for all $x > 0$.

(3) *Fréchet domain of attraction* ($\gamma > 0$): $\theta_F = +\infty$ and for some $\gamma > 0$ it holds that $\lim_{t \to \infty} (1 - F(tx))/(1 - F(t)) = x^{-1/\gamma}$ for all $x > 0$.

Under assumption 2, the classical extreme value theorem of Gnedenko (1943) applies to the latent noiseless distribution F; see theorem 3.1 below.[7] This will serve as the basis for extending this extreme value convergence to the observed noisy data. Without assumption 2, we would not be able to conduct inference using the asymptotic behavior of the sample

[7] See chapters 1-2 in Resnick (1987) and chapters 1-5 in de Haan and Ferreira (2006) for a thorough treatment of extreme value theory for iid observations.

maximum even if we had access to the true θ_i. Assumption 2 is a fairly mild assumption, satisfied by almost all textbook continuous distributions and many discrete ones.

θ_i and $\varepsilon_{i,T}$ will typically be related in a complex and unobservable manner outside of tightly controlled experimental settings (see remark 1 below). In addition, the distribution of $\varepsilon_{i,T}$ might not be well-approximated by a normal distribution or be centered at zero even in the limit. The following examples illustrate the issue.

Example 2 (Continued)**.** Consider the setting of example 2. The distribution of $\varepsilon_{i,T}$ depends not only on $\theta_i = g_i(x_0)$, but also on the behavior of g_i in a neighborhood of x_0, in particular on the derivatives of g_i. The value of $g_i(x_0)$ does not restrict the derivatives of g_i and vice versa. Thus, a given distribution of $\vartheta_{i,T}$ may be associated with multiple distinct distributions of θ_i, driven by the existence of a large variety of conditional distributions of $\varepsilon_{i,T}$ given θ_i. Moreover, the difficulty persists in the limit. If h is picked to minimize the MSE, the asymptotic mean of $\varepsilon_{i,T}$ is of the same magnitude as the asymptotic variance. Both mean and variance are controlled by the unknown $g_i'(x_0)$ and $g_i''(x_0)$; these derivatives may have arbitrary distribution for a given distribution of θ.

Example 3 (Continued)**.** In the setting of example 3, the distribution of $\varepsilon_{i,T}$ can be dissimilar to the normal distribution even for large T if z_{it} is not a strong instrument. As Nelson and Startz (1990) show, even if u_{it} is normally distributed, the distribution of $\varepsilon_{i,T}$ can be skewed, bimodal, and possess an infinite expected value.

A key feature of our analysis is that we place no restrictions on the dependence structure between θ_i and $\varepsilon_{i,T}$, motivated by the above difficulty. Instead, we only impose a weak assumption on the marginal distribution G_T of $\varepsilon_{i,T}$.

Assumption 3. $\{G_T\}$ form a tight family of distributions indexed by T.

Intuitively, assumption 3 does not impose any distributional restrictions on $\varepsilon_{i,T}$, but requires that $\varepsilon_{i,T}$ be defined in such a way that, as $T \to \infty$, the distributions G_T of $\varepsilon_{i,T}$ do not escape to infinity.[8] Assumption 3 holds automatically if $T^p(\vartheta_{i,T} - \theta_i)$ has a non-degenerate asymptotic distribution. Together with definition (1), assumption 3 implies that each $\vartheta_{i,T}$ is consistent for θ_i, but we allow $\vartheta_{i,T}$ to be biased in finite samples; for instance, all estimators of examples 1-4 are generally biased. In addition, we allow the mean of $\varepsilon_{i,T}$ to be nonzero in the limit; this occurs in example 2 with the MSE-minimizing choice of bandwidth h. Finally, we do not require G_T to be normal, as assumption of normality may be a poor approximation in some cases even for T large (as in example 3).

Remark 1 (On sources of dependence between θ_i and covariates)**.** Consider examples 1-4. Dependence between θ_i and x_{it} may arise if the decision maker chooses x_{it} based on knowledge of θ_i; x_{it} may be a continuous variable, a binary participation decision in a quasi-experimental setting, or generally any policy chosen. Dependence of this type is common (Heckman and Vytlacil, 1998; Heckman, 2001; Browning and Carro, 2007; Breitung and Salish, 2021).

3 Extreme Value Theory For Noisy Estimates

3.1 Extreme Value Theorem For Noisy Estimates

We begin our analysis by establishing distributional results for the sample maximum $\vartheta_{N,N,T} = \max\{\vartheta_{1,T}, \dots, \vartheta_{N,T}\}$. Our approach involves obtaining conditions under which $\vartheta_{N,N,T}$ inherits the limit properties of $\theta_{N,N} = \max\{\theta_1, \dots, \theta_N\}$. The potential asymptotic distributions of $\theta_{N,N}$ are characterized by the following standard result:

Theorem 3.1 (EVT for θ_i, (Gnedenko, 1943))**.** *Let assumption 1 hold. Let F satisfy assumption 2 with EV index γ. Then as $N \to \infty$ it holds that $(\theta_{N,N} - b_N)/a_N \Rightarrow X$, where*

($\gamma > 0$) $a_N = F^{-1}(1 - 1/N), b_N = 0$, and X has the Fréchet distribution: $P(X \leq x) = \exp(-x^{-1/\gamma})$ for $x \geq 0$ and 0 for $x < 0$.

[8] Assumption 3 is satisfied if $\varepsilon_{i,T}$ converges in probability to zero. However, it is desirable to define $\varepsilon_{i,T}$ in such a way that assumption 3 holds, but G_T does not collapse to a degenerate limit, as this leads to the mildest possible conditions in our results.

($\gamma = 0$) $a_N = \hat{f}(F - F^{-1}(1 - 1/N)), b_N = F^{-1}(1 - 1/N)$ *for* $\hat{f}$ *of assumption 2, and* X *has the Gumbel distribution:* $P(X \leq x) = \exp(-e^{-x})$ *for* $x \in \mathbb{R}$*.*

($\gamma < 0$) $a_N = \theta_F - F^{-1}(1 - 1/N), b_N = \theta_F$*, and* X *has the reverse Weibull distribution:* $P(X \leq x) = \exp(-(-x)^{-1/\gamma})$ *for* $x < 0$ *and* $P(X \leq x) = 1$ *for* $x \geq 0$*.*

Theorem 3.1 is the standard Fisher-Tippett-Gnedenko extreme value theorem (Gnedenko, 1943). The form of the asymptotic distribution of $\theta_{N,N}$ depends on the EV index γ.

The following theorem provides conditions under which the observed and the latent maxima $\vartheta_{N,N,T}$ and $\theta_{N,N}$ have the same asymptotic distribution.

Theorem 3.2 (EVT for noisy estimates)**.** *Let assumption 1 hold. Let constants* a_N, b_N *and a random variable* X *be as in theorem 3.1 so that* $(\theta_{N,N} - b_N)/a_N \Rightarrow X$ *as* $N \to \infty$*. Let for each* $\tau \in (0, \infty)$ *the following tail equivalence (TE) conditions hold: as* $N, T \to \infty$

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

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

Then as $N, T \to \infty$

$$\frac{\vartheta_{N,N,T} - b_N}{a_N} \Rightarrow X.$$

Conditions (TE-Inf) *and* (TE-Sup) *are sharp in the following sense: if at least one of the conditions fails, there exists a sequence of joint distributions of* $(\theta_i, \varepsilon_{i,T})$ *with given marginal distributions* F, G_T *such that* $(\vartheta_{N,N,T} - b_N)/a_N$ *weakly converges to a limit different from* X *or does not converge at all.*

(TE-Sup) and (TE-Inf) are pointwise asymptotic tail equivalence conditions between the distributions of $\vartheta_{i,T}$ and θ_i. This interpretation is straightforward. Observe that the $F^{-1} + T^{-p}G^{-1}$ term approximately corresponds to the quantiles of the noisy estimates, with the infimum and supremum adjusting for the unknown joint distribution of θ_i and $\varepsilon_{i,T}$. The $-F^{-1}$ term corresponds to the noiseless quantiles. If (TE-Inf) and (TE-Sup) hold, then the distance between the tails of F and the distribution of $\vartheta_{i,T}$ is shrinking, pointwise in τ.

Remark 2. It is natural to use the same normalization constants a_N, b_N for both $\vartheta_{N,N,T}$ and $\theta_{N,N}$. The distribution of $\vartheta_{i,T}$ is composed of two components: the G_T component that decays as $T \to \infty$, and the F component. As the F component does not decay, it is appropriate to use the same a_N. For inference, the centering constants b_N will typically be quantiles of F. Since they are the primary object of interest, it is natural to center at them.

Remark 3 (Impact of failure of conditions (TE-Inf) and (TE-Sup))**.** It is generally uncertain to what extent $\vartheta_{N,N,T}$ inherits the limit properties of $\theta_{N,N}$ if conditions (TE-Inf) and (TE-Sup) fail to hold. For some joint distributions of $(\theta_i, \varepsilon_{i,T})$, $\vartheta_{N,N,T}$ might still have the same asymptotic distribution as $\theta_{N,N}$. However, as the proof of theorem 3.2 shows, there exists at least one possible asymptotic distributions for $\vartheta_{N,N,T}$ that is different from the distribution of $\theta_{N,N}$. Moreover, $\vartheta_{N,N,T}$ may only converge to it along a subsequence and not overall.

A special knife-edge case arises if for each $\tau \in (0, \infty)$ the expressions in (TE-Sup) and (TE-Inf) converge to the same δ_τ. In this case $a_N^{-1}\left(\vartheta_{N,N,T} - b_N\right)$ does converge. However, if δ_τ is not a constant function of $\tau \in (0, \infty)$, this limit distribution has a shape different from the limit of $a_N^{-1}(\theta_{N,N} - b_N)$. Shape is only preserved if $\delta_\tau = \delta \neq 0$ for all τ. In this case $a_N^{-1}\left(\vartheta_{N,N,T} - b_N\right)$ converges to $X + \delta$ if $a_N^{-1}(\theta_{N,N} - b_N)$ converges to X. We note that this special case appears unlikely, as it would require that the contributions of the left and the right tails of G_T are balanced out in an exactly symmetrical manner, yet not negligible.

When do conditions (TE-Inf) and (TE-Sup) hold? As we presently show, sufficient conditions for (TE-Inf) and (TE-Sup) take form of rate conditions on N and T, given mild assumptions on the tails of G_T. The rate conditions restrict the magnitude of N relative to T as both $N, T \to \infty$.

To build intuition and motivate the form of our conditions, we first consider three examples. The distributions considered for F and G_T are quite general. First, if F satisfies assumption 2 with $\gamma \neq 0$, it differs from one of the cdfs of examples 5 and 7 below only by a slowly varying function. Example 6 provides a standard example of a distribution F that satisfies assumption 2 with $\gamma = 0$. Second, we consider two different cases for G_T: G_T only assumed to possess a given number of finite moments, and G_T normal. The first case reflects the typical practice in econometrics of only assuming that a limited number of moments exists; the second case reflects the situation in which the distribution $T^p(\vartheta_{i,T} - \theta_i)$ is well approximated by a normal distribution. In practice, we expect G_T to lie between the two extremes. All details of the examples are verified in the Appendix A.

Example 5 ($\gamma > 0$)**.** Let the cdf of θ_i be given by $F_{Fr,\kappa}(\theta) = 1 - (\theta+1)^{-\kappa}, \kappa > 0, \theta \in [0, \infty)$. $F_{Fr,\kappa}$ satisfies assumption 2 with $\gamma = 1/\kappa > 0$. By theorem 3.2, $a_N = N^{1/\kappa} - 1, b_N = 0$, and $\theta_{N,N}/(N^{1/\kappa} - 1)$ is asymptotically distributed as a Fréchet random variable.

We consider two specifications for G_T. First, let $G_{\beta,T}$ be defined by its density $g_\beta(x) = (\beta/2)(1 + \mathrm{sgn}(x - \mu_T)(x - \mu_T))^{-\beta-1}, \beta > 0, x \in \mathbb{R}$ where μ_T is a bounded sequence. $G_{\beta,T}$ is a symmetric distribution with median μ_T and moments of order $< \beta$. Second, let $G_{Normal,T}$ be $N(\mu_T, \sigma_T^2)$ where (μ_T, σ_T^2) is a bounded sequence.

First, let $G_T = G_{\beta,T}$. Then (TE-Inf) and (TE-Sup) hold if $N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, it is sufficient that $\sqrt{\log(N)}/N^{1/\kappa}T^p \to 0$ for (TE-Inf) and (TE-Sup) to hold.

Observe that there is no restriction on magnitudes of N and T if $\varepsilon_{i,T}$ has more than κ moments. This result is intuitive, as in this case F has a heavier tail which is more pronounced in the data and dominates the tail of G_T.

Example 6 ($\gamma = 0$)**.** Let $F = F_{Gu,\lambda}$ be the exponential distribution with parameter λ. $F_{Gu,\lambda}$ satisfies assumption 2 with $\gamma = 0$, $a_N = 1, b_N = \log N/\lambda$; by theorem 3.2 $(\theta_{N,N} - \log N/\lambda)$ is asymptotically distributed as a Gumbel random variable. Proceeding as in the preceding example, we obtain that if $G_T = G_{\beta,T}$, conditions (TE-Inf) and (TE-Sup) hold if $N^{1/\beta}(\log(T))^{1/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, the condition relaxes to $\sqrt{\log(N)}/T^p \to 0$. Unlike example 5, there are rate restrictions on N and T even if G_T has exponentially light tails. How stringent the rate conditions are depends on how many moments $\varepsilon_{i,T}$ is assumed to have. For example, if we are only willing to assume that $\varepsilon_{i,T}$ has 8 moments and if $p = 1/2$, it is sufficient that $N\log(T)/T^4 \to 0$.

Example 7 ($\gamma < 0$)**.** Let $\theta_F < \infty$ and let the cdf of θ_i be given by $F_{W,\alpha}(\theta) = 1 - ((\theta_F - \theta)/\theta_F)^\alpha, \alpha > 0, \theta \in [0, \theta_F]$, $\theta_F = F^{-1}(1) < \infty$. $F_{W,\alpha}$ satisfies assumption 2 with

$\gamma = -1/\alpha < 0$, $a_N = \theta_F/N^{1/\alpha}, b_N = \theta_F$; by theorem 3.2 $N^{1/\alpha}(\theta_{N,N} - \theta_F)/\theta_F$ is asymptotically distributed as a Weibull random variable. If $G_T = G_{\beta,T}$, then (TE-Inf) and (TE-Sup) hold if $N^{1/\beta+1/\alpha}(\log(T))^{1/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, it is sufficient that $N^{1/\alpha}\sqrt{\log(N)}/T^p \to 0$. Since $\alpha > 0$, the rate conditions of this example impose a stronger restriction on magnitude of N than conditions of examples 5 and 6. For example, let $\alpha = 1$, then $F_{W,1}$ is the uniform distribution on $[0, \theta_F]$. If estimation noise is normal and $p = 1/2$, then the conditions are satisfied if $N^2\sqrt{\log(N)}/T \to 0$

The rate restrictions derived in examples 5-7 generally cannot be verified directly in practice, as they depend on the unknown distribution F through its EV index γ. To see this, observe that if $G_T = G_{\beta,T}$, all conditions can be written in a common form $N^{1/\beta-\gamma}(\log(T))^{1/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, the conditions can be equivalently written as $N^{-\gamma}\sqrt{\log(N)}/T^p \to 0$.

However, a sufficient condition is possible if a lower bound γ' for γ is available. This is intuitive in light of the above examples. Distributions with larger values of γ have heavier and longer tails. This heavier tail of F will be more pronounced in the data relative to the contribution of the noise, and so a larger value of N will be allowable. The following proposition formalizes this idea.

Proposition 3.3. *Let assumptions 1 and 3 hold. 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* $\varepsilon_{i,T} \sim N(\mu_T, \sigma_T^2)$*, and let* $N^{-\gamma'}\sqrt{\log(N)}/T^p \to 0$ *for some* γ'*.*

In addition, let F *satisfy assumption 2 with EV index* $\gamma > \gamma'$*. Then* (TE-Inf) *and* (TE-Sup) *hold for* F *and* G_T*.*

As proposition 3.3 shows, two key factors determine how stringent the rate restrictions are. First, as highlighted above, a heavier right tail of F (larger γ') allows a larger value of N relative to T. Second, lighter tails of G_T also lead to milder restrictions on N. The conditions on tail of G_T are captured by (1) and (2), which follow the pattern of examples 5-7. Condition (1) only requires G_T to have uniformly bounded βth moments; (2) is a considerably stronger assumption of exact normality, it may be seen as a limiting version of (1) as $\beta \to \infty$.

Remark 4. An appropriate value of γ' might often be apparent in a given application. For example, Gabaix (2009, 2016) documents that many economic relations follow a power law and outlines some general theoretical mechanisms under which a power law arises. If such a mechanism is likely to hold in a given situation, it is reasonable to assume that the distribution of the data is well approximated by a power law. Since a power law distribution has to have $\gamma > 0$, it is sufficient to check the hypothesis of proposition 3.3 with $\gamma' = 0$.

Remark 5. It is intuitive that the sufficient conditions for (TE-Inf) and (TE-Sup) restrict the magnitude on N relative to T. If N is large relative to T, there is a higher chance that the observed maximum is due to a large realization of $\varepsilon_{i,T}$. In contrast, when T is large, we can have high confidence that a given large realization can be attributed to θ_i. Conditions (TE-Inf) and (TE-Sup) ensure that asymptotically large realizations are due only to θ_i.

3.2 Intermediate Order Statistics

To conduct inference on extreme quantiles, we also need to develop an asymptotic theory for intermediate order statistics. In this section we establish an intermediate value theorem for $\vartheta_{i,T}$ under a von Mises assumption.

Formally, $\theta_{N-k(N),N}$ is called an intermediate (or intermediate extreme) order statistic if $k(N) \to \infty$ as $N \to \infty$ and $k(N) = o(N)$.[9] Intuitively, intermediate order statistics asymptotically stay in the tail, but are not the top statistics. We will generally suppress dependence of k on N.

To derive asymptotic properties of intermediate order statistics, we impose an additional assumption on F that refines assumption 2.

Assumption 4 (F satisfies a first order von Mises condition)**.** *F is twice differentiable with density f, f positive in some left neighborhood of θ_F (θ_F finite or infinite), and for some*

[9] See chapter 2 of de Haan and Ferreira (2006) for a detailed discussion of asymptotic theory for intermediate order statistics in an iid setting.

$\gamma \in \mathbb{R}$

$$\lim_{t \uparrow \theta_F} \left(\frac{1-F}{f} \right)' (t) = \gamma.$$

Assumption 4 is a slight strengthening of assumption 2 and implies it. See Dekkers and de Haan (1989) and chapters 1 and 2 of de Haan and Ferreira (2006) for a discussion of the condition and its plausibility. The distributions of examples 5-7 satisfy assumption 4.

We now state a theorem describing the asymptotic behavior of intermediate order statistics:

Theorem 3.4 (Intermediate value theorem (IVT))**.** *Let assumptions 1 and 4 hold and let* $k = k(N) \to \infty, k = o(N)$ *as* $N \to \infty$*. Define* c_N *as the derivative of the inverse of* $1/(1-F)$ *evaluated at* N/k *and multiplied by* N/k*, that is,* $c_N = (N/k) \times \left[\left((1/(1-F))^{-1} \right)' (N/k) \right]$*. Let* $U_1, \dots, U_N$ *be iid Uniform[0, 1] random variables. Let the following tail equivalence conditions hold as* $N, T \to \infty$

$$\sup_{u \in \left[0, 1-U_{k,N}\right]} \frac{\sqrt{k}}{c_N} \left(F^{-1}(u) + \frac{1}{T^p} G_T^{-1} \left(1 - U_{k,N} - u\right) - F^{-1}\left(1 - U_{k,N}\right) \right) \xrightarrow{p} 0, \tag{2}$$

$$\inf_{u \in \left[1-U_{k,N}, 1\right]} \frac{\sqrt{k}}{c_N} \left(F^{-1}(u) + \frac{1}{T^p} G_T^{-1} \left(1 + 1 - U_{k,N} - u\right) - F^{-1}\left(1 - U_{k,N}\right) \right) \xrightarrow{p} 0. \tag{3}$$

Then as $N, T \to \infty$

$$\sqrt{k} \frac{\vartheta_{N-k,N,T} - F^{-1}\left(1 - k/N\right)}{c_N} \Rightarrow N(0,1). \tag{4}$$

Conditions (2) *and* (3) *are sharp in the following sense: if at least one of them fails, there exists a sequence of joint distributions of* $(\theta_i, \varepsilon_{i,T})$ *with given marginal distributions* F, G_T *such that* $\sqrt{k}\left(\vartheta_{N-k,N,T} - F^{-1}\left(1 - k/N\right)\right)/c_N$ *converges to a limit different from* $N(0,1)$ *or does not converge at all.*

Theorem 3.4 is the intermediate order counterpart of theorem 3.2; conditions (2) and (3) are analogs of conditions (TE-Sup) and (TE-Inf). The key difference between the two pairs of conditions is the region where tail equivalence is imposed. Theorem 3.2 concerns quantiles of the form $1-\tau/N$, τ fixed, whereas theorem 3.4 looks at quantiles of the order $1-k/N$ (as we establish in the proofs, $U_{k,N} = O_p(k/N)$). Since $k \to \infty$, the two regions are asymptotically distinct. A second (minor) point of difference between the two pairs of conditions is that conditions of theorem 3.4 take a randomized form. As we show in the examples and proposition 3.5 below, this randomness is easy to control and exits asymptotically, giving deterministic sufficient conditions. For completeness, in Appendix B we provide a deterministic sufficient condition for (2) and (3).

As for (TE-Inf) and (TE-Sup), sufficient conditions for (2) and (3) take form of rate restrictions on N and T. Before providing a general sufficient condition, we again consider the examples given after theorem 3.2. It is easy to verify that all three example cdfs satisfy assumption 4 (all verifications are given in Appendix A).

Example 5 ($\gamma > 0$, continued)**.** The normalizing constants of theorem 3.4 are $F_{Fr,\kappa}^{-1}(1 - k/N) = (1 - k/N)\left(N/k\right)^{1/\kappa} - 1$ and $c_N = \kappa^{-1}\left(N/k\right)^{1/\kappa}$. Suppose we set $k = N^{\delta}, \delta \in (0,1)$. If $G_T = G_{\beta,T}$, then (2) and (3) hold if for some $\nu > 0$ $N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta-1/\kappa)+\nu/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, then it is sufficient that $N^{\delta/2+(1-\delta)(-1/\kappa)}\sqrt{\log(N)}/T^p \to 0$ for conditions (2) and (3) to be satisfied.

Example 6 ($\gamma = 0$, continued)**.** In this case the constants are simple: $F_{Gu,\lambda}^{-1}(1 - k/N) = \log(N/k)/\lambda$ and $c_N = \lambda^{-1}$. Let $k = N^{\delta}, \delta \in (0,1)$. If $G_T = G_{\beta,T}$, a sufficient condition for conditions (2) and (3) is that for some $\nu > 0$ $N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta)+\nu/\beta}/T^p \to 0$. Similarly, if $G_T = G_{normal,T}$, it is sufficient that $N^{\delta/2}\sqrt{\log(N)}/T^p \to 0$.

Example 7 ($\gamma < 0$, continued)**.** The normalizing constants are given by $F_{W,\alpha}^{-1}(1 - k/N) = \theta_F - \theta_F\left(k/N\right)^{1/\alpha}$ and $c_N = (\theta_F/\alpha)\left(k/N\right)^{1/\alpha}$. Let $k = N^{\delta}, \delta \in (0,1)$. If $G_T = G_{\beta,T}$, it is sufficient that $N^{\delta/2(1+1/\beta)+(1-\delta)(1/\alpha+1/\beta)+\nu/\beta}/T^p \to 0$ for conditions (2) and (3) to hold; if $G_T = G_{normal,T}$ it is sufficient that $N^{\delta/2+(1-\delta)(1/\alpha)}\sqrt{\log(N)}/T^p \to 0$.

Observe that all conditions in the above example can be written in a common form:

$$G_{\beta,T} : \frac{N^{\delta/2(1+1/\beta)+(1-\delta)(-\gamma+1/\beta)+\nu/\beta}}{T^p} \to 0, \quad G_{normal,T} : \frac{N^{\delta/2+(1-\delta)(-\gamma)}\sqrt{\log(N)}}{T^p} \to 0, \quad (5)$$

where ν is any positive number.

Like conditions for theorem 3.2, conditions (5) depend on the underlying distribution F through the EV index γ. However, a sufficient condition is possible if a lower bound for γ is available. The following proposition is an analog of proposition 3.3 for conditions (2) and (3).

Proposition 3.5. *Let assumptions 1 and 3 hold. Let $\delta \in (0,1)$. Let one of the following conditions hold:*

(1) Let $\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$ for some $\beta > 0$, and $N^{\delta/2(1+1/\beta)+(1-\delta)(-\gamma'+1/\beta)+\nu/\beta}/T^p \to 0$ for some γ' and $\nu > 0$.

(2) For all T let $\varepsilon_{i,T} \sim N(\mu_T, \sigma_T^2)$, and let $N^{\delta/2+(1-\delta)(-\gamma')}\sqrt{\log(N)}/T^p \to 0$ for some γ'.

In addition, let F satisfy assumption 4 with EV index $\gamma > \gamma'$. Then conditions (2) *and* (3) *hold for F and G_T for $k = N^\delta$.*

Remark 6 (Comparison of sufficient conditions for the EVT and the IVT)**.** Depending on γ, the rate conditions for the EVT may be more or less restrictive than the conditions for the IVT. For example, let $G_T = G_{normal,T}$. In this case for the EVT to hold it is sufficient that $N^{-\gamma}\sqrt{\log(N)}/T^p \to 0$; for the IVT to hold it is sufficient that $N^{\delta/2+(1-\delta)(-\gamma)}\sqrt{\log(N)}/T^p \to 0$. If $\gamma \leq -1/2$, then the EVT condition implies the IVT condition; the opposite holds if $\gamma > -1/2$. In particular, if $\gamma > 0$, for the EVT there are no restrictions on relative sizes of N and T, but there are restrictions for the IVT.

Remark 7 (Dependence on δ)**.** Conditions (5) depend on δ, the parameter that determines the magnitude of $k = N^\delta$. If δ is close to zero, conditions for the IVT are close to those for the EVT. Intuitively, in this case conditions (2) and (3) require asymptotic tail equivalence in approximately the same section of the tail as conditions (TE-Sup) and (TE-Inf), and so the resulting sufficient conditions are similar.[10] As δ grows, the region controlled by conditions (2) and (3) becomes distinct from the right endpoint of the distribution (while still staying the tails by requirement that $k = o(N)$).

[10]In the limit as $\delta \to 0$, the two sufficient rate conditions become the same (modulo the $N^{\nu/2}$ and $(\log(T))^{1/\beta}$ terms).

4 Inference On Extreme Quantiles Using Noisy Estimates

We now turn to the problem of conducting inference. In this section, we introduce confidence intervals and statistics based on extreme and intermediate order asymptotic approximations. In addition, we briefly discuss the central order approximations of Jochmans and Weidner (2022). The fundamental difference between the three approximations lies in how the quantile of interest is modeled. This choice determines the resulting limit distributions, whose critical values are used to form the confidence intervals of interest or to provide corrected estimators for quantiles.

4.1 Inference Using Extreme Order Approximations

Extreme order approximations use theorem 3.2 as the basis for inference. In order to apply the theorem, we first resolve two challenges. First, we need to allow more general centering constants b_N to accommodate the extreme quantiles of interest. Second, the scaling constants a_N are unknown and cannot be estimated consistently. We focus on those in turn.

For extreme order approximations, the quantile of interest is modeled as drifting to 1 at a rate proportional to N^{-1}, an approach that Müller and Wang (2017) call fixed-k asymptotics. To conduct inference on extreme quantiles, we replace the constants b_N of theorem 3.1 by $F^{-1}(1 - l/N)$ for $l > 0$ fixed; the value of l will generally depend on the quantile of interest and sample size.

Example 8 (Choosing l)**.** Suppose we are interested in the 95th percentile of F. Let $N = 200$. Then l can be obtained by solving $1 - l/N = 0.95$ for $l = 10$ and using that l as an approximation.

With these centering constants, we obtain the following extension of theorem 3.2.

Lemma 4.1. *Let assumptions of theorem 3.2 hold. Let $l > 0$ be fixed.*

(1) Let Fr be a Fréchet random variable with $P(Fr \leq x) = \exp(-x^{-1/\gamma})$ for $x \geq 0$ and 0 for $x < 0$. If F satisfies assumption 2 with EV index $\gamma > 0$, then as $N, T \to \infty$

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

(2) Let Gu be a Gumbel random variable with $P(Gu \leq x) = \exp(-e^{-x})$ for $x \in \mathbb{R}$. If F satisfies assumption 2 with EV index $\gamma = 0$, then as $N, T \to \infty$

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

(3) Let W be a reverse Weibull random variable with $P(W \leq x) = \exp(-(-x)^{-1/\gamma})$ for $x < 0$ and $P(W \leq x) = 1$ for $x \geq 0$. If F satisfies assumption 2 with EV index $\gamma < 0$, then as $N, T \to \infty$

$$\frac{1}{\theta_F - F^{-1}\left(1-\frac{1}{N}\right)}\left(\vartheta_{N,N,T} - F^{-1}\left(1-\frac{l}{N}\right)\right) \Rightarrow W + \left(\frac{1}{l}\right)^{\gamma}.$$

Lemma 4.1 cannot be used for inference directly, as the scaling constants involved are unknown. These constants involve the $(1-1/N)$th quantile of F, and so are not covered by sample information.

To address this challenge, we first establish an intermediate result. The following lemma extends theorem 3.2 to cover a vector of q order statistics for q fixed.

Lemma 4.2 (Joint EVT)**.** *Let assumptions of theorem 3.2 hold. Let q be a fixed natural number and $E_1^*, \dots, E_{q+1}^*$ be iid standard exponential random variables.*

(1) If F satisfies assumption 2 with EV index $\gamma > 0$, then as $N, T \to \infty$

$$\left(\frac{\vartheta_{N,N,T}}{F^{-1}(1-1/N)}, \frac{\vartheta_{N-1,N}}{F^{-1}(1-1/N)}, \dots, \frac{\vartheta_{N-q,N}}{F^{-1}(1-1/N)}\right)$$
$$\Rightarrow \left((E_1^*)^{-\gamma}, (E_1^* + E_2^*)^{-\gamma}, \dots, (E_1^* + E_2^* + \dots + E_{q+1}^*)^{-\gamma}\right).$$

(2) If F satisfies assumption 2 with EV index $\gamma = 0$, then as $N, T \to \infty$

$$\left(\frac{\vartheta_{N,N,T} - F^{-1}\left(1-\frac{1}{N}\right)}{\hat{f}\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}, \frac{\vartheta_{N-1,N} - F^{-1}\left(1-\frac{1}{N}\right)}{\hat{f}\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}, \dots, \frac{\vartheta_{N-q,N} - F^{-1}\left(1-\frac{1}{N}\right)}{\hat{f}\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}\right)$$
$$\Rightarrow \left(-\log(E_1^*), -\log(E_1^* + E_2^*), \dots, -\log(E_1^* + E_2^* + \dots + E_{q+1}^*)\right).$$

(3) If F satisfies assumption 2 with EV index $\gamma < 0$, then as $N, T \to \infty$

$$\left(\frac{\vartheta_{N,N,T} - \theta_F}{\theta_F - F^{-1}(1 - 1/N)}, \frac{\vartheta_{N-1,N} - \theta_F}{\theta_F - F^{-1}(1 - 1/N)}, \dots, \frac{\vartheta_{N-q,N} - \theta_F}{\theta_F - F^{-1}(1 - 1/N)}\right)$$
$$\Rightarrow \left(-(E_1^*)^{-\gamma}, -(E_1^* + E_2^*)^{-\gamma}, \dots, -(E_1^* + E_2^* + \dots + E_{q+1}^*)^{-\gamma}\right).$$

Remark 8. Observe that this result agrees with theorem 3.2: the first coordinate in each case is a standard Fréchet, Gumbel, or reverse Weibull random variable.

Lemma 4.2 allows us to solve the issue of unknown scaling rates by using an observation due to Chernozhukov and Fernández-Val (2011). By taking the ratio of two elements in the joint EVT 4.2, we eliminate scaling factors completely, while the form of the limit is explicitly known up to the EV index γ.

Combining lemmas 4.1 and 4.2, we obtain the following version of the EVT that can be used to conduct inference on extreme quantiles under tail equivalence conditions:

Theorem 4.3 (Feasible EVT)**.** *Let assumptions of theorem 3.2 hold, in particular, let F have EV index $\gamma \in \mathbb{R}$. Let $q \geq 1, r \geq 0$ be fixed natural numbers and $l > 0$ be a fixed real number; let $E_1^*, E_2^*, \dots$ be iid standard exponential RVs. Then as $N, T \to \infty$*

$$\frac{\vartheta_{N,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^*)^{-\gamma} - \left(\frac{1}{l}\right)^{\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}, \tag{6}$$
$$\frac{\vartheta_{N-r,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^* + \dots + E_{r+1}^*)^{-\gamma} - \left(\frac{1}{l}\right)^{\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}, \tag{7}$$

where for $\gamma = 0$

$$\frac{(E_1^*)^{-\gamma} - \left(\frac{1}{l}\right)^{\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}} \equiv \frac{\log(E_1^* + \dots + E_{r+1}^*) - \log(l)}{\log(E_1^* + \dots + E_{q+1}^*) - \log(E_1^*)},$$
$$\frac{(E_1^* + \dots + E_{r+1}^*)^{-\gamma} - \left(\frac{1}{l}\right)^{\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}} \equiv \frac{\log(E_1^*) - \log(l)}{\log(E_1^* + \dots + E_{q+1}^*) - \log(E_1^*)}.$$

In addition, if F satisfies assumption 2 with $\gamma < 0$, then also as $N, T \to \infty$

$$\frac{\vartheta_{N,N,T} - \theta_F}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \Rightarrow \frac{(E_1^*)^{-\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}. \tag{8}$$

Theorem 4.3 allows us to construct confidence intervals and hypothesis tests for extreme quantiles with no knowledge of the value of γ. The left hand sides of eqs. (6)-(8) do not change depending on γ. While the right hand side limit distributions of eqs. (6)-(8) are non-pivotal and do depend on γ, in the next section we show how to estimate the associated critical values. The critical values can can be estimated consistently by subsampling (data resampling) or simulation after plugging in a consistent estimator of γ.

We illustrate the application of theorem 4.3 with a series of examples.

Example 9 (Confidence interval and a median unbiased estimator for maximum)**.** Let $\gamma < 0$, in which case $\theta_F = F^{-1}(1) < \infty$. Suppose we are interested in θ_F. We use convergence relation (8) as the basis of inference. Let $q \geq 1$ be fixed, see remark 9 on choice of q. Let $\hat{c}_\alpha$ be a consistent estimate of the αth quantile of $(E_1^*)^{-\gamma}/((E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma})$. Then

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

is a $(1-\alpha) \times 100\%$ asymptotic confidence interval for θ_F. Since both critical values $\hat{c}_.$ are non-negative, the lower bound of the CI is always above the sample maximum.

In addition, we can also obtain a median-unbiased estimator for θ_F. By theorem 4.3 $P\left((\vartheta_{N,N,T} - F^{-1}(1))/(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}) \leq \hat{c}_{1/2}\right) \to 1/2$. Rearranging, we obtain that the estimator

$$\mathcal{M}_{N,T,1} = \vartheta_{N,N,T} - \hat{c}_{1/2}(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}) \tag{9}$$

is consistent and asymptotically median-unbiased (see Chernozhukov and Fernández-Val (2011) for a similar construction in a quantile regression setting). The effect is obtained by pushing the sample maximum up.

Remark 9 (Choice of q)**.** q can be any positive integer. However, based on simulation evidence, we recommend picking q in the range 2-10. In the context of our simulations, we find that $q = 2$ works best if $\hat{c}_\alpha$ is estimated by subsampling; $q = 3$ performs best with simulation-based critical values.

Example 10 (Estimator and confidence interval for 95th percentile)**.** Suppose we are interested in $F^{-1}(0.95)$ and $N = 200$. There are only 10 observations to the right of the sample quantile, and it is appropriate to use extreme order approximations described by theorem 4.3. Eqs. (6) and (7) provide two slightly different approaches in this case. First, consider eq. (6), we call this approach "max-only" in the simulation study below. As in example 8, set $l = 10$, and let $q \geq 1$ be fixed. Let $\tilde{c}_\alpha$ be a consistent estimator of the αth quantile of $[(E_1^*)^{-\gamma} - 10^{-\gamma}] \, / \left[(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}\right]$. Then let the confidence interval be given by:

$$\widetilde{CI}_\alpha = \left[\vartheta_{N,N,T} - \tilde{c}_{\alpha/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right), \vartheta_{N,N,T} - \tilde{c}_{1-\alpha/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)\right].$$

The second approach is based on eq. (7), we call this the "mixed" approach. In contrast to the max-only CI, here we choose r so that the resulting CI is centered on the sample 95th percentile, and not the sample maximum. This corresponds to $r = 10$. Then let $l = 10$ and q be as before. Let $\hat{c}_\alpha$ be a consistent estimator of the αth quantile of $\left[(E_1^* + \cdots + E_{11}^*)^{-\gamma} - (10)^{-\gamma}\right] / \left[(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}\right]$. Then, using eq. (7) as an approximation, we set as our confidence interval

$$CI_\alpha = \left[\vartheta_{N-10,N,T} - \hat{c}_{\alpha/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right), \vartheta_{N-10,N,T} - \hat{c}_{1-\alpha/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)\right].$$

Care must be exercised in interpreting the asymptotic properties of $\widetilde{CI}_\alpha$ and CI_α: both are $(1-\alpha) \times 100\%$ asymptotic confidence intervals for $F^{-1}(1 - 10/N)$. The target quantity shifts with N, thus l must be picked accordingly to obtain the correct finite sample approximation.

In addition, $\mathcal{M}_{N,T,0.95} = \vartheta_{N-10,N,T} - \hat{c}_{1/2}(\vartheta_{N-q,N,T} - \vartheta_{N,N,T})$ is an asymptotically median-unbiased estimator for $F^{-1}(1 - 10/N)$. In the simulations reported in Appendix B, we find that $\mathcal{M}_{N,T,\cdot}$ slightly outperforms the raw sample quantile and other adjusted estimators considered in terms of mean absolute error. Observe that by construction $\mathcal{M}_{N,T,\cdot}$ is always contained in CI_α.

Example 11 (Hypothesis testing about support)**.** Let $\gamma < 0$. We can also use theorem 4.3 to test hypotheses about the support of F. Suppose we wish to test $H_0 : F^{-1}(1) \leq C$ vs. $H_1 : F^{-1}(1) > C$. Define the test statistic as

$$W_C = \frac{\vartheta_{N,N,T} - C}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}}.$$

The test rejects H_0 if $W_C < \hat{c}_\alpha$ where $\hat{c}_\alpha$ is a consistent estimator of the αth quantile of $(E_1^*)^{-\gamma} - l^{-\gamma}/(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}$. The test is then asymptotically size α and consistent against point alternatives, since $P(W_C < \hat{c}_\alpha | \theta_F = C) \to \alpha$, and for any $\delta > 0$, $P(W_C < \hat{c}_\alpha | \theta_F = C - \delta) \to 0$, $P(W_C < \hat{c}_\alpha | \theta_F = C + \delta) \to 1$.

Remark 10. The statistics of theorem 4.3 are location-scale invariant, hence the confidence intervals based on them are location-scale equivariant. This is a desirable property since θ may be only identified up to location and scale or may depend on scaling of y and x. For example, in the context of example 1, θ_i are sensitive to scale of x and y.

4.2 Estimating Critical Values and γ

We provide two ways to estimate quantiles of the limiting distributions in theorem 4.3: by resampling the data using subsampling and by simulation using a consistent estimator for γ.

Let $q > 1$ and $l \geq 0$ and define $J(x)$ to be the limit distribution in eq. (6) (if $l > 0$) or eq. (8) (if $\gamma < 0$ and $l = 0$):

$$J(x) = P\left(\frac{(E_1^*)^{-\gamma} - l^{-\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}} \leq x \right).$$

Quantiles of distribution of eq. (7) can be obtain by analogous methods.

The first approach to estimating quantiles of J is subsampling the data (Politis and Romano, 1994; Politis et al., 1999). To define the subsampling estimator, split the set of units $\{1, \ldots, N\}$ into all possible subsamples of size b and index the subsamples by s, $s = 1, \ldots, \binom{N}{b}$. Let $\vartheta_{b-k,b,T}^{(s)}$ be the $(b-k)$th order statistic in subsample s. Define the subsampling estimator $L_{b,N,T}$ for J as

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

For eq. (7), $W_{s,b,N,T}$ is replaced by

$$W_{s,b,N,T} = \frac{\vartheta_{b-r,b}^{(s)} - \vartheta_{N-Nl/b,N,T}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}}.$$

Observe that the subsample statistic $W_{s,b,N,T}$ is centered at $\vartheta_{N-Nl/b,N,T}$. This corresponds to the $(1-l/b)$th quantile, giving the correct centering to the subsampled statistics. If we are interested in θ_F, then $l=0$, and the statistic is centered at $\vartheta_{N,N,T}$.

Define the estimated critical value $\hat{c}_\alpha$ as the αth quantile of $L_{b,N,T}$. The following result shows that $\hat{c}_\alpha$ is consistent for the true critical values of interest for all $\alpha \in (0,1)$.

Theorem 4.4 (Consistency of subsampling)**.** *Let $b = N^m, m \in (0,1)$. If $l > 0$, let conditions of propositions 3.3 and 3.5 hold with $\delta = 1-m$. If $\gamma < 0$ and $l = 0$, then let conditions of proposition 3.3 hold. Then the subsampling estimator $L_{b,N,T}(x) \xrightarrow{p} J(x)$ at all $x \geq 0$ and $\hat{c}_\alpha \xrightarrow{p} c_\alpha = J^{-1}(\alpha)$ for all $\alpha \in (0,1)$*

Remark 11. The assumptions imposed and the proof of theorem 4.4 crucially depend on whether we are interested in $\theta_F = F^{-1}(1)$ or in $F^{-1}(q)$ for $q < 1$. If our interest lies in θ_F, consistency of subsampling only requires checking whether the extreme value theorem applies. If we are instead interested in $F^{-1}(q)$ for $q < 1$, the subsample statistics are centered at $\vartheta_{N-Nl/b,N,T}$, which is an intermediate order statistic.[11] It is therefore necessary for both the EVT and the IVT to hold.

The second approach to estimating quantiles of J involves plugging in a consistent estimator for γ and simulating from the resulting distribution. Formally, set $\hat{c}_\alpha = \hat{J}^{-1}(\alpha)$ where $\hat{\gamma}$ is a consistent estimator of γ and

$$\hat{J}(x) = P\left(\frac{(E_1^*)^{-\hat{\gamma}} - l^{-\hat{\gamma}}}{(E_1^* + \cdots + E_{q+1}^*)^{-\hat{\gamma}} - (E_1^*)^{-\hat{\gamma}}} \leq x\right). \tag{10}$$

Since $(E_1^*, \ldots, E_q^*)$ are iid standard exponential variables, it is straightforward to draw samples from $\hat{J}(x)$.

As $\hat{\gamma}$ we recommend using the Pickands (1975) estimator given by

$$\hat{\gamma}_P = \frac{1}{\ln 2} \ln \frac{\vartheta_{N-k,N} - \vartheta_{N-2k,N}}{\vartheta_{N-2k,N} - \vartheta_{N-4k,N}},$$

[11]Centering with an intermediate statistic is similar to the case examined by Chernozhukov and Fernández-Val (2011) in the context of regression.

where $k = k(N)$ satisfies $k \to \infty, k = o(N)$. The following theorem shows that $\hat{\gamma}_P$ is consistent.

Theorem 4.5. *Let $k = k(N) \to \infty, k = o(N)$ and let assumption of theorem 3.4 hold for k. Then $\hat{\gamma}_P \xrightarrow{p} \gamma$ as $N, T \to \infty$.*

Remark 12. $\hat{\gamma}_P$ has several properties that are desirable in our framework. First, it is consistent regardless of the value of γ. Second, it is scale and location invariant. Finally, as shown by theorem 4.5, the consistency requirement in our noisy case are just the conditions of theorem 3.4, requiring no extra rate conditions.

Remark 13. If $\gamma < 0$ and $l = 0$, it may be desirable to replace positive values of $\hat{\gamma}_P$ by $-1/N$. This truncation does not affect consistency and may prevent erroneous division by zero in $l^{-\hat{\gamma}_P}$ if $\hat{\gamma}_P > 0$.

4.3 Inference Using Intermediate Order Approximations

The intermediate order approximation of theorem 3.4 provides an alternative approach to inference that is based on convergence of intermediate order statistics (eq. (4)). In this case, the quantile of interest is modeled as drifting to 1 at a rate k/N where $k \to \infty, k = o(N)$ as $N \to \infty$; this rate is slower than the rate N^{-1} of extreme order approximations. The resulting statistic is asymptotically standard normal. Intermediate order approximations should not be applied to $F^{-1}(\tau)$ for τ very close to 1 (see remark 14 below).

As the scaling rate c_N of statistic (4) is unknown, we follow an approach similar to the feasible EVT 4.3 and use a second intermediate order statistics to eliminate c_N.[12] By using a pair of statistics which converge to the same random variable, we can exploit a knife-edge

[12] There are alternative approaches to inference using intermediate order statistics. For example, the subsampling approach of Bertail et al. (1999, 2004) can be used to account for unknown scaling rate c_N. However, the presence of a slowly varying component in c_N requires using multiple subsampling sizes, which may be problematic if N is not extremely large. Alternatively, for example, see ch. 4 of de Haan and Ferreira (2006) for inference under a second-order condition in a setting without estimation noise.

bias created by the difference in scaling factors. The following theorem first establishes that such a technique works in the noiseless case, which may be of independent interest; the result is then transferred to noisy observables.

Theorem 4.6. *Let assumptions 1 and 4 hold. Let $k = o(N), k \to \infty$. Let $f = F'$ be non-increasing or non-decreasing in some left neighborhood of θ_F ($\theta_F \le \infty$).*

(1) Then

$$\frac{\theta_{N-k,N} - F^{-1}\left(1 - \frac{k}{N}\right)}{\theta_{N-k,N} - \theta_{N-k-\lfloor\sqrt{k}\rfloor,N}} \Rightarrow N(0,1), \quad N \to \infty.$$

(2) In addition, let conditions of theorem 3.4 hold when evaluated at k and $k + \sqrt{k}$. Then

$$\frac{\vartheta_{N-k,N} - F(1 - k/N)}{\vartheta_{N-k,N} - \vartheta_{N-k-\lfloor\sqrt{k}\rfloor,N}} \Rightarrow N(0,1), \quad N, T \to \infty.$$

In addition to the von Mises condition, we also require monotonicity of the density in some region around the endpoint. All the distributions of examples 5-7 satisfy this condition.

The choice of k is determined by the quantile of interest. When k is chosen, the denominator is uniquely determined by k. This is in contrast to the free choice of the parameter q in the statistics of theorem 4.3.

Example 12. Suppose that we are interested in the 92nd percentile of F, and let $N = 200$. Then k is obtained by solving $1 - k/N = 0.92$ for $k = 16$. By theorem 4.6, a confidence interval for $F^{-1}(0.92)$ is given by

$$\left[\vartheta_{N-k,N} - z_{1-\alpha/2}\left(\vartheta_{N-k,N} - \vartheta_{N-k-\lfloor\sqrt{k}\rfloor,N}\right), \vartheta_{N-k,N} - z_{\alpha/2}\left(\vartheta_{N-k,N} - \vartheta_{N-k-\lfloor\sqrt{k}\rfloor,N}\right)\right]$$

for $N = 200, k = 16, \sqrt{k} = 4$, and z_α the αth quantile of the standard normal distribution. This is an asymptotic $(1-\alpha)$ confidence interval for $F^{-1}(1 - k/N)$ where k is parameterized so that $k(200) = 16$ and $k \to \infty, k = o(N)$ as $N \to \infty$.

Remark 14. The approximation of theorem 4.6 should not be used if k is small. If $\sqrt{k}$ is small, $\vartheta_{N-k,N}$ and $\vartheta_{N-k-\sqrt{k},N}$ will be close, and the statistic will be unstable. In addition, theorem 4.6 cannot be used for inference on the maximum, as this violates the assumption that k is intermediate.

4.4 Inference Using Central Order Approximations

The third method of inference is based on the central limit theorem for quantiles. The quantile of interest $F^{-1}(\tau)$ is modeled as fixed and independent from sample size (N, T), in contrast to the extreme and intermediate approximations given above. Typically, such "central" order approximations require that a sufficient number of observations be available on both sides of the corresponding sample order statistic $\vartheta_{\lfloor N\tau \rfloor, N, T}$. In particular, this approach cannot be applied to $\theta_F = F^{-1}(1)$.

Jochmans and Weidner (2022) derive such approximations in the context of our problem, and we briefly state their results. They study a version of (1) given by $\vartheta_{i,T} = \theta_i + T^{-1/2}\varepsilon_i$ (that is, $p = 1/2$ and $G_T = G$ for all T). We introduce some additional notation: let K be a kernel function, h a bandwidth parameter, and define

$$\sigma_i^2 = \operatorname{Var}(\varepsilon_i|\theta_i), \quad b_F(x) = \left[\frac{\mathbb{E}\left[\sigma_i^2|\theta_i = x\right] f(t)}{2}\right]'$$
$$\hat{b}_F = -\frac{(nh^2)^{-1}\sum_{i=1}^n \sigma_i^2 K'((\vartheta_{i,T} - \theta)/h)}{2}, \quad \hat{\tau}^* = \tau + \frac{\hat{b}_F(\vartheta_{\lfloor N\tau \rfloor, N, T})}{T}.$$

σ_i^2 is assumed known and invariant over time.

Theorem 4.7 (Propositions 2 and 4 of Jochmans and Weidner (2022))**.** *Let conditions of proposition 3 in Jochmans and Weidner (2022) hold, and in particular let for all T $\varepsilon_{i,T} = \varepsilon_i$, $\mathbb{E}(\varepsilon_i|\theta_i) = 0$, and ε_i be independent from θ_i given* $\operatorname{Var}(\varepsilon_i|\theta_i)$*. Let $\tau \in (0, 1)$.*

(1) If $N/T^2 \to c < \infty$, then as $N, T \to \infty$

$$\sqrt{N}\left(\vartheta_{\lfloor N\tau \rfloor, N, T} - F^{-1}(\tau) + \frac{1}{T}\frac{b_F(F^{-1}(\tau))}{f(F^{-1}(\tau))}\right) \Rightarrow N\left(0, \frac{\tau(1-\tau)}{f(F^{-1}(\tau))^2}\right).$$

(2) If $N/T^4 \to 0$, then as $N, T \to \infty$

$$\sqrt{N}\left(\vartheta_{\lfloor N\hat{\tau}^* \rfloor, N, T} - F^{-1}(\tau)\right) \Rightarrow N\left(0, \frac{\tau(1-\tau)}{f(F^{-1}(\tau))^2}\right).$$

As theorem 4.7 shows, the sample τth quantile $\vartheta_{\lfloor N\tau \rfloor, N, T}$ is a consistent and asymptotically normal estimator for $F^{-1}(\tau)$ with standard asymptotic variance. However, $\vartheta_{\lfloor N\tau \rfloor, N}$ is subject to bias of leading order $1/T$. Jochmans and Weidner (2022) show that this bias can be reduced by instead considering the sample $\hat{\tau}^*$th quantile: $\vartheta_{\lfloor N\hat{\tau}^*\rfloor,N,T}$ is consistent and asymptotically normal with the same variance, but the leading order of the bias is instead given by $1/T^2$. This bias is eliminated if $\sqrt{N}/T^2 \to 0$.

Remark 15 (Comparison of rate conditions)**.** The rate condition $N/T^4 \to 0$ of theorem 4.7 is the same for a broad class of distributions. In contrast, the rate conditions of theorem 3.2 fundamentally depend on F through γ and a_N. As a consequence, the rate conditions for central order approximations may be stronger or weaker than those for extreme approximations (see also remark 6). For example, let F have an infinite tail ($\gamma \geq 0$), $p = 1/2$, and let G_T have more than 8 moments. Then the condition $N/T^4 \to 0$ of theorem 4.7 is stronger than the sufficient conditions of proposition 3.3. In this case an extreme order approximation will apply even if a central order approximation does not. If $\gamma < 0$, the opposite may happen.

Remark 16. For central order approximations, the order of the bias incurred by using $\vartheta_{\lfloor N\tau\rfloor,N,T}$ in place of $\theta_{\lfloor N\tau\rfloor,N}$ is the same for a broad class of distributions, and equal to 1/T. This invariance of bias order enables construction of the debiased estimator $\vartheta_{\lfloor N\hat{\tau}^*\rfloor,N,T}$. The situation is more complex for extreme and intermediate order approximations. The magnitude of the impact of estimation noise is determined by the interaction of a_N and T in theorem 3.2; a_N itself may behave like N^γ for $\gamma \in \mathbb{R}$ depending on F, up to slowly varying components.

5 Simulations

5.1 Design

We illustrate the performance of our confidence intervals with a simulation study. We consider the setup of example 1:

$$y_{it} = \alpha_i + \eta_i x_{it} + \theta_i z_{it} + \sqrt{\frac{\mathrm{Var}(\theta_i)}{\mathrm{Var}(u_{it})}} \times u_{it}.$$

The parameter of interest θ_i is the coefficient on z_{it}. We are interested in coverage of a nominal 95% confidence for a collection of quantiles close to 1. For distributions with infinite endpoints we examine the 0.85-0.999th quantiles. For distributions with a finite endpoint, we consider the 0.85-1st quantiles. In Appendix B we also explore performance of the corrected estimators formed as in eq. (9).

We consider $\theta_i \sim F_{Fr,\kappa}, F_{Gu,\lambda}, F_{W,\alpha}$, where the distributions are defined in examples 5-7. As highlighted before example 5 in section 3, these are prototypical examples of all F that satisfy assumption 2. Thus, the results of this simulation study have implications for all F that satisfy assumption 2. We take $\kappa = 4, \lambda = 1, \alpha = 4$; for $F_{W,\alpha}$ we set $\theta_F = 10$. Coefficients $(\alpha_i, \eta_i, \theta_i)$ are drawn from a trivariate Gaussian copula with correlation 0.3 and the same marginals F. Covariates x_{it} are generated as $0.3\eta_i + \sigma_x \left(1 + 0.3 \left\|(\alpha_i, \eta_i, \theta_i)\right\|\right)^{1/2} (0.1 + U)$ where U is a Uniform[0, 1]; z_{it} are generated similarly with θ_i in place of η_i. This is a stylized setup in which UIH and covariates are positively dependent. u_{it} is sampled independently from the coefficients and the covariates from G_β, where $G_\beta = G_{\beta,T}$ with $\mu_T = 0$, $G_{\beta,T}$ is defined in example 5, $\mu_T = 0$, and $\beta = 3$. In Appendix B we also report results for $u_{it} \sim N(0, 1)$. Observe that we rescale u_{it} so that its variance matches that of θ_i. Coefficients are estimated using OLS. As a result, the estimation noise $\varepsilon_{i,T}$ is not independent from θ_i. All results are based on 15000 Monte Carlo samples.

To assess the importance of rate conditions we derive, we consider two moderate sample sizes: $N = 200, T = 15$ and $N = 200, T = 30$. The rate conditions of theorem 4.3 hold safely for $F = F_{Fr,\kappa}, F_{Gu,\lambda}$ for both $T = 15, 30$ (by examples 5 and 6 there are effectively no restrictions in these two cases). However, for $F = F_{W,\alpha}$ the associated ratio of N and T is quite large for $T = 15$; in this case theorem 4.3 might not hold. We judge the ratio to be small for $T = 30$, so that theorem 4.3 and other results can be applied safely.

We construct the confidence intervals using approximations given in section 4:

(1) Subsampled: based on the feasible EVT 4.3. Quantiles of the limit distribution are estimated using subsampling as in theorem 4.4. We report the two CIs proposed in example 10. The "max only" CI is centered at the maximum for all quantiles, and only the critical values track the target quantile. The "mixed" CI is instead centered at the corresponding sample quantile for each quantile of interest. We pick $q = 2$.

(2) Simulated: based on the feasible EVT 4.3. Quantiles of the limit distribution are estimated by simulating from eq. (10) with γ estimated as in theorem 4.5. We only report the "mixed" CI in the main text. We set $q = 3$.

(3) Intermediate: based on the normal approximation of theorem 4.6

(4) Central: raw data. We construct the CI using the raw $\vartheta_{i,T}$ using a binomial CI.[13]

(5) Central with the analytical correction of Jochmans and Weidner (2022). Based on theorem 4.7: we report the CI centered $\vartheta_{\lfloor N\hat{\tau}^* \rfloor, N, T}$. As advocated by Jochmans and Weidner (2022), we report bootstrap confidence intervals with 1000 bootstrap samples.

For estimating $F^{-1}(1)$, only the first two methods may be applied.

5.2 Results

Figures 1 and 2 visually state our results. We plot coverage rates for the above confidence intervals for $T = 15$ for all three F considered; in addition, for the case of $F = F_{W,\alpha}$ ($\gamma < 0$), we plot $T = 30$. In Appendix B we provide full results for $T = 15, 30$ for all combinations of distributions and additional methods of constructing confidence intervals.

For inference on extreme quantiles, we recommend using extreme approximations with quantiles of the limiting distribution estimated by subsampling. Coverage of such confidence intervals is represented by lines "Subsampling: max only" and "Subsampling: mixed" on figures 1 and 2. The two confidence intervals offer favorable coverage performance for all quantiles considered when the associated rate conditions holds (fig. 1 and bottom panel of fig. 2). Based on length properties reported in Appendix B, the mixed approach displays best overall performance, and we recommend using it. We advise against using simulated quantiles, as such intervals have uniformly worse coverage. This effect is primarily due to the relatively small N, with the asymptotic distributions of theorem 4.3 providing a somewhat poor approximation to the finite sample case.

The rate conditions of theorem 4.3 are essential for correct performance of our approximations when noise is present. This is evident on fig. 2. When $T = 15$, estimation noise

[13] Same approach is implemented in the Stata command `centile`, for example.

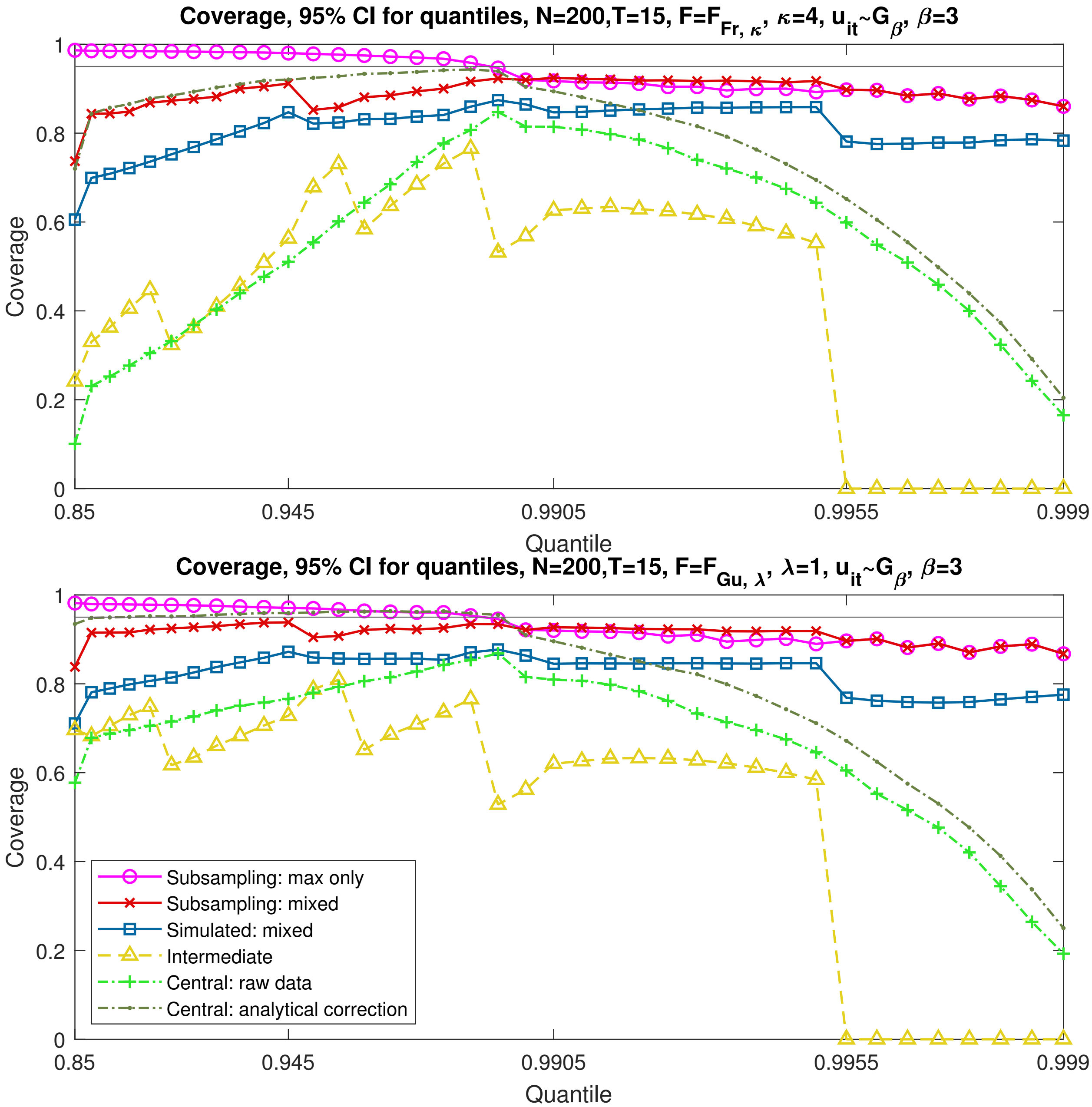


Figure 1: Coverage for 95% nominal confidence interval, different approximations. Distributions with infinite tail. Top: $F = F_{Fr,\kappa}, \kappa = 4$ (power law tail). Bottom: $F = F_{Gu,\lambda}, \lambda = 1$ (exponentially light tail). Both panels: $u_{it} \sim G_\beta, \beta = 3$, $N = 200, T = 15$.

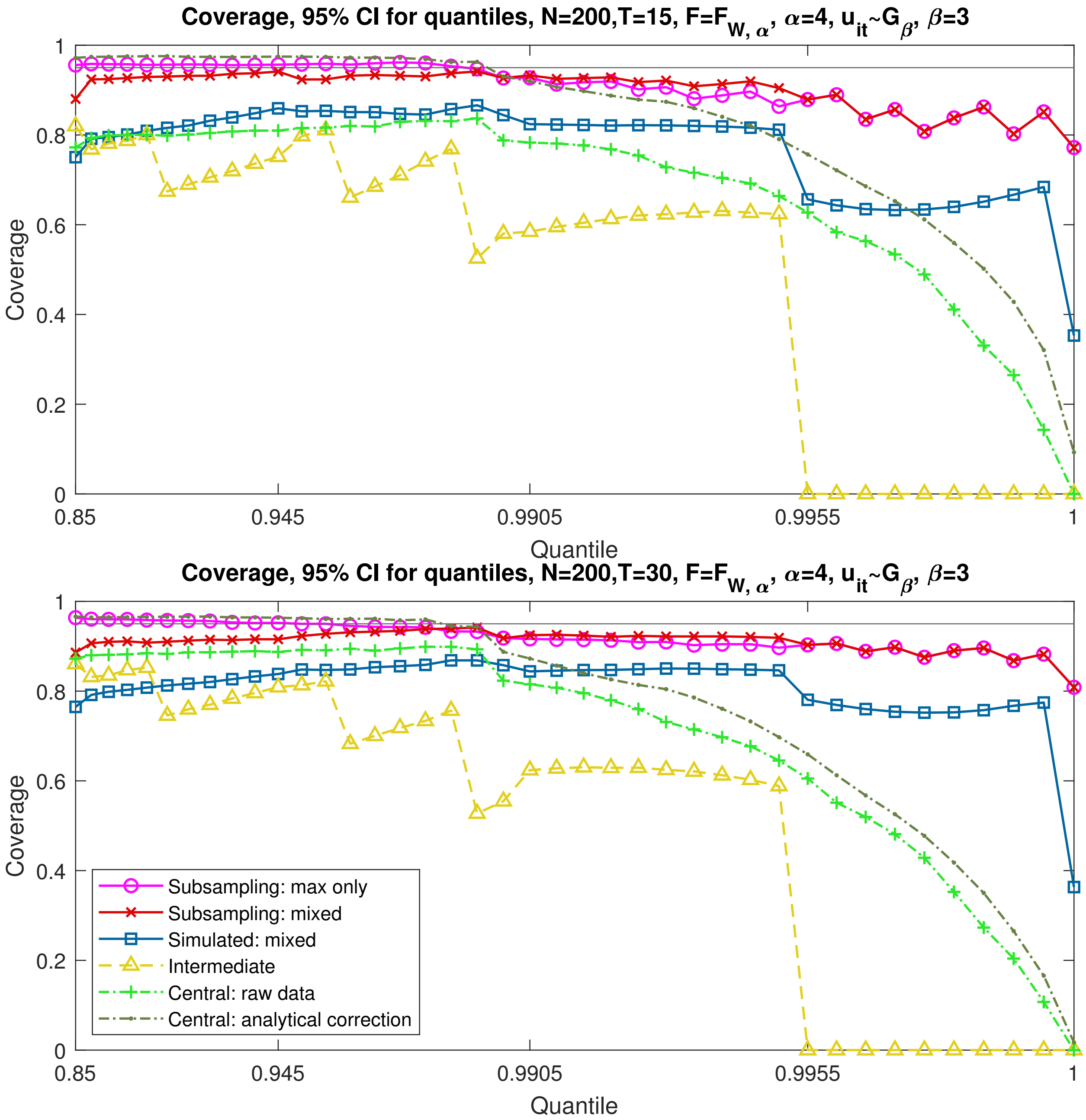


Figure 2: Coverage for 95% nominal confidence interval, different approximations. Distribution with finite tail. Both panels: $F = F_{W,\alpha}, \alpha = 4,\ u_{it} \sim G_\beta, \beta = 3, N = 200$. Top: $T = 15$. Bottom: $T = 30$.

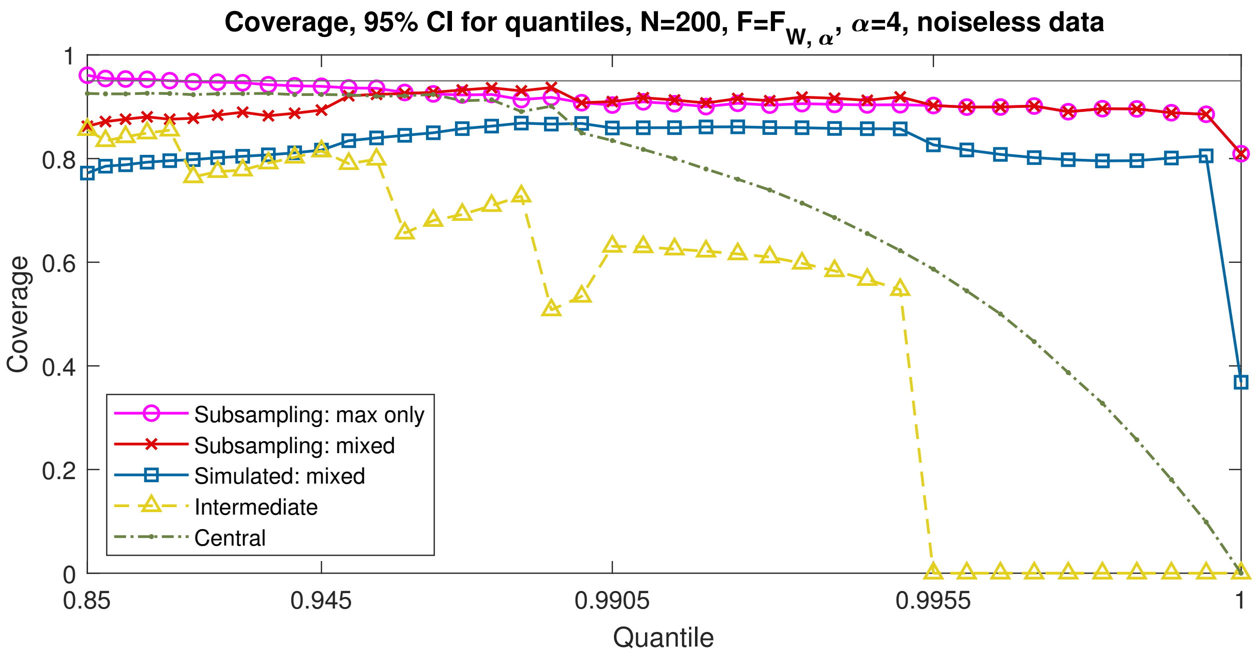


Figure 3: Noiseless data: coverage for 95% nominal confidence interval, different approximations. Distribution with finite tail. $F = F_{W,\alpha}, \alpha = 4$. Note: analytical correction and raw data CIs are identical due to lack of estimation noise.

exercises a noticeable effect on the coverage of extreme approximations. However, when $T = 30$, this effect is largely eliminated. The remaining size distortions are minor and are due to small cross-sectional size effects; raising T further does not improve performance of extreme approximations. This effect is evident on fig. 3 where the different approximations are applied directly to θ_i for $\theta_i \sim F_{W,\alpha}$ (effectively $T = \infty$).[14]

Central approximations suffer from the two problems highlighted in the introduction: effect of estimation noise and failure of central approximations in the tail. The first issue is evident in generally poor performance of CIs based on raw data for quantiles below 0.99. Applying the analytical correction of Jochmans and Weidner (2022) uniformly improves coverage for all such quantiles. The second issue is common to both using raw data and using the correction, and coverage and length of confidence intervals based on central approximations collapse to 0 for high quantiles.

[14]We report results using noiseless data for all distribution of θ_i considered in Appendix B.

Confidence intervals based on the intermediate approximation of theorem 3.4 generally fare poorly. We conjecture that this is so for two reasons. First, it is well-established in the EV theory literature that approximations based on intermediate order statistics suffer from bias when used to estimate quantiles (e.g. Li et al. (2011); Cai et al. (2013) and ch. 3 and 4 in de Haan and Ferreira (2006)). Including a correction term in the numerator might potentially improve coverages. Second, the rate of convergence in distribution for the distribution of the IVT is controlled by $k = o(N)$, and in samples considered here the values of k available might be insufficient for normality to be a good approximation.

6 Concluding Remarks

In this paper, we consider the issue of conducting inference on extreme quantiles of unobserved individual heterogeneity (UIH). This question arises in analysis of economic panel data and meta-analysis. Inference is challenging, as only estimates of UIH are available and the associated estimation noise may have a complex relationship with the heterogeneity of interest. We establish sharp conditions under which the observed noisy estimates can be used to conduct inference on the unobserved quantiles of interest. Building on these conditions, we develop extreme and intermediate extreme order approximations, and propose confidence intervals, hypothesis tests, and corrected estimators. In simulations, we find that confidence intervals based extreme order approximations have favorable coverage properties even with moderate sample sizes, provided the associated critical values are estimated by subsampling.

References

M. Arellano and S. Bonhomme. Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79(3):987–1020, 2012.

L. Barras, P. Gagliardini, and O. Scaillet. Skill, Scale and Value Creation in the Mutual Fund Industry. 2021.

P. Bertail, D. N. Politis, and J. P. Romano. On Subsampling Estimators with Unknown Rate of Convergence. *Journal of the American Statistical Association*, 94(446):569–579, 1999.

P. Bertail, C. Haefke, D. N. Politis, and H. White. Subsampling The Distribution of Diverging Statistics with Applications to Finance. *Journal of Econometrics*, 120(2):295–326, 2004.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987.

J. Breitung and N. Salish. Estimation of Heterogeneous Panels with Systematic Slope Variations. *Journal of Econometrics*, 220(2):399–415, 2021.

M. Browning and J. M. Carro. Heterogeneity and Microeconometrics Modeling. *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume III*, pages 47–74, 2007.

M. Browning and J. M. Carro. Heterogeneity in dynamic discrete choice models. *The Econometrics Journal*, 13(1):1–39, 2010.

J. J. Cai, L. de Haan, and C. Zhou. Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201, 2013.

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

V. Chernozhukov and I. 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.

P. P. Combes, G. Duranton, L. Gobillon, D. Puga, and S. Roux. The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6): 2543–2594, 2012.

L. de Haan and A. Ferreira. *Extreme Value Theory*. Springer, 2006.

A. L. Dekkers and L. de Haan. On the Estimation of the Extreme-Value Index and Large Quantile Estimation. *The Annals of Statistics*, 17(4):1795–1832, 1989.

B. Efron. Tweedie's Formula and Selection Bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Y. Fan and S. S. Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010.

S. Firpo and G. Ridder. Partial identification of the treatment effect distribution and its functionals. *Journal of Econometrics*, 213(1):210–234, 2019.

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

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

B. V. Gnedenko. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

J. Heckman and E. Vytlacil. Instrumental variables methods for the correlated random coefficient model. *Journal of Human Resources*, 33(4):974–987, 1998.

J. J. Heckman. Micro Data, Heterogeneity and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy*, 109(4):673–748, 2001.

J. J. Heckman, J. Smith, and N. 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.

J. P. Higgins, S. G. Thompson, and D. J. Spiegelhalter. A Re-Evaluation of Random-Effects Meta-Analysis. *Journal of the Royal Statistical Society: Series A*, 172(1):137–159, 2009.

K. Jochmans and M. Weidner. Inference on a Distribution From Noisy Draws. *Econometric Theory*, pages 1–38, 2022.

M. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer Series in Statistics, 1983.

D. Li, L. Peng, and X. Xu. Bias reduction for endpoint estimation. *Extremes*, (April 2010): 393–412, 2011.

G. Makarov. Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed. *Theory of Probability Amd Its Applications*, 26(4):803–806, 1981.

U. K. Müller and Y. Wang. Fixed-k Asymptotic Inference About Tail Properties. *Journal of the American Statistical Association*, 112(519):1334–1343, 2017.

K. Nagashima, H. Noma, and T. A. Furukawa. Prediction Intervals for Random-Effects Meta-Analysis: A Confidence Distribution Approach. *Statistical Methods in Medical Research*, 28(6):1689–1702, 2019.

C. R. Nelson and R. Startz. Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 58(4):967–976, 1990.

R. Okui and T. Yanagi. Kernel estimation for panel data with heterogeneous dynamics. *The Econometrics Journal*, 23(1):156–175, 2019a.

R. Okui and T. Yanagi. Panel Data Analysis with Heterogeneous Dynamics. *Journal of Econometrics*, 212(2):451–475, 2019b.

M. H. Pesaran and R. P. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 6061:473–477, 1995.

J. Pickands. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3 (1):119–131, 1975.

D. N. Politis and J. P. Romano. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994.

D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer Series in Statistics, 1999.

A. Rényi. On The Theory Of Order Statistics. *Acta Mathematica Academiae Scientiarum Hungarica*, 4:191–231, 1953.

S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 1987.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Ltd, 1980.

A. Weinstein, Z. Ma, L. D. Brown, and C. H. Zhang. Group-Linear Empirical Bayes Estimates for a Heteroscedastic Normal Mean. *Journal of the American Statistical Association*, 113 (522):698–710, 2018.

Y. Zhang. Extremal Quantile Treatment Effects. *The Annals of Statistics*, 46(6):3707–3740, 2018.

Appendix A

Proofs of Results in the Main Text

A.1 Distributional Results

Proof of theorem 3.1. Theorem 3.1 in the form stated is given by corollary 1.2.4 in de Haan and Ferreira (2006), to whom we refer for a proof. □

A.1.1 Proof of Theorem 3.2

We now turn to the proof of theorem 3.2. We begin by stating some auxiliary results.

Let H_T be the cdf of $\vartheta_{i,T} = \theta_i + \varepsilon_{i,T}$. Observe that $\{\vartheta_{i,T}\}$ form a triangular array with rows indexed by T, and number of entries in each row given by N. In each row, the entries are iid and distributed according to the cdf H_T.

Define the auxiliary function U_T as

$$U_T = \left(\frac{1}{1-H_T}\right)^{-1}. \tag{A.1.1}$$

Using U_T greatly simplifies notation in subsequent proofs. Observe the following useful connection between U_T and H_T. Let $\tau \in (0, \infty)$. Let N be large enough so that $N\tau > 1$. Then

$$\begin{aligned} U_T(N\tau) &= \inf\left\{x : \frac{1}{1-H_T(x)} \geq N\tau\right\} \\ &= H_T^{-1}\left(1 - \frac{1}{N\tau}\right). \end{aligned} \tag{A.1.2}$$

Similarly, define U_F with respect to F, the cdf of θ_i:

$$U_F = \left(\frac{1}{1-F}\right)^{-1}. \tag{A.1.3}$$

We begin with a technical lemma that connects convergence of the normalized sample maximum $\vartheta_{N,N,T} = \max\{\vartheta_{1,T}, \dots, \vartheta_{N,T}\}$ and convergence of the quantiles of H_T.

Lemma A.1.1. *Let assumption 1 hold. The following are equivalent:*

1. *As* $N, T \to \infty$*, for some constants* a_N, b_N *the random variable* $(\vartheta_{N,N,T} - b_N)/a_N$ *converges weakly to a random variable* X *with non-degenerate cdf* $Q(x)$*.*

2. *As* $N, T \to \infty$*, for all* $\tau \in (0, \infty)$ *such that* τ *is a continuity of point* $Q^{-1}(\exp(-1/\tau))$*, it holds that*
$$\frac{U_T(N\tau) - b_N}{a_N} \Rightarrow Q^{-1}(\exp(-1/\tau))$$

Observe that the lemma is also true for $\theta_{N,N}$ with U_F in place of U_T.

The proof is completely analogous to the proof of theorem 1.1.2 in de Haan and Ferreira (2006), which establishes a similar result for the iid case.

The following lemma provides a condition on U_T under which $\vartheta_{N,N,T}$ inherits the limit properties of $\theta_{N,N}$.

Lemma A.1.2. *If*

1. *Assumption 1 holds*

2. a_N, b_N *are such that as* $N \to \infty$ *the normalized noiseless maximum* $a_N^{-1}(\theta_{N,N} - b_N)$ *converges weakly to a non-degenerate random variable* X*.*

3. *For each* $\tau \in (0, \infty)$
$$\frac{U_T(N\tau) - U_F(N\tau)}{a_N} \to 0 \text{ as } N, T \to \infty,$$

then as $N, T \to \infty$
$$\frac{\vartheta_{N,N,T} - b_N}{a_N} \Rightarrow X.$$

Proof. Let Q be the cdf of X. Let $\tau \in (0, \infty)$ be a continuity point of $Q^{-1}(\exp(-1/\tau))$. Then by lemma A.1.1 $a_N^{-1}(U_F(N\tau) - b_N)$ converges to $Q^{-1}(\exp(-1/\tau))$. By assumption (3) of the lemma for all $\tau \in (0,\infty)$ continuity points of $Q^{-1}(\exp(-1/\tau))$ it holds that

$$\frac{U_T(N\tau)-b_N}{a_N}-\frac{U_F(N\tau)-b_N}{a_N}=\frac{U_T(N\tau)-U_F(N\tau)}{a_N}\to 0 \text{ as } N,T\to\infty.$$

From this we conclude that $a_N^{-1}(U_T(N\tau)-b_N)\to Q^{-1}(\exp(-1/\tau))$ for all $\tau\in(0,\infty)$ continuity points of $Q^{-1}(\exp(-1/\tau))$. By lemma A.1.1, this implies that $a_N^{-1}(\vartheta_{N,N,T}-b_N)\Rightarrow X$ as $N,T\to\infty$. □

We will make use of the following quantile inequalities due to Makarov (1981).

Lemma A.1.3 (Makarov quantile inequalities; eqs. (1) and (2) in Makarov (1981))**.** *Suppose that $X\sim F_X$ and $Y\sim F_Y$ are a pair of random variables whose joint distribution is not restricted, and consider their sum $X+Y\sim F_{X+Y}$. The following inequalities hold: for all $w\in[0,1]$*

$$F_{X+Y}^{-1}(w)\le \inf_{u\in[w,1]}\left(F_X^{-1}(u)+F_Y^{-1}(1+w-u)\right),$$
$$F_{X+Y}^{-1}(w)\ge \sup_{u\in[0,w]}\left(F_X^{-1}(u)+F_Y^{-1}(w-u)\right).$$

The bounds are pointwise sharp in the following sense: for each w there exists joint distributions of X and Y such that the wth quantile of $X+Y$ attains the lower/upper bound at w.

We now give a proof of theorem 3.2.

Proof of theorem 3.2. Suppose that for some sequence of constants a_N, b_N the noiseless maximum $a_N^{-1}(\theta_{N,N}-b_N)$ converges to some nondegenerate random variable X.

Let H_T be the cdf of $\vartheta_{i,T}$. In the statement of Makarov's inequalities (lemma A.1.3), take $X=\theta_i$ and $Y=T^{-p}\varepsilon_{i,T}$. By Makarov's inequalities we obtain that at any $w\in[0,1]$

$$\begin{aligned}&\sup_{u\in[0,w]}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}(w-u)\right)\\ \le& H_T^{-1}(w)\\ \le& \inf_{u\in[w,1]}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}(1+w-u)\right).\end{aligned}\tag{A.1.4}$$

Fix $\tau \in (0, \infty)$. Let N be large enough so that $N\tau > 1$. Suppose that $a_N > 0$ for N large enough (if not, simply reverse all inequalities below; also recall that a_N is monotonic). Take $w = 1 - 1/N\tau$, subtract $F^{-1}(1 - 1/N\tau)$ on all sides, and multiply by a_N^{-1} to obtain

$$\begin{aligned}
&\sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1 - \frac{1}{N\tau} - u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right) \\
\leq & \frac{1}{a_N}\left(H_T^{-1}\left(1 - \frac{1}{N\tau}\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right) \qquad (A.1.5)\\
\leq & \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1 + 1 - \frac{1}{N\tau} - u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right) .
\end{aligned}$$

Recall definitions of U_T and U_F (eqs. (A.1.1) and (A.1.3)) and observe that by eq. (A.1.2)

$$\frac{1}{a_N}\left(H_T^{-1}\left(1 - \frac{1}{N\tau}\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right) = \frac{U_T(N\tau) - U_F(N\tau)}{a_N}.$$

Substitute this expression in eq. (A.1.5). Then conditions (TE-Inf) and (TE-Sup) imply that

$$\frac{U_T(N\tau) - U_F(N\tau)}{a_N} \to 0 \qquad (A.1.6)$$

for all $\tau \in (0, \infty)$. Then by lemma A.1.2 it follows that

$$\frac{\vartheta_{N,N,T} - b_N}{a_N} \Rightarrow X.$$

Now we turn to the second assertion. Suppose that at least one of (TE-Inf) and (TE-Sup) fails. We show that the limit properties of $a_N^{-1}(\vartheta_{N,N,T} - b_N)$ differ from the limit properties of $a_N^{-1}(\theta_{N,N} - b_N)$ for some sequence of joint distributions of $(\theta_i, \varepsilon_{i,T})$. Suppose it is the infimum condition (TE-Inf) that fails to hold for some τ; the argument for (TE-Sup) is identical. Then along some subsequence of (N, T) it holds that $\inf_{u\in[1-1/N\tau,1]} a_N^{-1}(\cdots) = \delta_{N,T}$ such that $\delta_{N,T}$ are bounded away from zero. Suppose that it is possible to extract a further subsequence such that along it $\delta_{N,T}$ converges to some $\delta \neq 0$. Pass to that subsubsequence. Theorem 2 in Makarov (1981) establishes that for each (N, T) there exists a joint distribution of θ_i and $\varepsilon_{i,T}$ such that the resulting H_T attains the upper bound in inequality (A.1.4):

$$H_T^{-1}(w) = \inf_{u\in[w,1]}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}(1 + w - u)\right)$$

$$= a_N \delta_{N,T} + F^{-1}\left(1 - \frac{1}{N\tau}\right).$$

If $(U_F(N\tau) - b_N)/a_N \to Q(\exp(-1/\tau))$, then for such a sequence of H_T

$$\frac{U_T(N\tau) - b_N}{a_N} \to Q^{-1}(\exp(-1/\tau)) + \delta$$

Suppose that $(\vartheta_{N,N,T} - b_N)/a_N$ converges to a random variable with distribution $\tilde{Q}$ along the same subsequence. The above discussion shows that $\tilde{Q}^{-1}(\exp(-1/\tau)) = Q^{-1}(\exp(-1/\tau)) + \delta \neq Q^{-1}(\exp(-1/\tau))$. Thus, either the limit distribution of $\vartheta_{N,N,T}$ is different from that of $\theta_{N,N}$, or $\vartheta_{N,N,T}$ does not converge.

If we cannot extract a subsequence of $\delta_{N,T}$ converging to some finite δ, then $\delta_{N,T}$ is unbounded. In this case it is possible to extract a further monotonically increasing subsequence. Proceeding as above, we obtain that along that subsequence it holds that

$$\frac{U_T(N\tau) - b_N}{a_N} \to Q^{-1}(\exp(-1/\tau)) + \infty,$$

and so $a_N^{-1}(\vartheta_{N,N,T} - b_N)$ does not converge. □

A.1.2 Proof of Proposition 3.3

Before proving proposition 3.3, we state two useful results.

First, let RV_γ be the class of non-negative functions of regular variation with parameter γ, that is, those measurable $f : \mathbb{R}_+ \to \mathbb{R}_+$ that satisfy $\lim_{t\to\infty} f(tx)/f(t) = x^\gamma$ for any $x > 0$. RV_0 is the class of slowly varying functions.

Lemma A.1.4 (Karamata Characterization Theorem; Theorem 1.4.1 in Bingham et al. (1987))**.** *Let $f \in RV_\gamma$. Then there exists a slowly varying function L (that is, $L \in RV_0$) such that for all x*

$$f(x) = x^\gamma L(x).$$

Second, we observe that theorem 3.1 is equivalent to the following statement: for some sequences α_N, β_N and all $x > 0$

$$\lim_{N\to\infty} \frac{U_F(Nx) - U_F(N)}{\alpha_N} = \frac{x^\gamma - 1}{\gamma} \text{ (meaning } \log(x) \text{ for } \gamma = 0), \tag{A.1.7}$$

Since the left hand side is monotonic in x and the right hand side is continuous, convergence in (A.1.7) is locally uniform in x (that is, for any $0 < a, b < \infty$ convergence in (A.1.7) is uniform on $[a, b]$). See theorem 1.1.6 and corollary 1.2.4 in de Haan and Ferreira (2006).

We call the constants the α_N, β_N *canonical* normalization constants. They are related to the constants a_N, b_N of theorem 3.1 as follows.

1. For $\gamma < 0$, by lemma 1.2.9 in de Haan and Ferreira (2006)

$$\frac{a_N}{\alpha_N} \to -\frac{1}{\gamma}, \tag{A.1.8}$$

$$\frac{\beta_N - b_N}{a_N} = \frac{U_F(N) - \theta_F}{\theta_F - U_F(N)} = -1. \tag{A.1.9}$$

2. For $\gamma > 0$ by lemma 1.2.9 in de Haan and Ferreira (2006)

$$\frac{a_N}{\alpha_N} \to \frac{1}{\gamma}, \tag{A.1.10}$$

$$\frac{\beta_N - b_N}{a_N} = \frac{U_F(N) - 0}{U_F(N)} = 1. \tag{A.1.11}$$

3. For $\gamma = 0$

$$a_N = \alpha_N, \tag{A.1.12}$$

$$b_N = \beta_N.$$

We now give a proof of proposition 3.3.

Proof of proposition 3.3. Let a_N be the normalizing constants of theorem 3.1 for the distribution F. Fix $\tau \in (0, \infty)$ and define $s_{\tau,N,T}$ and $S_{\tau,N,T}$ as

$$S_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1 + 1 - \frac{1}{N\tau} - u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right),$$

$$s_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1 - \frac{1}{N\tau} - u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right)\right).$$

Condition (TE-Inf) for F and G_T can be written as $\inf_{u \in [1-1/N\tau, 1]} S_{\tau,N,T}(u) \to 0$, while condition (TE-Sup) can be written as $\sup_{u \in [0, 1-1/N\tau]} s_{\tau,N,T}(u) \to 0$.

Observe that by lemma A.1.3

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

Suppose that a_N is eventually positive (if not, simply reverse all inequalities below; also observe that a_N is monotonic). The above inequality implies that $\sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \leq \inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u)$. If $u_{s,\tau,N,T} \in [0, 1-1/N\tau]$ and $u_{S,\tau,N,T} \in [1-1/N\tau, 1]$, then the following chain of inequalities holds:

$$s_{\tau,N,T}(u_{s,\tau,N,T}) \leq \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) \leq \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u) \leq S_{\tau,N,T}(u_{S,\tau,N,T}) \ . \qquad \text{(A.1.13)}$$

Let F satisfy assumption 2 with EV index $\gamma > \gamma'$. Define

$$u_{S,\tau,N,T} = 1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1} \in \left[1-\frac{1}{N\tau}, 1\right].$$

Suppose that under the conditions of the proposition it holds that

$$\frac{1}{a_N}\left(F^{-1}\left(u_{S,\tau,N,T}\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right) \to 0, \qquad \text{(A.1.14)}$$

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}\left(1+1-\frac{1}{N\tau}-u_{S,\tau,N,T}\right) \to 0. \qquad \text{(A.1.15)}$$

Under (A.1.14) and (A.1.15) it holds that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$ as $N, T \to \infty$. By inequality (A.1.13) we conclude that $\limsup_{N,T\to\infty} \inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u) \leq 0$. An identical argument shows that $s_{\tau,N,T}(u_{s,\tau,N,T}) \to 0$ where $u_{s,\tau,N,T} = 1-(1/N\tau)((\log(T)+1)/\log(T)) \in [0, 1-1/N\tau]$; this implies that $\liminf_{N,T\to\infty} \sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \geq 0$. Inequality (A.1.13) also implies that $\limsup_{N,T\to\infty} \sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \leq \liminf_{N,T\to\infty} \inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u)$. Combining the results, we obtain that both $\inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u)$ and $\sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u)$ tend to 0, proving the result.

It remains to establish eqs. (A.1.14) and (A.1.15). First consider (A.1.14). We split the proof by sign of γ.

Suppose $\gamma > 0$. By theorem 3.1, $a_N = F^{-1}(1-1/N) = U_F(N)$, where the second equality follows by eq. (A.1.2). By corollary 1.2.10 in de Haan and Ferreira (2006) $(U_F(x))^{-1} \in RV_{-\gamma}$.

Then by lemma A.1.4 we can write $1/U_F(x) = x^{-\gamma}L(x)$ where L is a slowly varying function (that depends on F). In particular, $1/a_N = 1/U_F(N) = N^{-\gamma}L(N)$. Then at our choice of $u_{S,\tau,N,T}$, again using eq. (A.1.2), it holds that

$$\begin{aligned}
&\frac{1}{a_N}\left(F^{-1}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)-F^{-1}\left(1-\frac{1}{N\tau}\right)\right)\\
&=\frac{1}{a_N}\left(U_F\left(N\tau\frac{\log(T)+1}{\log(T)}\right)-U_F(N\tau)\right)\\
&=N^{-\gamma}L(N)\left(\left(N\tau\frac{\log(T)+1}{\log(T)}\right)^{\gamma}\frac{1}{L\left(N\tau\frac{\log(T)+1}{\log(T)}\right)}-(N\tau)^{\gamma}\frac{1}{L(N\tau)}\right)\\
&\propto\left(\frac{\log(T)+1}{\log(T)}\right)^{\gamma}\frac{L(N)}{L\left(N\tau\frac{\log(T)+1}{\log(T)}\right)}-\frac{L(N)}{L(N\tau)}\\
&\to 0.
\end{aligned}$$

Convergence follows since L is slowly varying on infinity: $L(N)/L(N\tau) \to 1$. By local uniform convergence (proposition 0.5 in Resnick (1987)) $L(N)/L\left(N\tau(\log(T)+1)/\log(T)\right) \to 1$.

Suppose $\gamma < 0$, which implies that $\theta_F = U(\infty) < \infty$ and that $a_N = \theta_F - F^{-1}(1-1/N) = U_F(\infty) - U_F(N)$. By corollary 1.2.10 in de Haan and Ferreira (2006), $(U_F(\infty) - U(x))^{-1} \in RV_{-\gamma}$ and we can write $1/(U_F(\infty) - U(x)) = x^{-\gamma}L(x)$ for some slowly varying function L by lemma A.1.4. In particular, $1/a_N = 1/(U_F(\infty) - U(N)) = N^{-\gamma}L(N)$. The F terms satisfy

$$\begin{aligned}
&\frac{1}{a_N}\left(F^{-1}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)-F^{-1}\left(1-\frac{1}{N\tau}\right)\right)\\
&=\frac{1}{a_N}\left(U_F\left(N\tau\frac{\log(T)+1}{\log(T)}\right)-U_F(N\tau)\right)\\
&=\frac{1}{U_F(\infty)-U_F(N)}\left(\left(U_F\left(N\tau\frac{\log(T)+1}{\log(T)}\right)-U_F(\infty)\right)-(U_F(N\tau)-U(\infty))\right)\\
&=N^{-\gamma}L(N)\left(\left(N\tau\frac{\log(T)+1}{\log(T)}\right)^{\gamma}\frac{1}{L\left(N\tau\frac{\log(T)+1}{\log(T)}\right)}-(N\tau)^{\gamma}\frac{1}{L(N\tau)}\right)\\
&\propto\left(\frac{\log(T)+1}{\log(T)}\right)^{\gamma}\frac{L(N)}{L\left(N\tau\frac{\log(T)+1}{\log(T)}\right)}-\frac{L(N)}{L(N\tau)}\\
&\to 0,
\end{aligned}$$

where convergence follows as above.

Suppose $\gamma = 0$. By eq. (A.1.7) $\lim_{N\to\infty} (U_F(Nx) - U_F(N))/\alpha_N = \log(x)$ locally uniformly in x, where α_N are the canonical constants (as defined on p. 45). In addition, by eq. (A.1.12), in this case $\alpha_N = a_N$. Then

$$\begin{aligned}
&\frac{1}{a_N}\left(F^{-1}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right)\\
&= \frac{1}{\alpha_N}\left(U_F\left(N\tau\frac{\log(T)+1}{\log(T)}\right) - U_F(N\tau)\right)\\
&\to \log\left(\lim_{T\to\infty}\frac{\log(T)+1}{\log(T)}\right) = 0.
\end{aligned}$$

Now we focus on the G_T^{-1} term in eq. (A.1.15). First suppose that $\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$. Then by Markov's inequality we obtain for any $\tau \in (0,1)$ that $G_T^{-1}(\tau) \leq \left(\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta/(1-\tau)\right)^{1/\beta}$. Hence

$$\frac{1}{T^p}G_T^{-1}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) = O\left(\frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^p}\right) \tag{A.1.16}$$

where the O term is uniform in T.

First suppose that $\gamma \neq 0$. As above, $a_N^{-1} = x^{-\gamma}L(x)$ where L is a slowly varying function (that depends on F). For eq. (A.1.15) to hold it is sufficient that $N^{1/\beta-\gamma}L(x)(\log(T))^{1/\beta}/T^p \to 0$ (by eq. (A.1.16)). Write $\gamma = \gamma' + \delta$, $\delta > 0$. Then

$$\frac{N^{1/\beta-\gamma}L(x)(\log(T))^{1/\beta}}{T^p} = \left[\frac{N^{1/\beta-\gamma'}(\log(T))^{1/\beta}}{T^p}\right]\left[\frac{L(x)}{N^\delta}\right] \to 0$$

since the condition holds for γ' and $L(x)/x^\delta \to 0$ for any $\delta > 0$ (L is slowly varying).

Let $\gamma = 0$. Then (A.1.15) holds if $N^{1/\beta}\hat{f}(U_F(N))(\log(T))^{1/\beta}/T^p \to 0$. By lemma 1.2.9 in de Haan and Ferreira (2006) $\hat{f}(U_F(N))/U_F(N) \to 0$ and $U_F(N) \in RV_0$. Thus

$$\begin{aligned}
\frac{N^{1/\beta}\hat{f}(U_F(N))(\log(T))^{1/\beta}}{T^p} &= \frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^p}\frac{\hat{f}(U_F(N))}{U_F(N)}U_F(N)\frac{N^{-\gamma'}}{N^{-\gamma'}}\\
&= \frac{N^{-1/\beta-\gamma'}(\log(T))^{1/\beta}}{T^p}\frac{\hat{f}(U_F(N))}{U_F(N)}\frac{U_F(N)}{N^{-\gamma'}} \to 0.
\end{aligned}$$

The first fraction tends to zero by the assumption of the claim, and last one because $U_F(N)$ is slowly varying and $-\gamma' > 0$ (since by assumption $\gamma' < \gamma = 0$).

The proof is identical if $G_T \sim N(\mu_T, \sigma_T^2)$. Assumption 3 implies that μ_T and σ_T^2 are bounded. In this case

$$\frac{1}{T^p} G_T^{-1}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) = O\left(\frac{\sqrt{\log(N)}}{T^p}\right),$$

where the O term is uniform in T. The rest of the argument proceeds as above. $\square$

A.1.3 Proof of Theorem 3.4

Proof of theorem 3.4. By theorem 2.2.1 in de Haan and Ferreira (2006)

$$\sqrt{k}\frac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} \Rightarrow N(0,1). \tag{A.1.17}$$

where $(N/k)U_F'(N/k) = (N/k) \times \left((1/(1-F))^{-1}\right)'(N/k) \equiv c_N$ (by eqs. (A.1.2) and (A.1.3)). We now transfer this convergence property to $\vartheta_{N-k,N}$.

It is convenient for the purposes of the proof to replace the uniform random variables U_i with $1/U_i$. Let $Y_1, \dots, Y_N$ be iid random variables with cdf $F_Y(y) = 1 - 1/y, y > 1$, $F_Y(y) = 0$ for $y \le 1$. Observe that $Y_i \stackrel{d}{=} 1/U_i$ where U_i is Uniform[0, 1]. Let $Y_{1,N} \le \cdots \le Y_{N,N}$ be the order statistics, then $U_{k,N} \stackrel{d}{=} 1/Y_{N-k,N}$. As pointed out by de Haan and Ferreira (2006) (p. 50)

$$\begin{aligned} \theta_{N-k,N} &\stackrel{d}{=} U_F(Y_{N-k,N}), \\ \vartheta_{N-k,N,T} &\stackrel{d}{=} U_T(Y_{N-k,N}). \end{aligned} \tag{A.1.18}$$

Let $c_N = (N/k)U_F'(N/k)$, as in the statement of the theorem. Then

$$\begin{aligned} &\sqrt{k}\frac{\vartheta_{N-k,N,T} - U_F\left(\frac{N}{k}\right)}{c_N} \\ &\stackrel{d}{=} \sqrt{k}\frac{U_T\left(Y_{N-k,N}\right) - U_F\left(\frac{N}{k}\right)}{c_N} \\ &= \sqrt{k}\frac{U_F\left(Y_{N-k,N}\right) - U_F\left(\frac{N}{k}\right)}{c_N} + \frac{\sqrt{k}}{c_N}\left(U_T\left(Y_{N-k,N}\right) - U_F\left(Y_{N-k,N}\right)\right) \end{aligned}$$

By eqs. (A.1.17) and (A.1.18) it follows that as $N \to \infty$

$$\sqrt{k}\frac{U_F\left(Y_{N-k,N}\right) - U_F\left(\frac{N}{k}\right)}{c_N} \stackrel{d}{=} \sqrt{k}\frac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{c_N} \Rightarrow N(0,1)$$

The conclusion of the theorem follows if

$$\frac{\sqrt{k}}{c_N}\left(U_T\left(Y_{N-k,N}\right)-U_F\left(Y_{N-k,N}\right)\right)\xrightarrow{p}0\ . \tag{A.1.19}$$

We establish eq. (A.1.19) by an argument similar to the one used in the proof of theorem 3.2. By lemma A.1.3 $U_T(Y_{N-k,N})=F^{-1}\left(1-1/Y_{N-k,N}\right)$ can be bounded as

$$\begin{aligned}
&\sup_{u\in\left[0,1-\frac{1}{Y_{N-k,N}}\right]} F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1-\frac{1}{Y_{N-k,N}}-u\right)\\
&\leq U_T(Y_{N-k,N})\\
&\leq \inf_{u\in\left[1-\frac{1}{Y_{N-k,N}},1\right]}\left(F^{-1}+\frac{1}{T^p}G_T^{-1}\left(1+1-\frac{1}{Y_{N-k,N}}-u\right)\right)
\end{aligned}$$

Subtract $U_F(Y_{N-k,n})=F^{-1}\left(1-1/Y_{N-k,N}\right)$ on all sides and multiply by $\sqrt{k}/c_N$ to obtain

$$\begin{aligned}
&\sup_{u\in\left[0,1-\frac{1}{Y_{N-k,N}}\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1-\frac{1}{Y_{N-k,N}}-u\right)-F^{-1}\left(1-\frac{1}{Y_{N-k,N}}\right)\right)\\
&\leq\frac{\sqrt{k}}{c_N}\left(U_T(Y_{N-k,N})-U_F(Y_{N-k,N})\right)\\
&\leq\inf_{u\in\left[1-\frac{1}{Y_{N-k,N}},1\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1+1-\frac{1}{Y_{N-k,N}}-u\right)-F^{-1}\left(1-\frac{1}{Y_{N-k,N}}\right)\right),
\end{aligned}$$

if c_N is non-negative; the opposite inequalities hold if c_N is negative. Since $1/Y_{N-k,N}\stackrel{d}{=}U_{k,N}$, we obtain

$$\begin{aligned}
&\inf_{u\in\left[1-\frac{1}{Y_{N-k,N}},1\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1+1-\frac{1}{Y_{N-k,N}}-u\right)-F^{-1}\left(1-\frac{1}{Y_{N-k,N}}\right)\right)\\
&\stackrel{d}{=}\inf_{u\in\left[1-U_{k,N},1\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1+1-U_{k,N}-u\right)-F^{-1}\left(1-U_{k,N}\right)\right)\\
&\xrightarrow{p}0
\end{aligned}$$

by condition (3). An identical argument applies to the supremum condition. Eq. (A.1.19) follows.

Sharpness of conditions (2) and (3) is established as in the proof of theorem 3.2. Suppose that condition (3) fails (the case for condition (2) is analogous). There is some subsequence of

(N,T) and some $\delta_{N,T}$ such that $\inf_{u\in\left[1-U_{k,N},1\right]}\sqrt{k}c_N^{-1}(\cdot)=\delta_{N,T}$ and $\delta_{N,T}$ is bounded away from zero. Suppose that it is possible to extract a further subsequence such that $\delta_{N,T}$ converges to some $\delta\neq 0$. By theorem 2 in Makarov (1981), there exists a joint distribution of θ_i and $\varepsilon_{i,T}$ such that the infimum is attained. Then along this subsequence for this joint distribution

$$\frac{\sqrt{k}}{c_N}\left(U_T\left(Y_{N-k,N}\right)-U_F\left(Y_{N-k,N}\right)\right)\xrightarrow{p}\delta,$$

from which we conclude that

$$\begin{aligned}
&\sqrt{k}\frac{\vartheta_{N-k,N,T}-U_F\left(\frac{N}{k}\right)}{c_N}\\
&=\sqrt{k}\frac{U_F\left(Y_{N-k,N}\right)-U_F\left(\frac{N}{k}\right)}{c_N}+\frac{\sqrt{k}}{c_N}\left(U_T\left(Y_{N-k,N}\right)-U_F\left(Y_{N-k,N}\right)\right)\\
&\Rightarrow N(\delta,1).
\end{aligned}$$

This convergence result may or may not hold for the overall original sequence.

If no convergent subsequence of $\delta_{N,T}$ can be extracted, $\delta_{N,T}$ is unbounded. Extract a further monotonically increasing subsequence. There exists a sequence of joint distributions of θ_i and $\varepsilon_{i,T}$ such that along that subsequence $\vartheta_{N-k,N,T}$ diverges:

$$\sqrt{k}\frac{\vartheta_{N-k,N,T}-U_F\left(\frac{N}{k}\right)}{c_N}\Rightarrow N(0,1)+\infty.$$

$\square$

A.1.4 Proof of Proposition 3.5

Proof of proposition 3.5. The proof proceeds similarly to that of proposition 3.3. As in the proof of proposition 3.3, it is sufficient to prove that for some $\tilde{u}_{S,N,T}\in[1-U_{k,N},1]$ and $\tilde{u}_{s,N,T}\in[0,1-U_{k,N}]$

$$\frac{\sqrt{k}}{c_N}\left(F^{-1}(\tilde{u}_{S,N,T})+\frac{1}{T^p}G_T^{-1}\left(1+1-U_{k,N}-\tilde{u}_{S,N,T}\right)-F^{-1}\left(1-U_{k,N}\right)\right)\xrightarrow{p}0,\tag{A.1.20}$$

$$\frac{\sqrt{k}}{c_N}\left(F^{-1}(\tilde{u}_{s,N,T})+\frac{1}{T^p}G_T^{-1}\left(1-U_{k,N}-\tilde{u}_{s,N,T}\right)-F^{-1}\left(1-U_{k,N}\right)\right)\xrightarrow{p}0.\tag{A.1.21}$$

We only show that eq. (A.1.20) holds, eq. (A.1.21) follows analogously.

Let $\rho = \delta/2 + \nu$, and set

$$\tilde{u}_{S,N,T} = 1 - U_{k,N} \frac{N^\rho}{N^\rho + 1} \in [1 - U_{k,N}, 1].$$

As in the proof of proposition 3.3, we first show that the scaled F^{-1} terms in eq. (A.1.20) decay, and then that the scaled G_T^{-1} term decays.

First we establish that

$$\frac{\sqrt{k}}{c_N} \left(F^{-1} \left(1 - U_{k,N} \frac{N^\rho}{N^\rho + 1} \right) - F^{-1} \left(1 - U_{k,N} \right) \right) \to 0. \tag{A.1.22}$$

Let $Y_1, \ldots, Y_N$ be iid random variables with cdf $F_Y(y) = 1 - 1/y, y > 1$, $F_Y(y) = 0$ for $y \leq 1$. We will use that $Y_i \stackrel{d}{=} 1/U_i$, and correspondingly $Y_{N-k,N} \stackrel{d}{=} 1/U_{k,N}$, as in the proof of theorem 3.4.

Observe that c_N can be written as $c_N = (N/k)U_F'(N/k)$. Then using eq. (A.1.2) we obtain that

$$\begin{aligned}
&\frac{\sqrt{k}}{c_N} \left(F^{-1} \left(1 - U_{k,N} \frac{N^\rho}{N^\rho + 1} \right) - F^{-1} \left(1 - U_{k,N} \right) \right) \\
&\stackrel{d}{=} \frac{\sqrt{k}}{\frac{N}{k} U_F'\left(\frac{N}{k}\right)} \left(U_F \left(Y_{N-k,N} \frac{N^\rho + 1}{N^\rho} \right) - U_F(Y_{N-k,N}) \right) \\
&= \frac{\sqrt{k}}{\frac{N}{k} U_F'\left(\frac{N}{k}\right)} \left(U_F \left(\frac{N}{k} \left(\frac{k}{N} Y_{N-k,N} \right) \frac{N^\rho + 1}{N^\rho} \right) - U_F \left(\frac{N}{k} \left(\frac{k}{N} Y_{N-k,N} \right) \right) \right) \\
&= \sqrt{k} \left(\frac{N^\rho + 1}{N^\rho} - 1 \right) \left(\frac{k}{N} Y_{N-k,N} \right) \frac{\frac{N}{k} U_F'\left(\frac{N}{k} x_N\right)}{\frac{N}{k} U_F'\left(\frac{N}{k}\right)}, \quad x_N \in \left[\frac{k}{N} Y_{N-k,N}, \left(\frac{k}{N} Y_{N-k,N} \right) \frac{N^\rho + 1}{N^\rho} \right]
\end{aligned}$$

where the last line follows by the mean value theorem. We now deal with the last two terms in the above expression. By corollary 2.2.2 in de Haan and Ferreira (2006), $(k/N)Y_{N-k,N} \xrightarrow{p} 1$. By corollary 1.1.10 in de Haan and Ferreira (2006) under assumptions 1 and 4 it holds that

$$\lim_{t \to \infty} \frac{U_F'(tx)}{U_F'(t)} = x^{\gamma - 1}$$

locally uniformly in x. Since $x_N \to 1$ as $N \to \infty$ and $k \to \infty, k = o(N)$, we conclude that

$$\frac{U_F'\left(\frac{N}{k} x_N\right)}{U_F'\left(\frac{N}{k}\right)} \to 1.$$

Combining these observations, we obtain that

$$\begin{aligned}
&\frac{\sqrt{k}}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}\left(U_F\left(\frac{N}{k}\left(\frac{k}{N}Y_{N-k,N}\right)\frac{N^\rho+1}{N^\rho}\right)-U_F\left(\frac{N}{k}\left(\frac{k}{N}Y_{N-k,N}\right)\right)\right)\\
&=O_p\left(\sqrt{k}\left[\frac{N^\rho+1}{N^\rho}-1\right]\right)\\
&=O_p(N^{\delta/2}N^{-(\delta/2+\nu)})\\
&=O_p(N^{-\nu})=o_p(1).
\end{aligned}$$

where we apply the assumption that $k=N^\delta$ in the third line.

Now we show that

$$\frac{\sqrt{k}}{c_N}\frac{1}{T^p}G_T^{-1}\left(1+1-U_{k,N}-\tilde{u}_{S,N,T}\right)\xrightarrow{p}0. \tag{A.1.23}$$

Suppose that $\sup_T\mathbb{E}|\varepsilon_{i,T}|^\beta<\infty$. As in eq. (A.1.16), we obtain that

$$\frac{1}{T^p}G_T^{-1}\left(1-U_{k,N}\left(1-\frac{N^\rho}{N^\rho+1}\right)\right)\sim O\left(\frac{N^{\rho/\beta}}{U_{k,N}^{1/\beta}T^p}\right)$$

By corollary 1.1.10 in de Haan and Ferreira (2006) $U_F'\in RV_{\gamma-1}$, hence for some slowly varying function L we can write $U_F'(x)=x^{\gamma-1}L(x)$ (be lemma A.1.4). Then

$$c_N=\frac{N}{k}U_F'\left(\frac{N}{k}\right)=\left(\frac{N}{k}\right)^\gamma L\left(\frac{N}{k}\right).$$

Hence

$$\begin{aligned}
&\frac{\sqrt{k}}{c_N}\frac{1}{T^p}G_T^{1}\left(1+1-U_{k,N}-\tilde{u}_{S,N,T}\right)\\
&=O\left(\sqrt{k}\frac{1}{\left(\frac{N}{k}\right)^\gamma L\left(\frac{N}{k}\right)}\frac{N^{\rho/\beta}}{U_{k,N}^{1/\beta}T^p}\right)\\
&=O_p\left(k^{1/2+\gamma-1/\beta}N^{-\gamma+\rho/\beta+1/\beta}\frac{1}{L\left(\frac{N}{k}\right)}\frac{1}{T^p}\right)\\
&=O_p\left(\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(-\gamma+1/\beta)+\nu/\beta}}{T^p}\frac{1}{L\left(\frac{N}{k}\right)}\right)\\
&=O_p\left(\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(-\gamma'+1/\beta)+\nu/\beta}}{T^p}\frac{N^{-\varkappa(1-\delta)}}{L\left(N^{1-\delta}\right)}\right)\\
&=o_p(1),
\end{aligned}$$

where in the third line we again use corollary 2.2.2 in de Haan and Ferreira (2006) to conclude that $(N/k) \times 1/U_{k,N} \stackrel{d}{=} (k/N)\, Y_{N-k,N} \xrightarrow{p} 1$; in the fourth line we write $\gamma = \gamma' + \varkappa$ where $\varkappa > 0$ by assumption, and last line follows by assumptions of the proposition and since L is slowly varying.

Proof in the case $G_T \sim N(\mu_T, \sigma_T^2)$ is analogous.

Combining together equations (A.1.22) and (A.1.23) shows that (A.1.20) holds. To prove that (A.1.21) holds, take

$$\tilde{u}_{s,N,T} = 1 - U_{k,N} \frac{N^\rho + 1}{N^\rho} \in [0, 1 - U_{k,N}]$$

and proceed as above. $\square$

A.2 Inference

A.2.1 Proof of Lemmas 4.1- 4.2 and Theorem 4.3

Proof of lemma 4.1. We split the proof by sign of γ.

For $\gamma > 0$

$$\begin{aligned}
&\frac{1}{F^{-1}\left(1 - \frac{1}{N}\right)} \left[\vartheta_{N,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)\right] \\
&= \frac{1}{U_F(N)} \left[\vartheta_{N,N,T} - U_F\left(\frac{N}{l}\right)\right] \\
&= \frac{1}{U_F(N)} \vartheta_{N,N,T} - \frac{U_F(N \times l^{-1})}{U_F(N)} \\
&\Rightarrow Fr - (l^{-1})^\gamma = Fr - \frac{1}{l^\gamma},
\end{aligned}$$

since by corollary 1.2.10 in de Haan and Ferreira (2006) $U_F(x) \in RV_\gamma$ and $\vartheta_{N,N,T}/U_F(N) \Rightarrow Fr$ by theorem 3.1.

For $\gamma = 0$

$$\begin{aligned}
&\frac{1}{\hat{f}\left(F^{-1}\left(1 - \frac{1}{N}\right)\right)} \left(\vartheta_{N,N,T} - F^{-1}\left(1 - \frac{l}{N}\right)\right) \\
&= \frac{1}{\hat{f}\left(F^{-1}\left(1 - \frac{1}{N}\right)\right)} \left(\vartheta_{N,N,T} - F^{-1}\left(1 - \frac{1}{N}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{\hat{f}\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}\left(F^{-1}\left(1-\frac{l}{N}\right) - F^{-1}\left(1-\frac{1}{N}\right)\right)\\
\Rightarrow\, & Gu - \log(l),
\end{aligned}
$$

since

$$
\begin{aligned}
&\frac{1}{f\left(F^{-1}\left(1-\frac{1}{N}\right)\right)}\left(F^{-1}\left(1-\frac{l}{N}\right) - F^{-1}\left(1-\frac{1}{N}\right)\right)\\
&= \frac{U_F(Nl^{-1}) - U_F(N)}{\alpha_N}\\
&\to -\log(l)
\end{aligned}
$$

by eq. (A.1.7).

For $\gamma < 0$

$$
\begin{aligned}
&\frac{1}{\theta_F - F^{-1}\left(1-\frac{1}{N}\right)}\left(\vartheta_{N,N,T} - F^{-1}\left(1-\frac{l}{N}\right)\right)\\
=&\frac{1}{\theta_F - F^{-1}\left(1-\frac{1}{N}\right)}\left(\vartheta_{N,N,T} - U_F(N/l)\right)\\
=&\frac{1}{\theta_F - F^{-1}\left(1-\frac{1}{N}\right)}\left(\vartheta_{N,N,T} - \theta_F\right) + \frac{\theta_F - U_F(N/l)}{\theta_F - U_F(N)}\\
\Rightarrow& W + \frac{1}{l^{\gamma}},
\end{aligned}
$$

since by corollary 1.2.10 in de Haan and Ferreira (2006) $\theta_F - U_F(x) \in RV_\gamma$. □

Proof of lemma 4.2. In this proof we use the canonical normalization constants α_N and β_N (see beginning of section A.1.2 for terminology and relation to constants of theorem 3.1), and at the end we convert our results to constants a_N and b_N of theorems 3.1-3.2.

First we show that for the canonical constants α_N, β_N

$$
\begin{aligned}
&\left(\frac{\vartheta_{N,N,T} - \beta_N}{\alpha_N}, \frac{\vartheta_{N-1,N,T} - \beta_N}{\alpha_N}, \ldots, \frac{\vartheta_{N-q,N,T} - \beta_N}{\alpha_N}\right)\\
&\qquad \Rightarrow \left(\frac{(E_1^*)^{-\gamma} - 1}{\gamma}, \frac{(E_1^* + E_2^*)^{-\gamma} - 1}{\gamma}, \ldots, \frac{(E_1^* + E_2^* + \cdots + E_{q+1}^*)^{-\gamma} - 1}{\gamma}\right). \quad \text{(A.2.1)}
\end{aligned}
$$

Let $E_1, \ldots, E_{q+1}$ be iid standard exponential RVs, and $E_1^*, \ldots, E_{q+1}^*$ another iid set of standard exponential RVs. Observe that

$$
P\left(U_T\left(\frac{1}{1-\exp(-E_i)}\right) \le x\right) = H_T(x).
$$

Then

$$
\begin{aligned}
&(\vartheta_{N,N,T}, \vartheta_{N-1,N,T}, \dots, \vartheta_{N-q,N,T}) \\
&\stackrel{d}{=} \left(U_T\left(\frac{1}{1-\exp(-E_{1,n})}\right), U_T\left(\frac{1}{1-\exp\left(-E_{2,n}\right)}\right), \dots, U_T\left(\frac{1}{1-\exp(-E_{q+1,n})}\right)\right) \\
&\stackrel{d}{=} \left(U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}\right)}\right), U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}-\frac{E_2^*}{N-1}\right)}\right), \dots, \right. \\
&\qquad \left. \dots, U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}-\frac{E_2^*}{N-1}-\dots-\frac{E_{q+1}^*}{N-q}\right)}\right)\right),
\end{aligned}
$$

where the second equality follows by the Rényi (1953) representation of order statistics from an exponential sample (see expression (1.9) in Rényi (1953)).

Hence

$$
\begin{aligned}
&\left(\frac{\vartheta_{N,N,T}-\beta_N}{\alpha_N}, \dots, \frac{\vartheta_{N-q,N}-\beta_N}{\alpha_N}\right) \\
&\stackrel{d}{=} \left(\frac{U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}\right)}\right)-\beta_N}{\alpha_N}, \quad \dots, \quad \frac{U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}-\frac{E_2^*}{N-1}-\dots-\frac{E_{q+1}^*}{N-q}\right)}\right)-\beta_N}{\alpha_N}\right). \quad \text{(A.2.2)}
\end{aligned}
$$

Examine the first coordinate in the above vector:

$$
\begin{aligned}
&\frac{U_T\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}\right)}\right)-\beta_N}{\alpha_N} \\
&= \frac{U_T\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-\beta_N}{\alpha_N} \\
&= \frac{U_F\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-\beta_N}{\alpha_N} + \frac{U_T\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-U_F\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)}{\alpha_N}.
\end{aligned}
$$

We can rewrite each term in eq. (A.2.2) as above by decomposing it into a U_F a component and a difference term involving U_T and U_F

We separately analyze the two terms.

First, by theorem 2.1.1 in de Haan and Ferreira (2006)

$$\left(\frac{U_F\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}\right)}\right)-\beta_N}{\alpha_N},\ \ldots,\ \frac{U_F\left(\frac{1}{1-\exp\left(-\frac{E_1^*}{N}-\frac{E_2^*}{N-1}-\cdots-\frac{E_{q+1}^*}{N-q}\right)}\right)-\beta_N}{\alpha_N}\right)$$
$$\Rightarrow\left(\frac{(E_1^*)^{-\gamma}-1}{\gamma},\frac{(E_1^*+E_2^*)^{-\gamma}-1}{\gamma},\ldots,\frac{(E_1^*+E_2^*+\cdots+E_{q+1}^*)^{-\gamma}-1}{\gamma}\right). \quad \text{(A.2.3)}$$

Second, the difference terms converge to zero. We show this for the first term only, the result follows analogously for the other terms. First, write the difference term as

$$\frac{U_T\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-U_F\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)}{\alpha_N}$$
$$=\frac{U_T\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-U_T(N)}{\alpha_N}$$
$$-\frac{U_F\left(\frac{N}{N\left(1-\exp\left(-\frac{E_1^*}{N}\right)\right)}\right)-U_F(N)}{\alpha_N}$$
$$+\frac{U_F(N)-U_T(N)}{\alpha_N}. \quad \text{(A.2.4)}$$

Define

$$\tilde{h}_{N,T}(y)=\frac{U_T\left(Ny\right)-U_T(N)}{\alpha_N}.$$

$\tilde{h}_{N,T}$ converges pointwise to $\tilde{h}(y)=(y^\gamma-1)/\gamma$ as $N,T\to\infty$ by theorem 3.2 and eq. (A.1.7) with U_T in place of U_F. Since the limit is continuous, and $\tilde{h}_{N,T}(y)$ is monotonic in y, convergence is locally uniform in y (see section 0.1 in Resnick (1987)). Define $y_N=N(1-\exp(-E_1^*/N))$. Then $y_N\to E_1^*$ and y_N is a bounded sequence. Then as $N,T\to\infty$ it holds that $\tilde{h}_{N,T}(y_N^{-1})\to\tilde{h}((E_1^*)^{-1})$ (observe that E_1^* does not depend on N or T).

Similarly, define

$$\tilde{f}_N(y)=\frac{U_F\left(Ny\right)-U_F(N)}{\alpha_N}.$$

$\tilde{f}_N(y)$ converges to the same limit $\tilde{h}(y)$ by theorem 3.1 and eq. (A.1.7). As $\tilde{f}_N(y)$ is monotonic, convergence is also locally uniform in y, so analogously $\tilde{f}_N(y_N^{-1})\to\tilde{h}((E_1^*)^{-1})$. Thus the difference between the first two terms in eq. (A.2.4) tends to zero.

Last, $(U(N) - U_T(N))/\alpha_N = o(1)$ by eq. (A.1.6) in the proof of theorem 3.2 (take $\tau = 1$ and recall that conditions of theorem 3.2 are assumed to hold).

Overall, we conclude that that the expression in eq. (A.2.4) is $o(1)$. Combining the above argument with eq. (A.2.3), we obtain eq. (A.2.1).

Last, we translate from the canonical normalizing constants (α_N, β_N) to the constants (a_N, b_N) of theorem 3.2. We only show this for the first coordinate in eq. (A.2.1), the argument for the other coordinates is identical.

For $\gamma < 0$

$$\begin{aligned}\frac{\vartheta_{N,N,T} - \theta_F}{\theta_F - U(N)} &= \frac{\vartheta_{N,N,T} - \beta_N}{\theta_F - U(N)} + \frac{\beta_N - \theta_F}{\theta_F - U(N)}\\ &= \frac{\vartheta_{N,N,T} - \beta_N}{\alpha_N}\frac{\alpha_N}{\theta_F - U(N)} - 1\\ &\Rightarrow -\frac{(E_1^*)^{-\gamma} - 1}{\gamma} \times \gamma - 1 = -(E_1^*)^{-\gamma}.\end{aligned}$$

The second line follows by eq. (A.1.9). The third line follows from equation (A.1.8).

For $\gamma > 0$

$$\begin{aligned}\frac{\vartheta_{N,N,T}}{U_F(N)} &= \frac{\vartheta_{N,N,T} - \beta_N}{U_F(N)} + \frac{\beta_N}{U_F(N)}\\ &= \frac{\vartheta_{N,N,T} - \beta_N}{\alpha_N}\frac{\alpha_N}{U_F(N)} + 1\\ &\Rightarrow \frac{(E_1^*)^{-\gamma} - 1}{\gamma} \times \gamma + 1 = (E_1^*)^{-\gamma}\end{aligned}$$

where the second line follows from eq. (A.1.11) and the third line from eq. (A.1.10).

For $\gamma = 0$ the constants in the theorem statement are the canonical constants (see section A.1.2). We only need to represent $((E_1^*)^{-\gamma} - 1))/\gamma$ in the form given in the theorem statement. Observe that that for $x > 0$ $(x^{-\gamma} - 1)/\gamma \to -\log(x)$ as $\gamma \to 0$ (note the minus). This implies that

$$\frac{1}{\hat{f}\left(F^{-1}\left(1 - \frac{1}{N}\right)\right)}\left(\vartheta_{N,N} - F^{-1}\left(1 - \frac{1}{N}\right)\right) \Rightarrow -\log(E_1^*).$$

□

Proof of theorem 4.3. Follows immediately from lemmas 4.1 and 4.2 and the continuous mapping theorem. □

A.2.2 Proof of Theorem 4.4

We begin by computing the order of the difference between the noisy and the noiseless maximum. Define $EA_{N,T} = \max\{T^{-p}|\varepsilon_{1,T}|, \ldots, T^{-p}|\varepsilon_{N,T}|\}$. Since $\vartheta_{i,T} = \theta_i + T^{-p}\varepsilon_{i,T}$, the following elementary inequality holds:

$$\theta_{N,N} - EA_N \leq \vartheta_{N,N,T} \leq \theta_{N,N} + EA_N. \tag{A.2.5}$$

This can be restated simply as $|\vartheta_{N,N,T} - \theta_{N,N}| \leq EA_{N,T}$.

Lemma A.2.1. *Let* $\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$ *for some* $\beta > 0$*, then* $\vartheta_{N,N,T} - \theta_{N,N}$ *is* $O_p\left(N^{1/\beta}/T^p\right)$*.*

Proof. Let $t > 0$. We compare $EA_{N,T}$ to N^s/T^p for $s \geq 0$:

$$\begin{aligned}
P\left(\frac{EA_{N,T}}{N^s/T^p} \geq t\right) &= P\left(\bigcup_{i=1}^{N}\left\{\frac{1}{T^p}\frac{|\varepsilon_{i,T}|}{N^s/T^p} \geq t\right\}\right) \\
&\leq \sum_{i=1}^{N} P\left(\frac{|\varepsilon_{i,T}|}{N^s/T^p} \geq tT^p\right) \\
&\leq NP\left(|\varepsilon_{i,T}| \geq tN^s\right) \\
&\leq N\frac{\mathbb{E}|\varepsilon_{i,T}|^\beta}{t^\beta N^{\beta s}} \\
&\leq \frac{\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta}{t^\beta} N^{1-\beta s},
\end{aligned}$$

where we use Markov's inequality in the penultimate step. Setting $s = 1/\beta$ shows that these probabilities are bounded, uniformly in T, hence $EA_{N,T} = O_p\left(N^{1/\beta}/T^p\right)$. The result follows by inequality (A.2.5). □

Lemma A.2.2. *Let assumption 3 hold and let* $\varepsilon_{i,T} \sim N(\mu_T, \sigma_T^2)$ *for all* T*. Then* $\vartheta_{N,N,T} - \theta_{N,N} = O_p(\sqrt{\log(N)}/T^p)$.

Proof. By assumption 3, (μ_T, σ_T^2) is a bounded sequence. $EA_{N,T}$ is a maximum of independent normal variables of bounded mean and variance. Then $EA_{N,T} = O_p\left(\sqrt{\log(N)}/T^p\right)$. The result then follows by inequality (A.2.5). □

Proof of theorem 4.4. The proof changes depending on whether $l = 0$ or $l > 0$.

We begin with $l = 0$ and $\gamma < 0$. Label

$$J(x) = P\left(\frac{(E_1^*)^{-\gamma}}{(E_1^* + \dots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}} \le x\right),$$
$$J_{N,T}(x) = P\left(\frac{\vartheta_{N,N,T} - \theta_F}{\vartheta_{N-q,N,T} - \vartheta_{N,N,T}} \le x\right),$$

using notation of theorem 4.3. Theorem 4.3 shows that $J_N \Rightarrow J$.

Add and subtract θ_F in $L_{b,N,T}$ to obtain

$$L_{b,N,T}(x) = \frac{1}{\binom{N}{b}} \sum_{s=1}^{\binom{N}{b}} \mathbb{I}\left\{\frac{\vartheta_{b,b,T}^{(s)} - \theta_F}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} + \frac{\theta_F - \vartheta_{N,N,T}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} \le x\right\}.$$

Fix an arbitrary $\epsilon > 0$ and define the event

$$E_{N,T} = \left\{\left|\frac{\theta_F - \vartheta_{N,N,T}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}}\right| \le \epsilon\right\}.$$

The goal is to show that $P(E_{N,T}) \to 1$ for any $\varepsilon > 0$ as $N, T \to \infty$.

Write

$$\frac{\theta_F - \vartheta_{N,N,T}}{\vartheta_{b-k,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} = \frac{\theta_F - \theta_{N,N}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} + \frac{\theta_{N,N} - \vartheta_{N,N,T}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}}. \tag{A.2.6}$$

We show that both terms are $o_p(1)$, which allows us to conclude that $P(E_{N,T}) \to 1$.

Focus on the first term in eq. (A.2.6), which we can further expand as

$$\frac{\theta_F - \theta_{N,N}}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} = \frac{\theta_F - \theta_{N,N}}{\theta_F - U_F(N)} \frac{\theta_F - U_F(b)}{\vartheta_{b-q,b,T}^{(s)} - \vartheta_{b,b,T}^{(s)}} \frac{\theta_F - U_F(N)}{\theta_F - U_F(b)}.$$

1. The first term is $O_p(1)$ by the noiseless extreme value theorem 3.1, case of $\gamma < 0$ (recall that $a_N = \theta_F - U_F(N)$ by eq. (A.1.2)). This term does not depend on T.

2. Second term is $O_p(1)$ by lemma 4.2. Lemma 4.2 applies to subsample s, since (N, T) satisfy conditions of proposition 3.3, and $b = o(N)$, hence (b, T) satisfy conditions of proposition 3.3 as well.

3. Last, $(\theta_F - U_F(t)) \in RV_\gamma$ by corollary 1.2.10 in de Haan and Ferreira (2006). By proposition 0.5 in Resnick (1987) $(\theta_F - U_F(xt))/(\theta_F - U_F(t)) \to x^\gamma$ uniformly on intervals of the form (b, ∞). Hence, using $b = N^m$, $m < 1, \gamma < 0$, we obtain

$$\frac{\theta_F - U_F(N)}{\theta_F - U_F(N^m)} = \frac{\theta_F - U_F((N^{1-m}N^m))}{\theta_F - U_F(N^m)} \sim (N^{1-m})^\gamma \to 0.$$

The last term is $o(1)$.

Overall the first term in eq. (A.2.6) is $o_p(1)$

Now focus on the second term in eq. (A.2.6). First, suppose that condition (1) of proposition 3.3 holds. Since $\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$, by lemma A.2.1 we conclude that $\theta_{N,N} - \vartheta_{N,N} = O_p(N^{1/\beta}/T^p)$. By lemma 4.2, $(\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T})/(\theta_F - U_F(b)) = O_p(1)$. In addition, by corollary 1.2.10 in de Haan and Ferreira (2006) $1/(\theta_F - U_F(t))$ is $RV_{-\gamma}$, so by lemma A.1.4 we can write $1/(\theta_F - U_F(t)) = t^{-\gamma}L(t)$ for some slowly varying L. Since $b = N^m$, we obtain

$$\begin{aligned}
\frac{\theta_{N,N} - \vartheta_{N,N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} &= (\theta_{N,N} - \vartheta_{N,N,T})\frac{\theta_F - U_F(b)}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}}\frac{1}{\theta_F - U_F(b)} \\
&= O_p\left(\frac{N^{1/\beta}}{T^p}\right) O_p(1) O_p\left(\frac{1}{\theta_F - U_F(N^m)}\right) \\
&= O_p\left(\frac{N^{1/\beta}}{T^p}\right) O_p\left(N^{-\gamma m}L(N^m)\right). \qquad \text{(A.2.7)}
\end{aligned}$$

$L(N^m)$ diverges at rate slower than any power of N^m. Then the expression in (A.2.7) is $o_p(1)$ if for some fixed $\kappa > 0$

$$\frac{N^{1/\beta - \gamma m + \kappa}}{T^p} \to 0.$$

However, such a $\kappa > 0$ exists since assumptions of proposition 3.3 hold and $\gamma < \gamma m$.

If instead condition (2) of proposition 3.3 holds, we instead use lemma A.2.2 to obtain $\theta_{N,N} - \vartheta_{N,N,T} = O(\sqrt{\log(N)}/T^p)$. Then proceeding as above

$$\begin{aligned}
\frac{\theta_{N,N} - \vartheta_{N,N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} &= O_p\left(\frac{\sqrt{\log(N)}}{T^p}\right) O_p\left(\frac{1}{\theta_F - U(N^m)}\right) \\
&= O_p\left(\frac{\sqrt{\log(N)}}{T^p}\right) O_p\left(N^{-\gamma m}L(N^m)\right).
\end{aligned}$$

Since condition (2) of proposition 3.3 holds, the second term of (A.2.6) is $o_p(1)$ in this case as well.

The remainder of the proof now proceeds as in Politis and Romano (1994). Define

$$\tilde{L}_{b,N,T} = \frac{1}{\binom{N}{b}} \sum_{s=1}^{\binom{N}{b}} \mathbb{I}\left\{\frac{\vartheta^{(s)}_{b,b,T} - \theta_F}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} \le x\right\}.$$

On the event $E_{N,T}$ it holds that

$$\tilde{L}_{b,N,T}(x-\epsilon) \leq L_{b,N,T}(x) \leq \tilde{L}_{b,N,T}(x+\epsilon).$$

Since $P(E_{N,T}) \to 1$, the above also holds with probability approaching one. Observe that

$$\mathbb{E}(\tilde{L}_{n,b}(x)) = J_{b,T}(x) \Rightarrow J(x).$$

$\tilde{L}_{b,N,T}$ is U-statistic of order b with kernel bounded between 0 and 1. By theorem A on p. 201 in Serfling (1980) it holds that $\tilde{L}_{b,N,T}(x) - J_{b,T}(x) \xrightarrow{p} 0$. Then, as in Politis and Romano (1994), for any $\epsilon > 0$ with probability approaching it holds that

$$J(x-\epsilon) - \epsilon \leq L_{b,N,T} \leq J(x+\epsilon) + \epsilon.$$

Letting $\epsilon \to 0$ shows that $L_{b,N,T}(x) \to J(x)$ at all continuity points x of $J(x)$.

This also shows that $\hat{c}_\alpha = L^{-1}_{b,N,T}(\alpha) \to J^{-1}(\alpha) = c_\alpha$ since weak convergence of cdfs is equivalent to weak convergence of quantiles.

Now consider the case of $l > 0$. Add and subtract $U_F(b/l)$ in the subsampling estimator:

$$L_{b,N,T}(x) = \frac{1}{\binom{N}{b}} \sum_{s=1}^{\binom{N}{b}} \mathbb{I}\left\{ \frac{\vartheta^{(s)}_{b,b,T} - U_F(b/l)}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} + \frac{U_F(b/l) - \vartheta_{N-Nl/b,N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} \leq x \right\}.$$

First, since b satisfies the conditions of theorem 4.3

$$\frac{\vartheta^{(s)}_{b,b,T} - U_F(b/l)}{\vartheta^{(s)}_{b-q,b,F} - \vartheta^{(s)}_{b,b,F}} \Rightarrow \frac{(E_1^*)^{-\gamma} + \left(\frac{1}{l,}\right)^{\gamma}}{(E_1^* + \cdots + E_{q+1}^*)^{-\gamma} - (E_1^*)^{-\gamma}}.$$

Fix some $\epsilon > 0$ and define the event

$$E_{N,T} = \left\{ \left| \frac{U_F(b/l) - \vartheta_{N-Nl/b,N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} \right| \leq \epsilon \right\}.$$

The goal is is to show that $P(E_{N,T}) \to 1$ for any $\epsilon > 0$ under the assumptions of the theorem. We show this separately for different signs of γ.

Let $\gamma < 0$ and write the expression of interest as

$$\begin{aligned} &\frac{U_F(b/l) - \vartheta_{N-\frac{Nl}{b},N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} \qquad \text{(A.2.8)} \\ &= \sqrt{Nl/b}\frac{U_F(b/l) - \vartheta_{N-Nl/b,N,T}}{\frac{b}{l}U'_F(\frac{b}{l})} \frac{U_F(\infty) - U_F(b)}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} \frac{\frac{b}{l}U'_F\left(\frac{b}{l}\right)}{\sqrt{\frac{Nl}{b}}(U_F(\infty) - U_F(b))}. \end{aligned}$$

1. By theorem 3.4
$$\sqrt{\frac{Nl}{b}}\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N,T}}{\frac{b}{l}U_F'(\frac{b}{l})}=O_p(1).$$
Conditions of theorem 3.4 hold, since $k=Nl/b\sim N^{1-m}$ and conditions of proposition 3.5 are assumed to hold for $\delta=1-m$.

2. Since conditions of proposition 3.3 hold for N, they hold for b, and lemma 4.2 applies. We conclude that
$$\frac{U_F(\infty)-U_F(b)}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}=O_p(1)$$

3. Finally, by corollary 1.1.14 in de Haan and Ferreira (2006) under assumption 4 $(b/l)U_F'\left(\frac{b}{l}\right)/(U_F(\infty)-U_F(b))\to-\gamma$. Multiplying this by $(Nl/b)^{-1/2}\to 0$ shows that overall the last term is $o(1)$.

We conclude that overall the expression in eq. (A.2.8) is $o_p(1)$.

For $\gamma>0$ instead write
$$\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N}}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}=\sqrt{\frac{Nl}{b}}\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N,T}}{\frac{b}{l}U_F'(\frac{b}{l})}\frac{U_F(b)}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}\frac{\frac{b}{l}U_F'\left(\frac{b}{l}\right)}{\sqrt{\frac{Nl}{b}}U_F(b)}.$$
The last term is $o(1)$ by corollary 1.1.12 in de Haan and Ferreira (2006), the other terms remain as above.

For $\gamma=0$ write the term of interest as
$$\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N,T}}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}=\sqrt{\frac{Nl}{b}}\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N}}{\frac{b}{l}U_F'(\frac{b}{l})}\frac{\frac{b}{l}U_F'\left(\frac{b}{l}\right)}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}\frac{1}{\sqrt{\frac{Nl}{b}}}.$$

1. As above
$$\sqrt{\frac{Nl}{b}}\frac{U_F(b/l)-\vartheta_{N-\frac{Nl}{b},N}}{\frac{b}{l}U_F'(\frac{b}{l})}=O_p(1).$$

2. By corollary 1.1.10 in de Haan and Ferreira (2006) under assumption 4
$$\lim_{t\to\infty}\frac{U_F(tx)-U_F(t)}{tU_F'(t)}=\log(x),$$
that is, $NU_F'(N)$ are the canonical scaling constants α_N that appear in eq. (A.1.7). Hence by lemma 4.2
$$\frac{\frac{b}{l}U_F'\left(\frac{b}{l}\right)}{\vartheta^{(s)}_{b-q,b,T}-\vartheta^{(s)}_{b,b,T}}=O(1)$$

3. The last term converges to zero.

This shows that for all values of γ

$$\frac{U_F(b/l) - \vartheta_{N-\frac{Nl}{b},N,T}}{\vartheta^{(s)}_{b-q,b,T} - \vartheta^{(s)}_{b,b,T}} = o_p(1).$$

From this we conclude that for any $\epsilon > 0$ $P(E_N) \to 1$. The rest of the proof now proceeds as above. $\square$

A.2.3 Proof of Theorem 4.5

Proof of theorem 4.5. Observe that under assumptions of theorem 3.2 relationship (A.1.7) also holds for U_T:

$$\lim_{T\to\infty} \frac{U_T(Nx) - U_T(N)}{\alpha_N} = \frac{x^\gamma - 1}{\gamma}.$$

Since $(U_T(Nx) - U_T(N))/\alpha_N$ is monotonic in x and the l.h.s. is continuous, convergence is locally uniform (as in the proof of lemma 4.2). Correspondingly, locally uniformly for $0 < x, y < \infty$ (see, for example, lemma 2.2 in Dekkers and de Haan (1989))

$$\frac{U_T(Nx) - U_T(N)}{U_T(Ny) - U_T(N)} \to \frac{x^\gamma - 1}{y^\gamma - 1}.$$

The same condition is also satisfied by U_F under assumptions 1 and 2. Then locally uniformly in x, y

$$\frac{U_T(Nx) - U_T(N)}{U_T(Ny) - U_T(N)} - \frac{U_F(Nx) - U_F(N)}{U_F(Ny) - U_F(N)} \to 0. \tag{A.2.9}$$

Like in proof of theorem 3.4, let $Y_1, \ldots, Y_N$ be iid with cdf $1 - 1/y, y > 1$. Let $Y_{1,N} \leq \cdots \leq Y_{N,N}$ be the order statistics. As in proof of theorem 3.4

$$\theta_{N-k,N} \stackrel{d}{=} U_F(Y_{N-k,N}),$$

$$\vartheta_{N-k,N} \stackrel{d}{=} U_T(Y_{N-k,N}).$$

Now examine

$$\frac{\vartheta_{N-k,N} - \vartheta_{N-2k,N}}{\vartheta_{N-2k,N} - \vartheta_{N-4k,N}}$$

$$
\begin{aligned}
&\overset{d}{=} \frac{U_T\left(Y_{N-k,N}\right)-U_T\left(Y_{N-4k,N}\right)}{U_T\left(Y_{N-2k,N}\right)-U_T\left(Y_{N-4k,N}\right)}-1\\
&=\frac{U_F\left(Y_{N-k,N}\right)-U_F\left(Y_{N-2k,N}\right)}{U_F\left(Y_{N-2k,N}\right)-U_F\left(Y_{N-4k,N}\right)}\\
&\quad+\frac{U_T\left(Y_{N-k,N}\right)-U_T\left(Y_{N-4k,N}\right)}{U_T\left(Y_{N-2k,N}\right)-U_T\left(Y_{N-4k,N}\right)}\\
&\quad-\frac{U_F\left(Y_{N-k,N}\right)-U_F\left(Y_{N-4k,N}\right)}{U_F\left(Y_{N-2k,N}\right)-U_F\left(Y_{N-4k,N}\right)},
\end{aligned}
\tag{A.2.10}
$$

where in the third equality the (-1) is absorbed into the first term.

By theorem 2.1 in Dekkers and de Haan (1989)

$$
\frac{U_F\left(Y_{N-k,N}\right)-U_F\left(Y_{N-4k,N}\right)}{U_F\left(Y_{N-2k,N}\right)-U_F\left(Y_{N-2k,N}\right)} \overset{d}{=} \frac{\theta_{N-k,N}-\theta_{N-2k,N}}{\theta_{N-2k,N}-\theta_{N-4k,N}} \xrightarrow{p} 2^{\gamma}.
$$

We now show that the remainder term in eq. (A.2.10) is $o_p(1)$. We do so by showing that any subsequence contains a further subsequence that converges to 0 almost surely. First, write the remainder as

$$
\begin{aligned}
&\frac{U_T\left(Y_{N-k,N}\right)-U_T\left(Y_{N-4k,N}\right)}{U_T\left(Y_{N-2k,N}\right)-U_T\left(Y_{N-4k,N}\right)}-\frac{U_F\left(Y_{N-k,N}\right)-U_F\left(Y_{N-4k,N}\right)}{U_F\left(Y_{N-2k,N}\right)-U_F\left(Y_{N-4k,N}\right)}\\
=&\frac{U_T\left(\frac{Y_{N-k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_T\left(Y_{N-4k,N}\right)}{U_T\left(\frac{Y_{N-2k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_T\left(Y_{N-4k,N}\right)}-\frac{U_F\left(\frac{Y_{N-k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_F\left(Y_{N-4k,N}\right)}{U_F\left(\frac{Y_{N-2k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_F\left(Y_{N-4k,N}\right)}.
\end{aligned}
$$

Pass to an arbitrary subsequence T_j, N_j, k_j.

By corollary 2.2.2 in de Haan and Ferreira (2006) $(k/N)Y_{N-k,N} \xrightarrow{p} 1$. Hence

$$
\begin{aligned}
\frac{Y_{N-k,N}}{Y_{N-4k}} &= 4\frac{(k/N)Y_{N-k,N}}{(4k/N)Y_{N-4k}} \xrightarrow{p} 4,\\
\frac{Y_{N-2k,N}}{Y_{N-4k}} &= 2\frac{(2k/N)Y_{N-2k,N}}{(4k/N)Y_{N-4k}} \xrightarrow{p} 2.
\end{aligned}
$$

We can pass along to a subsubsequence N_{j_l} along which convergence above is a.s.

Fix some $\epsilon \in (0,1)$. Starting from some (random) N^* it holds that $Y_{N_{j_l}-k_{j_l},N_{j_l}}/Y_{N_{j_l}-4k_{j_l}} \in (4(1-\epsilon),4(1+\epsilon))$ and $Y_{N_{j_l}-2k_{j_l},N_{j_l}}/Y_{N_{j_l}-4k_{j_l}} \in (2(1-\epsilon),2(1+\epsilon))$. Then along this subsequence locally uniform convergence in (A.2.9) kicks in and so *along* N_{j_l}

$$
\frac{U_T\left(\frac{Y_{N-k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_T\left(Y_{N-4k,N}\right)}{U_T\left(\frac{Y_{N-2k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_T\left(Y_{N-4k,N}\right)}-\frac{U_F\left(\frac{Y_{N-k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_F\left(Y_{N-4k,N}\right)}{U_F\left(\frac{Y_{N-2k,N}}{Y_{N-4k}}Y_{N-4k,N}\right)-U_F\left(Y_{N-4k,N}\right)} \xrightarrow{a.s.} 0.
$$

Since the original subsequence was arbitrary, the remainder is $o_p(1)$ and the result now follows from the continuous mapping theorem. □

A.2.4 Proof of Theorem 4.6

We begin by establishing several supporting lemmas.

Lemma A.2.3. *Let $U_{1,N} \leq \cdots \leq U_{N,N}$ be the order statistics from an iid sample of size N from a Uniform$[0,1]$ distribution. If $k = o(N)$, $s = \lfloor \sqrt{k} \rfloor$, and $k \to \infty$, then*

$$\begin{pmatrix} \sqrt{k}\left(\frac{N}{k}U_{k+1,N} - 1\right) \\ \sqrt{k}\left(\frac{N}{k+s}U_{k+s+1,N} - 1\right) \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right).$$

Proof. By lemma 2.2.3 in de Haan and Ferreira (2006)

$$\sqrt{k}\left(\frac{N}{k}U_{k+1,N} - 1\right) \Rightarrow N(0,1) \equiv Z.$$

To show the result, we only need to show that the suitable scaled difference between $U_{k+1,N}$ and $U_{k+s-1,N}$ converges to zero in probability. Consider

$$\begin{aligned}
&\sqrt{k}\left(\frac{N}{k+s}U_{k+s+1,N} - 1\right) - \sqrt{k}\left(\frac{N}{k}U_{k+1,N} - 1\right) \\
&= \sqrt{k}N\left(\frac{k}{k}\frac{1}{k+s}U_{k+s+1,N} - \frac{1}{k}U_{k+1,N}\right) \\
&= \frac{N}{\sqrt{k}}\left(U_{k+s+1} - U_{k+1} - \frac{s}{N+1} - \frac{s}{k+s}U_{k+s+1,N} + \frac{s}{N+1}\right) \\
&= \frac{N}{\sqrt{k}}\left(U_{k+s+1} - U_{k+1} - \frac{s}{N+1}\right) \\
&\quad - \frac{N}{\sqrt{k}}\left(\frac{s}{k+s}U_{k+s+1,N} - \frac{s}{k+s}\frac{k+s+1}{N+1}\right) \\
&\quad + \frac{N}{\sqrt{k}}\left(\frac{s}{N+1} - \frac{s}{k+s}\frac{k+s+1}{N+1}\right) \\
&\equiv (A) + (B) + (C).
\end{aligned}$$

We show that each of the three terms $(A), (B), (C)$ is $o_p(1)$.

First consider the deterministic term (C):

$$\frac{N}{\sqrt{k}}\frac{s}{N+1}\left(1 - \frac{k+s+1}{k+s}\right) \sim \frac{s}{\sqrt{k}}\left(1 - \frac{k+s+1}{k+s}\right) \to 0 \text{ as } s \sim \sqrt{k}, k \to \infty.$$

Consider (A). A difference of order statistics from the uniform distribution follows a beta distribution: if $p > r$, then $U_{p,N} - U_{r,N} \sim \text{Beta}(p - r, N - p + r + 1)$. Let $\delta > 0$, then

$$P(|(A)| \geq \delta) = P\left(\left|\frac{N}{\sqrt{k}}\left(U_{k+s+1,N} - U_{k+1,N} - \frac{s}{N+1}\right)\right| \geq \delta\right)$$

$$
\begin{aligned}
&= P\left(\left|\text{Beta}(s, N-s+1) - \mathbb{E}\left(\text{Beta}(s, N-s+1)\right)\right| \geq \frac{\sqrt{k}}{N}\delta\right) \\
&\leq \frac{\text{Var}(\text{Beta}(s, N-s+1))}{\delta^2 k/N^2} \\
&= \frac{\frac{s(N-s+1)}{(N+1)^2(N+2)}}{\delta^2 \frac{k}{N^2}} \\
s = o(N) \qquad &\sim \frac{\frac{s}{N^2}}{\delta^2 \frac{k}{N^2}} \\
&= \frac{s}{\delta^2 k}
\end{aligned}
$$

which tends to zero as $s \sim \sqrt{k}$, showing that $(A) = o_p(1)$.

Last, turn to (B). Observe that $U_{k+s+1,N} \sim$Beta$(k+s+1, n-k-s)$. Let $\delta > 0$, then

$$
\begin{aligned}
&P((B) \geq \delta) = P\left(\frac{N}{\sqrt{k}}\left|\frac{s}{k+s}U_{k+s+1,N} - \frac{s}{k+s}\frac{k+s+1}{N+1}\right| \geq \delta\right) \\
&= P\left(\left|\text{Beta}(k+s+1, N-k-s) - \mathbb{E}\left(\text{Beta}(k+s+1, N-k-s)\right)\right| \geq \delta\frac{\sqrt{k}(k+s)}{Ns}\right) \\
&\leq \text{Var}(\text{Beta}(k+s+1, N-k-s))\frac{N^2 s^2}{\delta^2 k(k+s)^2} \\
&= \frac{(k+s+1)(N-k-s)}{(N+1)^2(N+2)}\frac{N^2 s^2}{\delta^2 k(k+s)^2} \\
&\sim \frac{(k+s+1)s^2}{k(k+s)^2} \sim \frac{1}{k+s} \to 0,
\end{aligned}
$$

which shows that $(B) = o_p(1)$. □

Lemma A.2.4. *Let θ be sampled iid from F, let assumption 4 hold. Let $k = o(N), k \to \infty, s = \lfloor\sqrt{k}\rfloor$. Let $\theta_{N-k,N}, \theta_{N-k-s,N}$ be the order statistics from F. Then as $N \to \infty$*

$$
\sqrt{k}\begin{pmatrix} \dfrac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} \\ \dfrac{\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right).
$$

Proof. We use the Cramer-Wold device together with a technique used by de Haan and Ferreira (2006) in proving an asymptotic normality result for a single statistic under a von Mises condition (see proof of theorem 2.2.1 therein).

Observe that (in notation of lemma A.2.3)

$$(\theta_{N-k,N},\theta_{N-k-s,N})\overset{d}{=}\left(U_F\left(\frac{1}{U_{k+1,N}}\right),U_F\left(\frac{1}{U_{k+s+1,N}}\right)\right).$$

Le $(c_1,c_2)\in\mathbb{R}^2$ and examine

$$\begin{aligned}
&c_1\sqrt{k}\frac{\theta_{N-k,N}-U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}+c_2\sqrt{k}\frac{\theta_{N-k-s,N}-U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)}\\
\overset{d}{=}&c_1\sqrt{k}\frac{U_F\left(\frac{N}{k}\frac{k}{NU_{k+1,N}}\right)-U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}+c_2\sqrt{k}\frac{U_F\left(\frac{N}{k+s}\frac{k+s}{NU_{k+s+1,N}}\right)-U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)}\\
=&c_1\sqrt{k}\int_1^{k/(NU_{k+1N})}\frac{U_F'\left(\frac{N}{k}t\right)}{U_F'\left(\frac{N}{k}\right)}dt+c_2\sqrt{k}\int_1^{(k+s)/(NU_{k+s+1,N})}\frac{U_F'\left(\frac{N}{k+s}t\right)}{U_F'\left(\frac{N}{k+s}\right)}dt.
\end{aligned}\tag{A.2.11}$$

Under assumption 4 $U_F'\in RV_{\gamma-1}$ by corollary 1.1.10 in de Haan and Ferreira (2006) (up to sign). Then by Potter's inequalities (proposition B.1.9 (5) in de Haan and Ferreira (2006)) for any $\varepsilon,\varepsilon'>0$ starting from some N_0 for $t\geq 1$ it holds that

$$(1-\varepsilon)t^{\gamma-1-\varepsilon'}<\frac{U_F'\left(\frac{N}{k}t\right)}{U_F'\left(\frac{N}{k}\right)}<(1+\varepsilon)t^{\gamma-1+\varepsilon'},$$

$$(1-\varepsilon)t^{\gamma-1-\varepsilon'}<\frac{U_F'\left(\frac{N}{k+s}t\right)}{U_F'\left(\frac{N}{k+s}\right)}<(1+\varepsilon)t^{\gamma-1+\varepsilon'}.$$

Multiply by $\sqrt{k}$ and take integrals with limits of integration as in in eq. (A.2.11):

$$(1-\varepsilon)\sqrt{k}\frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma-\varepsilon'}-1}{\gamma-\varepsilon'}\leq\sqrt{k}\frac{U_F\left(\frac{1}{U_{k+1,N}}\right)-U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}\leq(1+\varepsilon)\sqrt{k}\frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma+\varepsilon'}-1}{\gamma+\varepsilon'},$$

$$(1-\varepsilon)\sqrt{k}\frac{\left(\frac{k+s}{NU_{k+s+1,N}}\right)^{\gamma-\varepsilon'}-1}{\gamma-\varepsilon'}\leq\sqrt{k}\frac{U_F\left(\frac{1}{U_{k+s+1,N}}\right)-U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k+s}\right)}\leq(1+\varepsilon)\sqrt{k}\frac{\left(\frac{k+s}{NU_{k+s+1,N}}\right)^{\gamma+\varepsilon'}-1}{\gamma+\varepsilon'},$$

hence if $c_1,c_2\geq 0$

$$\begin{aligned}
&c_1(1-\varepsilon)\sqrt{k}\frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma-\varepsilon'}-1}{\gamma-\varepsilon'}+c_2(1-\varepsilon)\sqrt{k}\frac{\left(\frac{k+s}{NU_{k+s+1,N}}\right)^{\gamma-\varepsilon'}-1}{\gamma-\varepsilon'}\\
&\qquad\leq c_1\sqrt{k}\frac{U_F\left(\frac{1}{U_{k+1,N}}\right)-U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}+c_2\sqrt{k}\frac{U_F\left(\frac{1}{U_{k+s+1,N}}\right)-U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k+s}\right)}\\
&\qquad\qquad\leq c_1(1+\varepsilon)\sqrt{k}\frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma+\varepsilon'}-1}{\gamma+\varepsilon'}+c_2(1+\varepsilon)\sqrt{k}\frac{\left(\frac{k+s}{NU_{k+s+1,N}}\right)^{\gamma+\varepsilon'}-1}{\gamma+\varepsilon'}.
\end{aligned}$$

Similar inequalities apply for different combinations of signs of c_1, c_2, though with $(1-\varepsilon)$ replaced by $(1+\varepsilon)$ for the terms with $c_i < 0$.

Since by lemma A.2.3

$$\sqrt{k}\begin{pmatrix} \frac{N}{k}U_{k+1,N} - 1 \\ \frac{N}{k+s}U_{k+s+1,N} - 1 \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right),$$

then by the delta method we get that for any $\varepsilon' > 0$ that satisfies $\varepsilon' \neq \pm\gamma$ it holds that

$$\sqrt{k}\begin{pmatrix} \frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma+\varepsilon'} - 1}{\gamma+\varepsilon'} \\ \frac{\left(\frac{k}{NU_{k+s+1,N}}\right)^{\gamma+\varepsilon'} - 1}{\gamma+\varepsilon'} \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right), \quad \sqrt{k}\begin{pmatrix} \frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma-\varepsilon'} - 1}{\gamma-\varepsilon'} \\ \frac{\left(\frac{k}{NU_{k+s+1,N}}\right)^{\gamma-\varepsilon'} - 1}{\gamma-\varepsilon'} \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right).$$

Since $\varepsilon > 0$ is arbitrary, we obtain that

$$c_1\sqrt{k}\frac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k+s}\right)}$$

has the same asymptotic distribution as

$$c_1\sqrt{k}\frac{\left(\frac{k}{NU_{k+1,N}}\right)^{\gamma\pm\varepsilon'} - 1}{\gamma \pm \varepsilon'} + c_2\sqrt{k}\frac{\left(\frac{k+s}{NU_{k+s+1,N}}\right)^{\gamma\pm\varepsilon'} - 1}{\gamma \pm \varepsilon'}.$$

Finally, by the Cramer-Wold device,

$$\sqrt{k}\begin{pmatrix} \frac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} \\ \frac{\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k+s}\right)} \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right).$$

$\square$

Finally, we provide a proof of theorem 4.6.

Proof of theorem 4.6. First, we focus on part (1) and establish the result for θ.

Since F is differentiable, so is U_F, and

$$U_F'(t) = \frac{(1 - F(U_F(t)))^2}{f(U_F(t))}.$$

Then since $F(U_F(t))$ is monotonic, the monotonicity assumption on f implies that eventually also U_F' is non-increasing/non-decreasing.

Let $s = \lfloor \sqrt{k} \rfloor$. Recall $F^{-1}(1 - k/N) = U_F(N/k)$, then

$$\begin{aligned}
&\frac{\theta_{N-k,N} - F^{-1}\left(1 - \frac{k}{N}\right)}{\theta_{N-k,N} - \theta_{N-k-s,N}} \\
=&\frac{\theta_{N-k,N} - U_F(N/k)}{\theta_{N-k,N} - \theta_{N-k-s,N}} \\
=&\frac{\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - U_F(N/k)\right)}{\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - \theta_{N-k-s,N}\right)}.
\end{aligned}$$

By lemma A.2.4 the numerator weakly converges to $Z \equiv N(0,1)$

We show that the denominator converges to 1 in probability. Rewrite the denominator as

$$\begin{aligned}
&\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - \theta_{N-k-s,N}\right) \\
=&\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)\right) - \frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)\right) \\
&+ \frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(U_F\left(\frac{N}{k}\right) - U_F\left(\frac{N}{k+s}\right)\right).
\end{aligned}$$

Examine the second term.

$$\begin{aligned}
&\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)\right) \\
&= \frac{\sqrt{k}}{\frac{N}{k+s}U_F'(N/(k+s))}\left(\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)\right)\frac{\frac{N}{k+s}U_F'(N/(k+s))}{\frac{N}{k}U_F'(N/k)}.
\end{aligned}$$

By assumption 4 and corollary 1.1.10 in de Haan and Ferreira (2006) $U_F'(tx)/U_F'(t) \to x^{\gamma-1}$ as $t \to \infty$ locally uniformly in $(0, \infty)$. Hence

$$\frac{\frac{N}{k+s}U_F'(N/(k+s))}{\frac{N}{k}U_F'(N/k)} = \frac{k}{k+s}\frac{U_F'\left(\frac{N}{k}\frac{k}{k+s}\right)}{U_F'\left(\frac{N}{k}\right)} \to 1.$$

since $k/k + s \to 1$. Thus, by lemma A.2.4

$$\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)\right) - \frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)\right) \xrightarrow{p} 0.$$

Last, we examine the residual. Observe that $U_F' \geq 0$. First suppose U_F' is eventually non-increasing, in which case

$$\left(U_F\left(\frac{N}{k}\right) - U_F\left(\frac{N}{k+s}\right)\right) = \int_{(N/k)\times k/(k+s)}^{N/k} U_F'(t)\,dt \leq \frac{s}{k+s}\frac{N}{k}U_F'\left(\frac{N}{k}\frac{k}{k+s}\right).$$

Using the above expression, we obtain an upper bound for the residual term

$$\begin{aligned}
&\frac{\sqrt{k}}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}\left(U_F\left(\frac{N}{k}\right) - U_F\left(\frac{N}{k+s}\right)\right) \\
&\leq \frac{\sqrt{k}}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}\left(\frac{s}{k+s}\frac{N}{k}U_F'\left(\frac{N}{k}\frac{k}{k+s}\right)\right) \\
&= \frac{\sqrt{k}s}{k+s}\frac{U_F'\left(\frac{N}{k}\frac{k}{k+s}\right)}{U_F'\left(\frac{N}{k}\right)} \to 1
\end{aligned}$$

since $s \sim \sqrt{k}$ and by local uniform convergence of the ratio of U_F'.

At the same time, since U_F' is eventually non-increasing, we obtain a lower bound

$$\int_{(N/k)\times k/(k+s)}^{N/k} U_F'(t)\,dt \geq \frac{s}{k+s}\frac{N}{k}U_F'\left(\frac{N}{k}\right)$$

which shows that

$$\frac{\sqrt{k}}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)}\left(U\left(\frac{N}{k}\right) - U\left(\frac{N}{k+s}\right)\right) \geq \frac{\sqrt{k}s}{k+s} \to 1.$$

Hence, the residual term converges to 1. If instead U_F' is eventually non-decreasing, swap the $N/(k+s)$ and N/k terms.

Thus, overall we get that the denominator satisfies

$$\begin{aligned}
&\frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)\right) - \frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)\right) \\
&+ \frac{\sqrt{k}}{\frac{N}{k}U_F'(N/k)}\left(U\left(\frac{N}{k}\right) - U\left(\frac{N}{k+s}\right)\right) \\
&\xrightarrow{p} 1.
\end{aligned}$$

We conclude that

$$\frac{\theta_{N-k,N} - U_F(N/k)}{\theta_{N-k,N} - \theta_{N-k-s,N}} \Rightarrow \frac{Z}{1} = Z \sim N(0,1)$$

Now turn to part (2) and consider the noisy estimates ϑ_i.

If it holds that

$$\sqrt{k}\begin{pmatrix} \dfrac{\vartheta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} \\ \dfrac{\vartheta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right), \tag{A.2.12}$$

then the proof of part (1) applies with order statistics of ϑ replacing their noiseless counterparts θ. In light of this, it is sufficient to establish (A.2.12). We proceed similarly to proof of theorem 3.4 and we apply the Cramer-Wold device.

Observe that

$$(\vartheta_{N-k,N}, \vartheta_{N-k-s,N}) \stackrel{d}{=} (U_T(Y_{N-k,N}), U_T(Y_{N-k-s,N}))\,.$$

for Y as in the proof of theorem 3.4

Let $c_1, c_2 \in \mathbb{R}$. Then

$$\begin{aligned}
&c_1\sqrt{k}\frac{\vartheta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{\vartheta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \\
\stackrel{d}{=}&c_1\sqrt{k}\frac{U_T\left(Y_{N-k,N}\right) - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{U_T\left(Y_{N-k-s,N}\right) - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \\
=&c_1\sqrt{k}\frac{U_F\left(Y_{N-k,N}\right) - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{U_F\left(Y_{N-k-s,N}\right) - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \\
&+ c_1\sqrt{k}\frac{U_T\left(Y_{N-k,N}\right) - U_F\left(Y_{N-k,N}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{U_T\left(Y_{N-k-s,N}\right) - U_F\left(Y_{N-k-s,N}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \\
\stackrel{d}{=}&c_1\sqrt{k}\frac{\theta_{N-k,N} - U_F\left(\frac{N}{k}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{\theta_{N-k-s,N} - U_F\left(\frac{N}{k+s}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)} \\
&+ c_1\sqrt{k}\frac{U_T\left(Y_{N-k,N}\right) - U_F\left(Y_{N-k,N}\right)}{\frac{N}{k}U_F'\left(\frac{N}{k}\right)} + c_2\sqrt{k}\frac{U_T\left(Y_{N-k-s,N}\right) - U_F\left(Y_{N-k-s,N}\right)}{\frac{N}{k+s}U_F'\left(\frac{N}{k+s}\right)}.
\end{aligned}$$

The result now follows: the first two terms converge to the desired limit by lemma A.2.4; the third and the fourth term converge to zero i.p., this convergence follows as in the proof of theorem 3.4 applied at k and $k+s$. $\square$

A.3 Verification Of Examples 5-7

A.3.1 Description

In this section we collect all the expressions that we use in verifying the examples further on.

Distributions The three example cdfs are

$$\begin{aligned}
&\gamma < 0 \qquad && F_{W,\alpha}(\theta) = 1 - \left(\frac{\theta_F - \theta}{\theta_F}\right)^{\alpha}, \quad \alpha > 0, \theta \in [0, \theta_F], \theta_F < \infty, \\
&\gamma > 0 \qquad && F_{Fr,\kappa}(\theta) = 1 - (\theta + 1)^{-\kappa}, \quad \kappa > 0, \theta \in [0, \infty), \\
&\gamma = 0 \qquad && F_{Gu,\lambda}(\theta) = 1 - e^{-\lambda x}, \quad \theta \in [0, \infty).
\end{aligned}$$

Densities:

$$\begin{aligned}
f_{W,\alpha}(\theta) &= \frac{\alpha}{\theta_F}\left(\frac{\theta_F - \theta}{\theta_F}\right)^{\alpha-1}, \quad \theta \in [0, \theta_F] \ , \\
f_{Fr,\kappa}(\theta) &= \kappa(\theta + 1)^{-\kappa-1} \ , \\
f_{Gu,\lambda} &= \lambda e^{-\lambda x}.
\end{aligned}$$

Two specifications for cdf of estimation noise:

$$\begin{aligned}
G_{\beta,T}(x) &= \begin{cases} 1 - \frac{1}{2}(1 + (x - \mu_T))^{-\beta}, & x \geq \mu_T \\ \frac{1}{2}(1 - (x - \mu_T))^{-\beta} & x < \mu_T, \end{cases} \\
G_{Normal,T}(x) &= \Phi\left(\frac{x - \mu_T}{\sigma_T}\right) \ .
\end{aligned}$$

The distribution $G_{\beta,T}$ can be equivalently specified through its density

$$g_{\beta}(x) = \frac{\beta}{2}\left(1 + \operatorname{sgn}(x - \mu_T)(x - \mu_T)\right)^{-\beta-1} \ .$$

Inverses We will also use the following expressions for quantiles of the functions we consider and the corresponding expressions for the auxiliary function U_F:

$$F_{W,\alpha}^{-1}(y) = \theta_F - \theta_F(1 - y)^{1/\alpha},$$

$$
\begin{aligned}
U_{F_{W,\alpha}}(y) &= F^{-1}\left(1-\frac{1}{y}\right) = \theta_F - \theta_F y^{-1/\alpha}, \\
F^{-1}_{Fr,\kappa}(y) &= (1-y)^{-1/\kappa} - 1, \\
U_{F_{Fr,\kappa}}(y) &= y^{1/\kappa} - 1, \\
F^{-1}_{Gu,\lambda}(y) &= -\frac{\log(1-y)}{\lambda}, \\
U_{Gu,\lambda} &= \frac{\log y}{\lambda}, \\
G^{-1}_{\beta,T}(\tau) &= \begin{cases} (2(1-\tau))^{-1/\beta} - 1 + \mu_T & \tau \geq \frac{1}{2}, \\ 1-(2\tau)^{-1/\beta} + \mu_T & \tau < \frac{1}{2}. \end{cases}
\end{aligned}
$$

A.3.2 Extreme Value Theorem, Examples 5, 6, 7, Page 13

Here we check the details of verifying theorem 3.2 for examples 5-7.

Fix $\tau \in (0,\infty)$ and define $s_{\tau,N,T}$ and $S_{\tau,N,T}$ as

$$
\begin{aligned}
S_{\tau,N,T}(u) &= \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1+1-\frac{1}{N\tau}-u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right), \quad \text{(A.3.1)} \\
s_{\tau,N,T}(u) &= \frac{1}{a_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1-\frac{1}{N\tau}-u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right)\right).
\end{aligned}
$$

We follow the same strategy for all three examples. The approach is similar to that of the proof of proposition 3.3. First, we construct a sequence $u_{S,\tau,N,T} \in [1-1/N\tau, 1]$ such that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$ under certain conditions on N and T. Since $\inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u) \leq S_{\tau,N,T}(u_{S,\tau,N,T})$, we obtain that

$$
\limsup_{N,T\to\infty} \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u) \leq 0.
$$

Second, we construct a sequence $u_{s,\tau,N,T} \in [0, 1-1/N\tau]$ such that $s_{\tau,N,T}(u_{s,\tau,N,T}) \to 0$, which implies that

$$
\liminf_{N,T\to\infty} \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) \geq 0.
$$

Last, in all cases $a_N > 0$, hence by lemma A.1.3 it holds that $\sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \leq \inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u)$. This implies that

$$
\limsup_{N,T\to\infty} \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) \leq \liminf_{N,T\to\infty} \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u).
$$

Combining the three observations, and the trivial observation that $\liminf\{\dots\} \leq \limsup\{\dots\}$, we conclude that

$$\lim_{N,T\to\infty} \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) = \lim_{N,T\to\infty} \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u) = 0.$$

The above holds for any $\tau \in (0,\infty)$, and so the conditions (TE-Inf) and (TE-Sup) of theorem 3.2 hold.

Example 5, page 13 Let $\theta \sim F_{Fr,\kappa}$. Then $a_N = N^{1/\kappa} - 1$, $b_N = 0$.

We examine the infimum condition for a fixed $\tau \in (0\,\infty)$. Pick

$$u_{S,\tau,N,T} = 1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1} \in \left[1 - \frac{1}{N\tau}, 1\right] \ . \tag{A.3.2}$$

We show that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$, where $S_{\tau,N,T}$ is defined in eq. (A.3.1).

We will separately show that the G_T term and the pair of $F^{-1}_{Fr,\kappa}$ terms decay to zero.

Using expressions for quantiles given in section A.3.1, we obtain that

$$\begin{aligned}
F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\right) &= (N\tau)^{1/\kappa} - 1, \\
F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) &= \left(N\tau\frac{\log(T)+1}{\log(T)}\right)^{1/\kappa} - 1.
\end{aligned}$$

Then as $N, T \to \infty$

$$\begin{aligned}
&\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\right) - F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right] \\
&= \tau^{1/\kappa} - \tau^{1/\kappa}\left(\frac{\log(T)+1}{\log(T)}\right)^{1/\kappa} \\
&\to 0 \ .
\end{aligned}$$

This convergence holds regardless of relative values of N and T.

Suppose $G_T = G_{\beta,T}$. Using the expression for quantiles of $G_{\beta,T}$ given is section A.3.1, we obtain that

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1 + 1 - \frac{1}{N\tau} - u_{S,\tau,N,T}\right)$$

$$
\begin{aligned}
&= \frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right)\\
&\sim \frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p}+\frac{\mu_T}{N^{1/\kappa}T^p}\ .
\end{aligned}
$$

Since μ_T is a bounded sequence, we conclude that for $\limsup_{N,T\to\infty}\inf_{u\in[1-1/N\tau,1]} S_{\tau,N,T}(u)$ to be equal to zero, it is sufficient that

$$
\frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p}\to 0. \tag{A.3.3}
$$

Observe that this condition does not depend on the value of τ.

Now suppose $G_T = G_{Normal,T}$. Since σ_T is a bounded sequence, we can use the following simple approximation for quantiles of a normal random variable: $G^{-1}_{Normal,T}(1-c/N)\sim\sqrt{\log(N)}+\mu_T$. Then

$$
\frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{Normal,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right)\sim\frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p}+\frac{\mu_T}{N^{1/\kappa}T^p}\ .
$$

Since μ_T is bounded, for the G_T term to decay in this case, it is sufficient that

$$
\frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p}\to 0.
$$

Now we turn to $s_{\tau,N,T}$, associated with the supremum condition. Pick

$$
u_{s,\tau,N,T} = 1-\frac{1}{N\tau}\frac{\log(T)+1}{\log(T)}. \tag{A.3.4}
$$

With this choice

$$
\begin{aligned}
&\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\right)-F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\frac{\log(T)+1}{\log(T)}\right)\right]\\
&=\tau^{1/\kappa}-\tau^{1/\kappa}\left(\frac{\log(T)}{\log(T)+1}\right)^{1/\kappa}\\
&\to 0\ .
\end{aligned}
$$

Now turn to the G_T turns. If $G_T = G_{\beta,T}$ for all T, we obtain as above

$$
\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-\frac{1}{N\tau}-u_{s,\tau,N,T}\right)
$$

$$
\begin{aligned}
&= \frac{1}{N^{1/\kappa}-1}\frac{1}{T^p} G^{-1}_{\beta,T}\left(\frac{1}{N\tau}\left(\frac{\log(T)+1}{\log(T)}-1\right)\right) \\
&\sim \frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T(\log(T))^{1/\beta}}{N^{1/\kappa}T^p} \ .
\end{aligned}
$$

Exactly as above, we conclude that for $\liminf_{N,T\to\infty} \sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u)$ to be equal to zero, it is sufficient that

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

The condition is the same as for the infimum.

If $G_T(x) = G_{Normal,T}$, we obtain exactly as above that

$$
s\frac{1}{N^{1/\kappa}-1}\frac{1}{T^p} G^{-1}_{Normal,T}\left(\frac{1}{N\tau}\left(\frac{\log(T)+1}{\log(T)}-1\right)\right) \sim \frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p} \ ,
$$

which again matches the condition derived for the infimum.

Finally, since $a_N > 0$, lemma A.1.3 implies that

$$
\limsup_{N,T\to\infty} \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) \le \liminf_{N,T\to\infty} \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u) \ .
$$

If the above rate conditions on N and T hold, it holds that $\liminf_{N,T\to\infty} \sup_{u\in\left[0,1-\frac{1}{N\tau}\right]} s_{\tau,N,T}(u) = 0, \limsup_{N,T\to\infty} \inf_{u\in\left[1-\frac{1}{N\tau},1\right]} S_{\tau,N,T}(u) = 0$. We conclude that conditions TE-Inf and TE-Sup hold.

On the role of the log factor The $\log(T)$ factor in the definitions of $u_{s,\tau,N,T}$ and $u_{S,\tau,N,T}$ (eqs. (A.3.2) and (A.3.4)) can by replaced by any other function $h(N,T)$ of N and T that diverges to infinity as $N,T\to\infty$. This can soften the $(\log(T))^{1/\beta}$ term arbitrarily; for example, if we use an iterated log instead, the condition for the scaled $G^{-1}_{\beta,T}$ term to decay becomes instead

$$
\frac{N^{1/\beta}(\log(\cdots\log(T))^{1/\beta}}{N^{1/\kappa}T^{1/2}} \to 0 \ .
$$

At the same time, such a function $h(N,T)$ is necessary to eliminate the F^{-1} terms in the limit. Too see this, consider again the F^{-1} terms in $S_{\tau,N,T}$. Pick $u_{S,\tau,N,T} = 1-1/(cN\tau)$ where $c>1$ is fixed constant. Then

$$
\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\right) - F^{-1}_{Fr,\kappa}\left(1-\frac{1}{cN\tau}\right)\right] = \tau^{1/\kappa} - \tau^{1/\kappa}c^{1/\kappa} \neq 0,
$$

and the $F^{-1}_{Fr,\kappa}$ terms do not decay.

Example 6, page 13 If θ_i are exponential(λ), then $a_N^{-1} = 1$ (see 1.7.2. in Leadbetter et al. (1983)). Consider $S_{\tau,N,T}$ and pick $u_{S,\tau,N,T}$ as in eq. (A.3.2). Then since

$$F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\right) = \frac{\log(N\tau)}{\lambda} \ ,$$
$$F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) = \frac{\log(N\tau)+\log\left(\frac{\log(T)+1}{\log(T)}\right)}{\lambda} \ ,$$

we obtain that

$$\frac{1}{a_N}\left[F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\right) - F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right] = \frac{\log\left(\frac{\log(T)+1}{\log(T)}\right)}{\lambda} \to 0.$$

Suppose that $G_T = G_{\beta,T}$. Then

$$\frac{1}{a_N T^p} G^{-1}_{\beta,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T(\log(T))^{1/\beta}}{T^p} \ .$$

Since μ_T is bounded, for the above expression to decay it is sufficient that

$$\frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^{1/2}} \to 0.$$

Suppose that $G_T = G_{normal,T}$. Then

$$\frac{1}{a_N T^p} G^{-1}_{normal,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{\sqrt{\log(N)}}{T^p} + \frac{\mu_T}{T^p},$$

where the order equivalence holds because σ_T is a bounded sequence. If

$$\frac{\sqrt{\log(N)}}{T^p} \to 0$$

the above term decays to zero for any τ, since μ_T is bounded.

The results for $s_{\tau,N,T}$ follow the same pattern and yield the same conditions on N and T.

Example 7, page 13 If $\theta \sim F_{W,\alpha}$, then $a_N^{-1} = N^{1/\alpha}/\theta_F$ (see 1.7.10 in Leadbetter et al. (1983)). Consider $S_{\tau,N,T}$ and let $u_{S,\tau,N,T}$ be as in eq. (A.3.2).

First examine the $F^{-1}_{W,\alpha}$ terms. Using the expressions for inverses given in section A.3.1, we obtain

$$F^{-1}_{W,\alpha}\left(1-\frac{1}{N\tau}\right) = \theta_F - \theta_F\left(\frac{1}{N\tau}\right)^{1/\alpha},$$
$$F^{-1}_{W\alpha}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) = \theta_F - \theta_F\left(\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)^{1/\alpha},$$

hence

$$\begin{aligned}
&\frac{N^{1/\alpha}}{\theta_F}\left(F^{-1}_{W,\alpha}\left(1-\frac{1}{N\tau}\right) - F^{-1}_{W,\alpha}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right)\\
&\propto \frac{1}{\tau^{1/\alpha}} - \frac{1}{\tau^{1/\alpha}}\left(\frac{\log(T)}{\log(T)+1}\right)^{1/\alpha}\\
&\to 0.
\end{aligned}$$

Now turn to the G_T term. First suppose that $G_T = G_{T,\beta}$. Then

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\alpha+1/\beta}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T N^{1/\alpha}}{T^p}.$$

Since μ_T is a bounded sequence, for the above expression to decay it is sufficient that

$$\frac{N^{1/\alpha+1/\beta}(\log(T))^{1/\beta}}{T^{1/2}} \to 0.$$

Now suppose that $G_T = G_{normal,T}$. Then exactly as in the preceding examples we get

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{normal,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\alpha}\sqrt{\log(N)}}{T^p} + \frac{\mu_T N^{1/\alpha}}{T^p}.$$

For this term to decay it is sufficient that

$$\frac{N^{1/\alpha}\sqrt{\log(N)}}{T^p} \to 0.$$

The results for $s_{\tau,N,T}$ are obtained similarly and yield the same conditions on N and T.

A.3.3 Intermediate Order Statistics, Examples 5, 6, 7, Page 17

Example cdfs satisfy assumption 4 First we establish that cdfs F of examples 5-7 satisfy assumption 4, hence theorem 3.4 can be applied to the examples.

First consider $F_{Fr,\kappa}$:

$$\frac{1-F_{Fr,\kappa}}{f_{Fr,\kappa}}(\theta) = \frac{(\theta+1)^{-\kappa}}{\kappa(\theta+1)^{-\kappa-1}} = \frac{1}{\kappa}(\theta+1),$$

so

$$\left(\frac{1-F_{Fr,\kappa}}{f_{Fr,\kappa}}\right)' = \frac{1}{\kappa} = \gamma.$$

Second, examine $F_{Gu,\lambda}$:

$$\frac{1-F_{Gu,\lambda}}{f_{Gu,\lambda}} = \frac{e^{-\lambda x}}{\lambda e^{-\lambda x}} = \frac{1}{\lambda},$$

so

$$\left(\frac{1-F_{Gu,\lambda}}{f_{Gu,\lambda}}\right)' = 0 = \gamma.$$

Last, turn to $F_{W,\alpha}$:

$$\frac{1-F_{W,\alpha}}{f_{W,\alpha}}(\theta) = \frac{\left(\frac{\theta_F-\theta}{\theta_F}\right)^{\alpha}}{\frac{\alpha}{\theta_F}\left(\frac{\theta_F-\theta}{\theta_F}\right)^{\alpha-1}} = \frac{1}{\alpha}(\theta_F-\theta),$$

from which it follow that

$$\left(\frac{1-F_{W,\alpha}}{f_{W,\alpha}}\right)' = -\frac{1}{\alpha} = \gamma$$

Approach to obtaining rate conditions We will convert the tail equivalence conditions (2) and (3) into rate restrictions on N and T, along with conditions on choice of k. The overall approach is the same as in section A.3.2. Define $\tilde{s}_{N,T}$ and $\tilde{S}_{N,T}$ similarly to section A.3.2:

$$\begin{aligned}
\tilde{S}_{N,T}(u) &= \frac{\sqrt{k}}{c_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1+1-U_{k,N}-u\right) - F^{-1}\left(1-U_{k,N}\right)\right), \\
\tilde{s}_{N,T}(u) &= \frac{\sqrt{k}}{c_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1-U_{k,N}-u\right) - F^{-1}\left(1-U_{k,N}\right)\right),
\end{aligned}$$

where c_N is as in theorem 3.4.

We construct a sequence $\tilde{u}_{S,N,T} \in [1-U_{k,N}, 1]$ such that $\tilde{S}_{N,T}(\tilde{u}_{S,N,T}) \to 0$ under certain conditions on N, T and k. As previously, since $\inf_{u\in\left[1-U_{k,N},1\right]} \tilde{S}_{N,T}(u) \leq \tilde{S}_{N,T}(\tilde{u}_{S,N,T})$, we obtain that $\limsup_{N,T\to\infty} \inf_{u\in\left[1-U_{k,N},1\right]} \tilde{S}_{N,T}(u) \leq 0$. The same argument can be applied to $s_{N,T}$ to conclude that $\liminf_{N,T\to\infty} \sup_{u\in[0,1-U_{k,N}]} \tilde{s}_{N,T}(u) \geq 0$. Proceeding as section A.3.2, we conclude that conditions (2) and (3) of theorem 3.4 hold.

First, we establish the following elementary lemma.

Lemma A.3.1. *Let $\gamma \in \mathbb{R}, \rho \geq 0$. $((N^\rho + 1)/N^\rho)^\gamma - 1 = O(N^{-\rho})$ as $N \to \infty$.*

Proof. If $\gamma = 0$, the result is immediate. Suppose that $\gamma \neq 0$. Observe that $((N^\rho + 1)/N^\rho)^\gamma - 1 = f(1/N)$ for $f(x) = (1 + x^\rho)^\gamma - 1$. Observe that $f(0) = 0$. Then by the mean value theorem

$$\left(1 + \frac{1}{N^\rho}\right)^\gamma - 1 = f\left(\frac{1}{N}\right) - 1 = \frac{1}{N} f'\left(\frac{\varkappa}{N}\right), \varkappa \in [0,1].$$

Derivative of f is given by $f'(x) = \rho\gamma\,(1 + x^\rho)^{\gamma-1}\, x^{\rho-1}$, and so

$$\left|\frac{1}{N} f'\left(\frac{\varkappa}{N}\right)\right| = O\left(\frac{1}{N^\rho}\left|\left(1 + \frac{\varkappa^\rho}{N^\rho}\right)^{\gamma-1}\right|\right) = O\left(N^{-\rho}\right).$$

□

Example 5, page 17 Let $F = F_{Fr,\kappa}$. Compute the normalizing functions of theorem 3.4.

$$F^{-1}_{Fr,\kappa}\left(1 - \frac{k}{N}\right) = U_{F_{Fr,\kappa}}\left(\frac{N}{k}\right) = \left(\frac{N}{k}\right)^{1/\kappa} - 1,$$

$$c_N = \frac{N}{k}\left(\left(\frac{1}{1 - F_{Fr,\kappa}}\right)^{-1}\right)'\left(\frac{N}{k}\right) = \frac{N}{k} U'_{F_{Fr,\kappa}}\left(\frac{N}{k}\right) = \frac{1}{\kappa}\left(\frac{N}{k}\right)^{1/\kappa}$$

Examine the infimum condition (3). Define for $\rho > 0$ (below we discuss how k influences choice of $\tilde{u}_{S,N,T}$, including ρ):

$$\tilde{u}_{S,N,T} = 1 - U_{k,N}\frac{N^\rho}{N^\rho + 1} \in [1 - U_{k,N}, 1]\,. \tag{A.3.5}$$

We show that $\tilde{S}_{N,T}(\tilde{u}_{S,N,T}) \to 0$ by separately showing that both F^{-1} terms decay and the G_T^{-1} term decays, exactly as in section A.3.2.

The F terms satisfy (we suppress the $1/\kappa$ multiplicative term of c_N):

$$\begin{aligned}
&\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\left[F^{-1}_{Fr,\kappa}\left(1 - U_{k,N}\frac{N^\rho}{N^\rho+1}\right) - F^{-1}_{Fr,\kappa}\left(1 - U_{k,N}\right)\right] \\
=&\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\left[\left(\frac{1}{U_{k,N}}\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa} - \left(\frac{1}{U_{k,N}}\right)^{1/\kappa}\right] \\
=&\left(\frac{N}{k}U_{k,N}\right)^{-1/\kappa}\sqrt{k}\left[\left(\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa} - 1\right].
\end{aligned}$$

By corollary 2.2.2 in de Haan and Ferreira (2006) $(N/k)U_{k,N} \xrightarrow{p} 1$.[15]. Then the $F^{-1}_{Fr,\kappa}$ terms decay if k is chosen such that

$$\sqrt{k}\left[\left(\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa} - 1\right] \to 0. \tag{A.3.6}$$

Also observe that the $((N^\rho+1)/N^\rho)^{1/\kappa} - 1 = O(N^{-\rho})$ by lemma A.3.1. Thus, k must satisfy $k = o(N^{-\rho})$. See below for choice of ρ and k.

Now we turn to the G_T terms. First suppose that $G_T = G_{\beta,T}$. In this case

$$\begin{aligned}
&\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1 - U_{k,N}\left(1 - \frac{N^\rho}{N^\rho+1}\right)\right)\\
&\sim \frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{N^{\rho/\beta}}{U_{k,N}^{1/\beta}T^p}\\
&= \left(\frac{N}{k}U_{k,N}\right)^{-1/\beta}\frac{k^{1/2+1/\kappa-1/\beta}N^{\rho/\beta}}{N^{1/\kappa-1/\beta}T^p}.
\end{aligned}$$

Since $(N/k)U_{k,N} \xrightarrow{p} 1$, the above expression decays if

$$\frac{k^{1/2+1/\kappa-1/\beta}}{N^{1/\kappa-1/\beta-\rho/\beta}T^p} \to 0. \tag{A.3.7}$$

If $G_T = G_{normal,T}$, then we obtain that

$$\begin{aligned}
&\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{normal,T}\left(1 - U_{k,N}\left(1 - \frac{N^\rho}{N^\rho+1}\right)\right)\\
&= \frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{normal,T}\left(1 - \left(\frac{N}{k}U_{k,N}\right)\frac{N^\delta}{N}\frac{1}{N^\rho+1}\right)\\
&\sim \frac{k^{1/2+1/\kappa}\sqrt{\log(N)}}{N^{1/\kappa}}\frac{1}{T^p}.
\end{aligned}$$

Thus, the scaled G term decays if

$$\frac{k^{1/2+1/\kappa}\sqrt{\log(N)}}{N^{1/\kappa}}\frac{1}{T^p} \to 0. \tag{A.3.8}$$

If the above restrictions on k, N, T hold, then $\tilde{S}_{N,T}(\tilde{u}_{S,N,T}) \to 0$. Same restrictions are implied by the requirement $\tilde{s}_{N,T}(\tilde{u}_{s,N,T}) \to 0$ where $\tilde{u}_{s,N,T} = 1 - U_{k,N}(N^\rho - 1)/N^\rho$. Then conditions (2) and (3) hold by the same argument as in section A.3.2.

[15] Corollary 2.2.2 in de Haan and Ferreira (2006) asserts that $(k/N)Y_{N-k,N} \xrightarrow{p} 1$ where $Y_{N-k,N}$ are order statistics of $\{Y_1, \dots, Y_N\}$, and $P(Y_i \le y) = 1 - 1/y$ if $y \ge 1$ and zero otherwise. Observe that $U_{k,N} \overset{d}{=} 1/Y_{N-k,N}$, the result then follows. See also the proof of theorem 3.4

Rate conditions if $k = N^\delta$ Conditions (A.3.6), (A.3.7), and (A.3.8) are general conditions that jointly restrict k, N, T. The conditions can be specialized based on the form of k. The leading standard choice is $k = N^\delta, \delta < 1$. In this case condition (A.3.6) transforms to the requirement that $N^{\delta/2-\rho} \to 0$, which holds if $\rho > \delta/2$. Condition (A.3.7) becomes

$$\frac{N^{\delta/2+\delta/\kappa-\delta/\beta}}{N^{1/\kappa-1/\beta-\rho/\beta}T^p} = \frac{N^{\delta/2+(1-\delta)(1/\beta-1/\kappa)+\rho/\beta}}{T^p} \to 0.$$

Similarly, condition (A.3.8) becomes

$$\frac{N^{\delta/2+(1-\delta)(-1/\kappa)}\sqrt{\log(N)}}{T^p} \to 0$$

Write $\rho = \delta/2 + \nu$ where $\nu > 0$, then the above condition transforms into

$$\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta-1/\kappa)+\nu/\beta}}{T^p} \to 0. \tag{A.3.9}$$

In particular, observe that ν can be taken arbitrarily close to 0.

Choice of $\tilde{u}_{S,N,T}$ The choice of $\tilde{u}_{S,N,T}$ and the resulting rate conditions is driven by the desired choice of k. This is most apparent in the derivation of (A.3.6). If we instead use $\tilde{u}_{S,N,T} = 1 - U_{k,N}\log(T)/(\log(T) + 1)$ (similarly to section (A.3.2)), then (A.3.6) is replaced by

$$\sqrt{k}\left[1 - \left(\frac{\log(T)}{\log(T)+1}\right)^{1/\kappa}\right] \to 0$$

which is compatible with k growing at most as $o(\log^2(T))$, typically a much stronger restriction than $k = o(N)$. This motivates our choice of $\tilde{u}_{S,N,T}$ in eq. (A.3.5) as that compatible with $k = N^\delta, \delta < 1$.

Comparison with rate conditions for the EVT We remark that the rate conditions for the extreme and intermediate value theorems will generally differ (compare eq. (A.3.9) to eq. (A.3.3)). The fundamental reason is that the two theorems require asymptotic tail equivalence to hold at different portions of the tail, with the discrepancy controlled by the magnitude of k. The smaller the value of δ, the closer condition (A.3.9) is to condition

(A.3.3). This effect is also visible in the choice of $u_{S,\tau,N,T}$ and $\tilde{u}_{S,N,T}$: choice of $u_{S,\tau,N,T}$ for the EVT has relatively little importance, while choice of $\tilde{u}_{S,N,T}$ for intermediate order statistics is tightly related to the chosen value of k, as remarked above.

Example 6, page 17 Now let $F = F_{Gu,\lambda}$. All the remarks above apply equally to this case, and we limit ourselves to obtaining the corresponding rate conditions for the infimum.

First we compute the normalizing functions of theorem 3.4

$$F^{-1}_{Gu,\lambda}\left(1-\frac{k}{N}\right) = U_{Gu,\lambda}\left(\frac{N}{k}\right) = \frac{\log(N/k)}{\lambda},$$
$$c_N = \frac{N}{k} \times \left(\left(\frac{1}{1-F_{Gu,\lambda}}\right)^{-1}\right)'\left(\frac{N}{k}\right) = \frac{N}{k}U'_{Gu,\lambda}\left(\frac{N}{k}\right) = \frac{1}{\lambda}\frac{N}{k}\frac{k}{N} = \frac{1}{\lambda}.$$

Pick $\tilde{u}_{S,N,T}$ as in eq. (A.3.5):

$$u = 1 - U_{k,N}\frac{N^\rho}{N^\rho+1}.$$

Then

$$\begin{aligned}
&\sqrt{k}\left(F^{-1}_{Gu,\lambda}\left(1-U_{k,N}\frac{N^\rho}{N^\rho+1}\right) - F^{-1}_{Gu,\lambda}\left(1-U_{k,N}\right)\right)\\
=&\sqrt{k}\left(\frac{\log U_{k,N} + \log\frac{N^\rho+1}{N^\rho}}{\lambda} - \frac{\log U_{k,N}}{\lambda}\right)\\
\sim&\sqrt{k}\log\left(1+\frac{1}{N^\rho}\right) \sim \sqrt{k}N^{-\rho}.
\end{aligned}$$

Let $k = N^\delta$, then the above expression decays if $\rho > \delta/2$.

Let $G_T = G_{T,\beta}$. In this case

$$\begin{aligned}
&\sqrt{k}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-U_{k,N}\left(1-\frac{N^\rho}{N^\rho+1}\right)\right)\\
&\sim k^{1/2}\frac{(N)^{\rho/\beta}}{U^{1/\beta}_{k,N}T^p}\\
&= \left(\frac{N}{k}U_{k,N}\right)^{-1/\beta}\frac{k^{1/2-1/\beta}}{N^{-1/\beta-\rho/\beta}T^p}.
\end{aligned}$$

Since $(N/k)U_{k,N} \xrightarrow{p} 1$, for the above expression to decay it is sufficient that

$$\frac{k^{1/2-1/\beta}}{N^{-1/\beta-\rho/\beta}T^p} \to 0.$$

With our choice of $k = N^\delta$ the condition resolves into

$$\frac{N^{\delta/2-\delta/\beta}}{N^{-1/\beta-\rho/\beta}T^p} = \frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta)+\nu/\beta}}{T^p} \to 0,$$

where ν is any fixed number > 0.

Let $G_T = G_{normal,T}$. Then

$$\sqrt{k}\frac{1}{T^p}G^{-1}_{normal,T}\left(1 - U_{k,N}\left(1 - \frac{N^\rho}{N^\rho+1}\right)\right) \sim k^{1/2}\frac{\sqrt{\log(N)}}{T^p}.$$

If $k = N^\delta$, then the above decays if

$$\frac{N^{\delta/2}\sqrt{\log(N)}}{T^p} \to 0.$$

Example 7, page 17 Finally, let $F = F_{W,\alpha}$. We proceed as in the previous two examples.

First compute the normalizing functions of theorem 3.4.

$$F^{-1}_{W,\alpha}\left(1 - \frac{k}{N}\right) = U_{W,\alpha}\left(\frac{N}{k}\right) = \theta_F - \theta_F\left(\frac{k}{N}\right)^{1/\alpha},$$

$$c_N = \frac{N}{k}\left(\left(\frac{1}{1-F_{W,\alpha}}\right)^{-1}\right)'\left(\frac{N}{k}\right) = \frac{N}{k}U'_{F_{W,\alpha}}\left(\frac{N}{k}\right) = \frac{N}{k}\frac{\theta_F}{\alpha}\left(\frac{N}{k}\right)^{-\frac{1}{\alpha}-1} = \frac{\theta_F}{\alpha}\left(\frac{k}{N}\right)^{1/\alpha}.$$

Pick $\tilde{u}_{S,N,T}$ as in eq. (A.3.5):

$$\tilde{u}_{S,N,T} = 1 - U_{k,N}\frac{N^\rho}{N^\rho+1}.$$

Then

$$\begin{aligned}
&\frac{N^{1/\alpha}}{\frac{\theta_F}{\alpha}k^{1/\alpha-1/2}}\left(F^{-1}_{W,\alpha}\left(1 - U_{k,N}\frac{N^\rho}{N^\rho+1}\right) - F^{-1}_{W,\alpha}\left(1 - U_{k,N}\right)\right)\\
\sim&\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}}\left[\left(U_{k,N}\right)^{1/\alpha} - \left(\frac{N^\rho}{N^\rho+1}\right)^{1/\alpha}\left(U_{k,N}\right)^{1/\alpha}\right]\\
=&\sqrt{k}\left(\frac{N}{k}U_{k,N}\right)^{-1/\alpha}\left[1 - \left(\frac{N^\rho}{N^\rho+1}\right)^{1/\alpha}\right].
\end{aligned}$$

As for the two previous cases, the above decays if k is such that

$$\sqrt{k}\left[1 - \left(\frac{N^\rho}{N^\rho+1}\right)^{1/\alpha}\right] \sim \sqrt{k}N^{-\rho} \to 0.$$

If $k = N^\delta$, then the above decays if $\rho > \delta/2$.

Now turn to the G_T terms. Let $G_T = G_{\beta,T}$. Then

$$\frac{1}{T^p} G_{\beta,T}^{-1}\left(1 - U_{k,N}\left(1 - \frac{N^\rho}{N^\rho + 1}\right)\right) \sim \frac{(N)^{\rho/\beta}}{U_{k,N}^{1/\beta} T^p}.$$

Multiplying by the scaling constants, we see that the $G_{\beta,T}$ term to decay it is sufficient that

$$\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}} \frac{N^{\rho/\beta}}{U_{k,N}^{1/\beta} T^p} = \frac{N^{1/\alpha+1/\beta+\rho/\beta}}{k^{1/\alpha+1/\beta-1/2} T^p} \left(\frac{N}{k} U_{k,N}\right)^{-1/\beta} \sim \frac{N^{1/\alpha+1/\beta+\rho/\beta}}{k^{1/\alpha+1/\beta-1/2} T^p} \to 0\ .$$

If $k = N^\delta$ and $\rho = \delta/2 + \nu, \nu > 0$, the above transforms into

$$\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\alpha+1/\beta)+\nu/\beta}}{T^p} \to 0.$$

If $G_T = G_{normal,T}$, then

$$\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}} G_{normal,T}^{-1}\left(1 - U_{k,N}\left(1 - \frac{N^\rho}{N^\rho + 1}\right)\right) \sim \frac{N^{1/\alpha}}{k^{1/\alpha-1/2}} \frac{\sqrt{\log(N)}}{T^p}.$$

If $k = N^\delta$, then the above expression decays if

$$\frac{N^{\delta/2+(1-\delta)(1/\alpha)}\sqrt{\log(N)}}{T^p} \to 0.$$

Appendix B

Additional Results

B.1 Unbalanced Data

In some applications, the data available may be unbalanced. For example, in panel data analysis, the time series for some units may be shorter or longer, while in a meta-analysis setting, individual studies may have varying sample sizes.

In this section we discuss how to handle unbalanced data under the assumption that the minimal individual sample size tends to infinity.

Formally, let the observable $\vartheta_{i,T}$ be generated as

$$\vartheta_{i,T} = \theta_i + \frac{1}{T_i^p}\varepsilon_{i,T_i}, \tag{B.1.1}$$

where T_i is the individual sample size of unit i and $\varepsilon_{i,T_i} = O_p(1)$. Define T, λ_i so that $T_i = \lambda_i T$, $\lambda_i \geq 1$ and $T \to \infty$. Observe that by construction $T_i \geq T$, and so T can be interpreted as a minimal sample size.

To conduct inference on extreme quantiles of F when the observable $\vartheta_{i,T}$ are generated by (B.1.1), we represent eq. (B.1.1) as a special case of the setup studied in the main text. We assume that the individual sample size $T_i = \lambda_i T$ is also random. Define $\tilde{\varepsilon}_{i,T} = \lambda_i^{-p}\varepsilon_{i,T_i}$. With this definition we can write

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\tilde{\varepsilon}_{i,T}.$$

The setup can be intuitively interpreted in a hierarchical manner: first θ_i is drawn, then the observed sample size T_i (λ_i) is drawn for i, then with that T_i we draw ε_{i,T_i} from G_{T_i}. Finally, the components are combined into the observable $\vartheta_{i,T}$ as in eq. (B.1.1).

Remark 17. The assumption that the sample size T_i is random is not restrictive. T_i can be related to θ_i in a complex manner, and we do not restrict the joint distribution of θ_i and

$\tilde{\varepsilon}_{i,T}$, similarly to the main text. Such dependence might be present in applications. Consider an economic example: let i index firms, and let θ_i be firm productivity. If firms with low productivity go bankrupt and exit the market at higher rates than more productive firms, then firms with lower θ_i will tend to have lower sample sizes T_i available. Our setup allows such relationships, under the assumption that minimal sample size is still appropriately large.

Define $\tilde{G}_T$ to be the cdf of $\tilde{\varepsilon}_{i,T}$. Then all the results established in the main text apply with $\tilde{G}_T$ in place of G_T and $T \to \infty$. It is also easy to apply sufficient conditions of propositions 3.3 and 3.5 if moment conditions are available for G_T. For example, let $\sup_T \mathbb{E}|\varepsilon_{i,T}|^\beta < \infty$ for some β. Since $\lambda_i^{-p} \leq 1$ a.s., we obtain that $\sup_T \mathbb{E}|\tilde{\varepsilon}_{i,T}| < \infty$ (intuitively, the bound on tails of G_T is uniform, and tails of $\tilde{G}_T$ are an average of tails of G_T, averaging over the distribution of sample sizes).

B.2 Deterministic Conditions For Intermediate Order Tail Equivalence Conditions

For completeness, we provide a deterministic sufficient condition for conditions (2) and (3) of theorem 3.4.

Proposition 1. *Let $U_1, \dots, U_N$ be iid Uniform[0, 1]. Let $\delta_{k,N,T}$ be such that as $T, N(T), k(N) \to \infty$, $k(N) = o(N)$*

$$P\left(\left|\frac{N}{k}U_{k,N} - 1\right| \geq \delta_{k,N,T}\right) \to 0. \tag{B.2.1}$$

Also let c_N be a sequence of constants that is eventually positive (if $c_N < 0$ for all N, the proposition holds with all signs of $\delta_{k,N,T}$ switched). Define

$$s_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}(u) + \frac{1}{T^p}G_T^{-1}\left(1 - \frac{k}{N}(1+\delta) - u\right) - F^{-1}\left(1 - \frac{k}{N}(1-\delta)\right)\right),$$

$$S_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}(u) + \frac{1}{T^p}G^{-1}\left(1 + 1 - \frac{k}{N}(1-\delta) - u\right) - F^{-1}\left(1 - \frac{k}{N}(1+\delta)\right)\right)$$

for all δ, u such that the above functions are well-defined. If

$$\sup_{u \in \left[0, 1 - \frac{k}{N}(1+\delta_{k,N,T})\right]} s_{s,k,N,T}(u, \delta_{k,N,T}) \to 0,$$

$$\inf_{u\in\left[1-\frac{k}{N}(1-\delta_{k,N,T}),1\right]} S_{s,k,N,T}(u,\delta_{k,N,T}) \to 0,$$

then

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

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

A sequence $\delta_{k,N,T}$ of eq. (B.2.1) always exists since $\frac{N}{k}U_{k,N}\xrightarrow{p}1$; $\delta_{k,N,T}$ depends only on k,T,N.

Proof of proposition 1. Define the event

$$A_N=\left\{\left|\frac{N}{k}U_{k,N}-1\right|\le\delta_{k,N,T}\right\}.$$

By assumption, $P(A_N)\to 1$. On A_N it holds that $(N/k)U_{k,N}\in(1-\delta_{k,N,T},1+\delta_{k,N,T})$.

Suppose that c_N is eventually positive (if not, switch $-\delta_{k,N,T}$ to $+\delta_{k,N,T}$ and vice versa in the main function). Then on A_N it is also true that

$$\begin{aligned}
&\inf_{u\in\left[1-U_{k,N},1\right]} \frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1+1-U_{k,N}-u\right)-F^{-1}\left(1-U_{k,N}\right)\right)\\
=&\inf_{u\in\left[1-U_{k,N},1\right]} \frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1+1-\frac{k}{N}\frac{N}{k}U_{k,N}-u\right)-F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}\right)\right)\\
\le&\inf_{u\in\left[1-\frac{k}{N}(1-\delta_{k,N,T}),1\right]} S_{s,k,N,T}(u,\delta_{k,N,T})\\
\to&\ \ 0.
\end{aligned}$$

To obtain the above inequality, we decrease the choice set for the inf. In G^{-1} we take $(1-\delta_{k,N,T})$, and in the second F^{-1} we take $(1+\delta_{k,N,T})$, this corresponds to largest possible value of the resulting expression for each u. Last line follows by the assumption of the proposition.

We proceed in the exact same manner for the supremum: on A_N it holds that

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

$$
\begin{aligned}
&= \sup_{u\in\left[0,1-U_{k,N}\right]} \frac{\sqrt{k}}{c_N}\left(F^{-1}(u)+\frac{1}{T^p}G_T^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}-u\right)-F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}\right)\right)\\
&\geq \sup_{u\in\left[0,1-\frac{k}{N}(1+\delta)\right]} s_{s,k,N,T}(u,\delta_{k,N,T})\\
&\quad \to 0.
\end{aligned}
$$

To obtain the inequality, we decrease the choice set for the supremum and choose the smallest values in the quantiles.

Finally, observe that the Makarov inequalities of lemma A.1.3 also show that for the original random supremum and infimum

$$
\sup_{u\in\left[0,1-U_{k,N}\right]}\{\dots\} \leq \inf_{u\in\left[1-U_{k,N},1\right]}\{\dots\}.
$$

This implies that on event A_N both the random supremum and the random infimum converge to zero. Since $P(A_N)\to 1$, this establishes convergence i.p. □

B.3 Additional Simulation Results

In this section we provide additional results related to the simulation study of section 5 of the main text. First, in section B.3.1 we consider the performance of different estimators for quantiles proposed in section 4. In section B.3.2 we report the full results of our simulation study, covering all confidence intervals and all combinations of distributions for θ_i and u_{it}. Finally, we repeat our simulation study for noiseless data in order to understand the importance of our rate conditions and the impact of estimation noise versus the impact of a limited cross-sectional size.

B.3.1 Corrected Estimators For Quantiles

We begin by briefly assessing the performance of adjusted estimators for quantiles proposed in example 10 and in section 4.4. Performance is compared in terms of mean absolute error. We work in the setup of section 5. We report results for both $T = 15, 30$. u_{it} is drawn from the heavy-tailed G_β distribution and from a $N(0,1)$ distribution.

We now describe the estimators compared. Suppose interest lies in $F^{-1}(q)$. To form median-unbiased estimators based on extreme approximations, let l solve $1 - l/N = q$, and set $r = \lfloor l \rfloor$. Let q be a fixed number. "Mixed" estimators are based on adjusting the sample quantile. Let $\hat{c}_\alpha$ be a consistent estimator of the αth quantile of $\left[(E_1^* + \cdots + E_{r+1})^{-\gamma} - l^{-\gamma}\right] / \left[\left(E_1^* + \cdots + E_{q+1}^*\right)^{-\gamma} - (E_1^*)^{-\gamma}\right]$.

$$\mathcal{M}_{N,T,q}^{mixed} = \vartheta_{N-r,N,T} - \hat{c}_{1/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right) \tag{B.3.1}$$

"Max only" estimators instead adjust the sample maximum using the corresponding limit distribution. Let $\tilde{c}_\alpha$ be an estimator of the αth quantile of $\left[(E_1^*)^{-\gamma} - l^{-\gamma}\right] / [\left(E_1^* + \cdots + E_{q+1}^*\right)^{-\gamma} - (E_1^*)^{-\gamma}]$. The estimator is then defined as

$$\mathcal{M}_{N,T,q}^{Max\ only} = \vartheta_{N,N,T} - \tilde{c}_{1/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)$$

$\hat{c}_\alpha$ and $\tilde{c}_\alpha$ are consistently estimated by subsampling and by simulation, as described in section 4.2. Finally, we also consider the adjusted estimator of Jochmans and Weidner (2022) $\hat{\vartheta}_{\lfloor N\hat{\tau}^* \rfloor,N,T}$, which is based on the central order approximations of section 4.4.

We plot the results graphically on figs. 4-9. In all cases we report the mean absolute error (MAE) relative to the unadjusted sample quantile. Values greater than 1 mean that the estimator performs worse than the sample quantiles, values below 1 signify better performance. On the top panel of each figure we plot relative MAE on a scale from 0.5 to 1; on the bottom panel we plot the relative MAE on a scale 0.5 to 15 in order to capture the magnitude of breakdown of several estimators.

Based on the results, we recommend using unadjusted sample quantiles or estimator (B.3.1) with quantiles $\hat{c}_\alpha$ estimated by subsampling (line "Subsampling: mixed" on figures). Estimator (B.3.1) performs similarly to the unadjusted sample quantile: it has relative MAE <1 for a slight majority of quantiles considered and does not suffer from significant breakdowns. We recommend against all other correction methods and against using quantiles estimated by simulation.

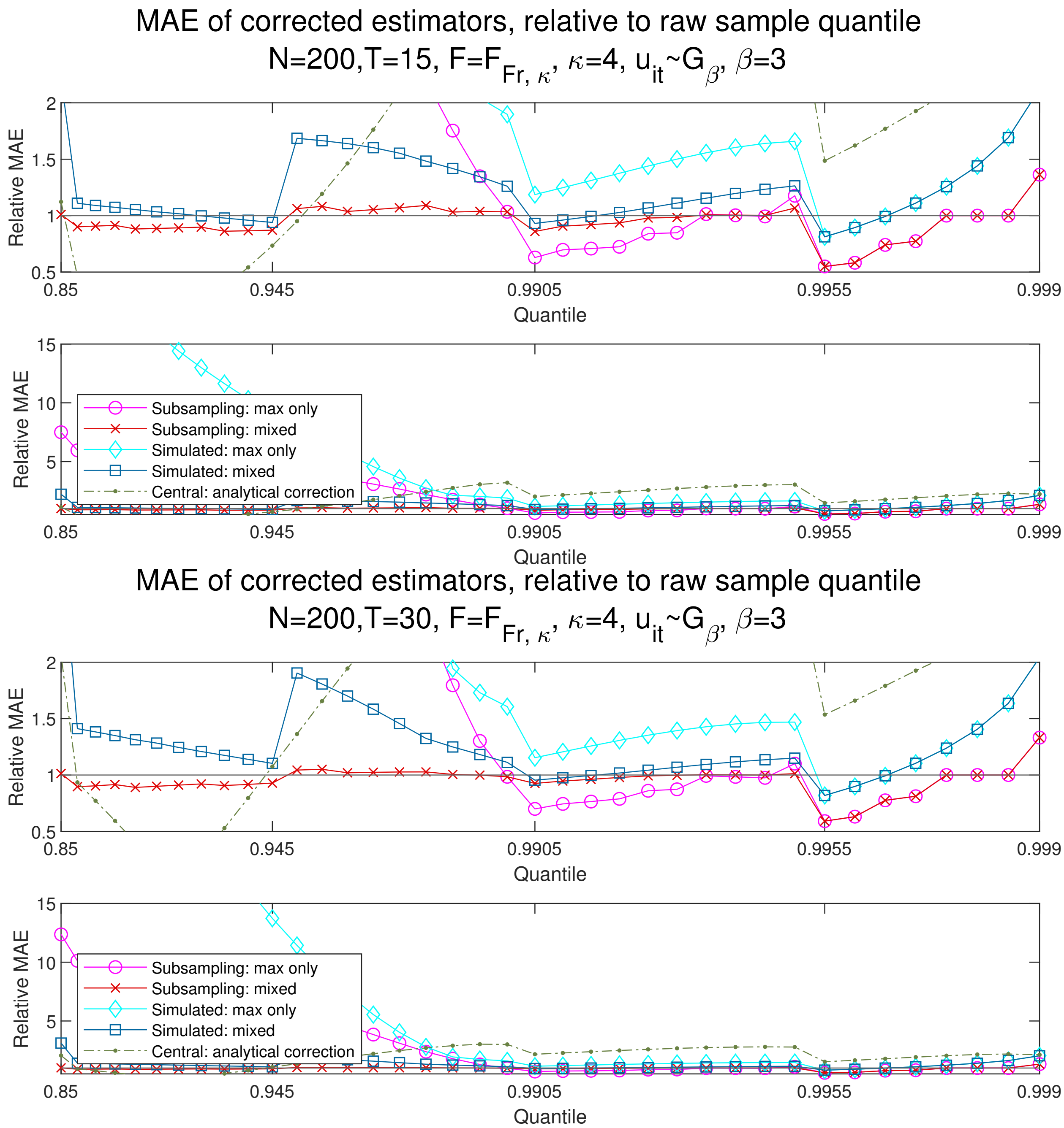


Figure 4: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 3$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

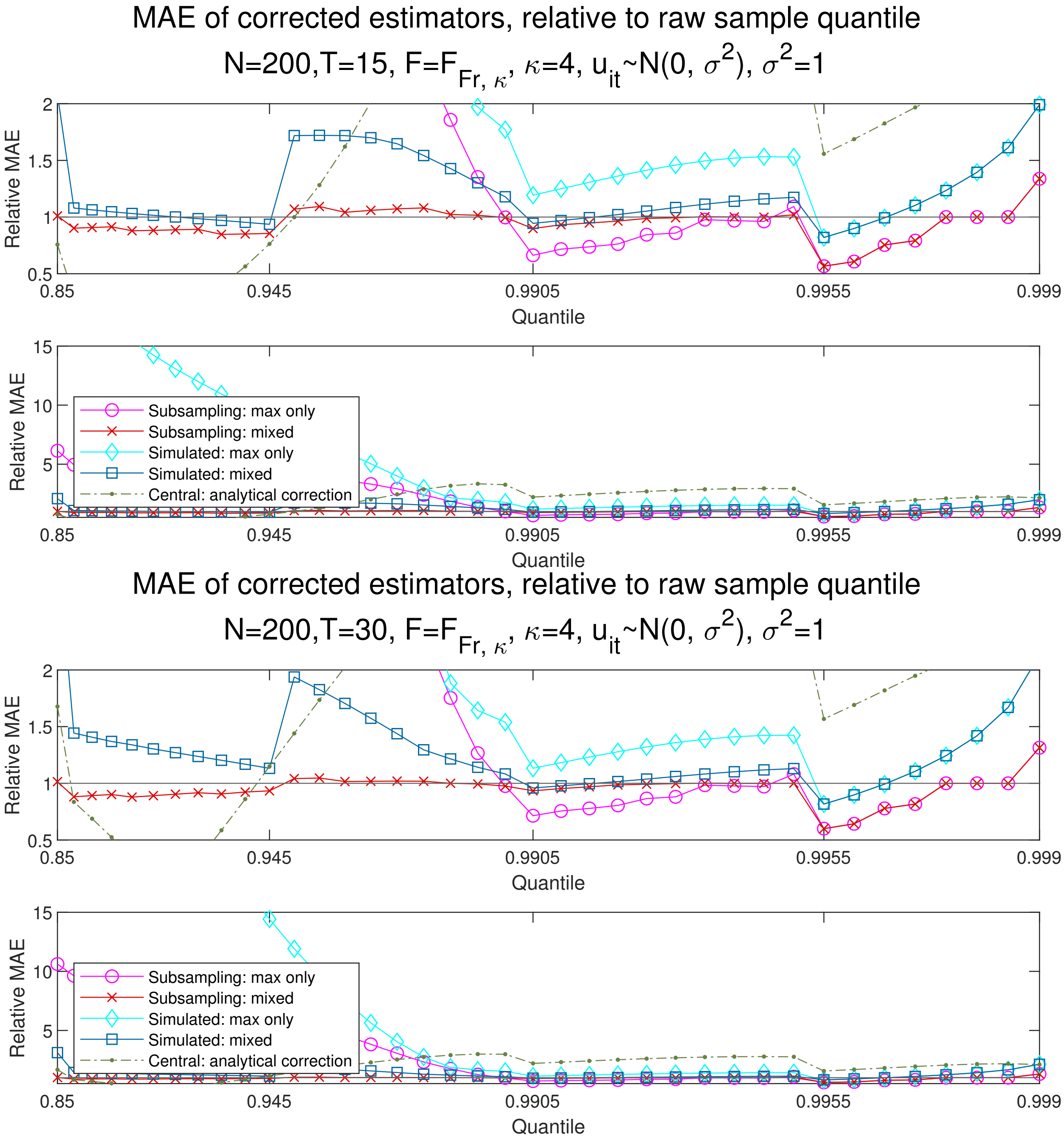


Figure 5: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

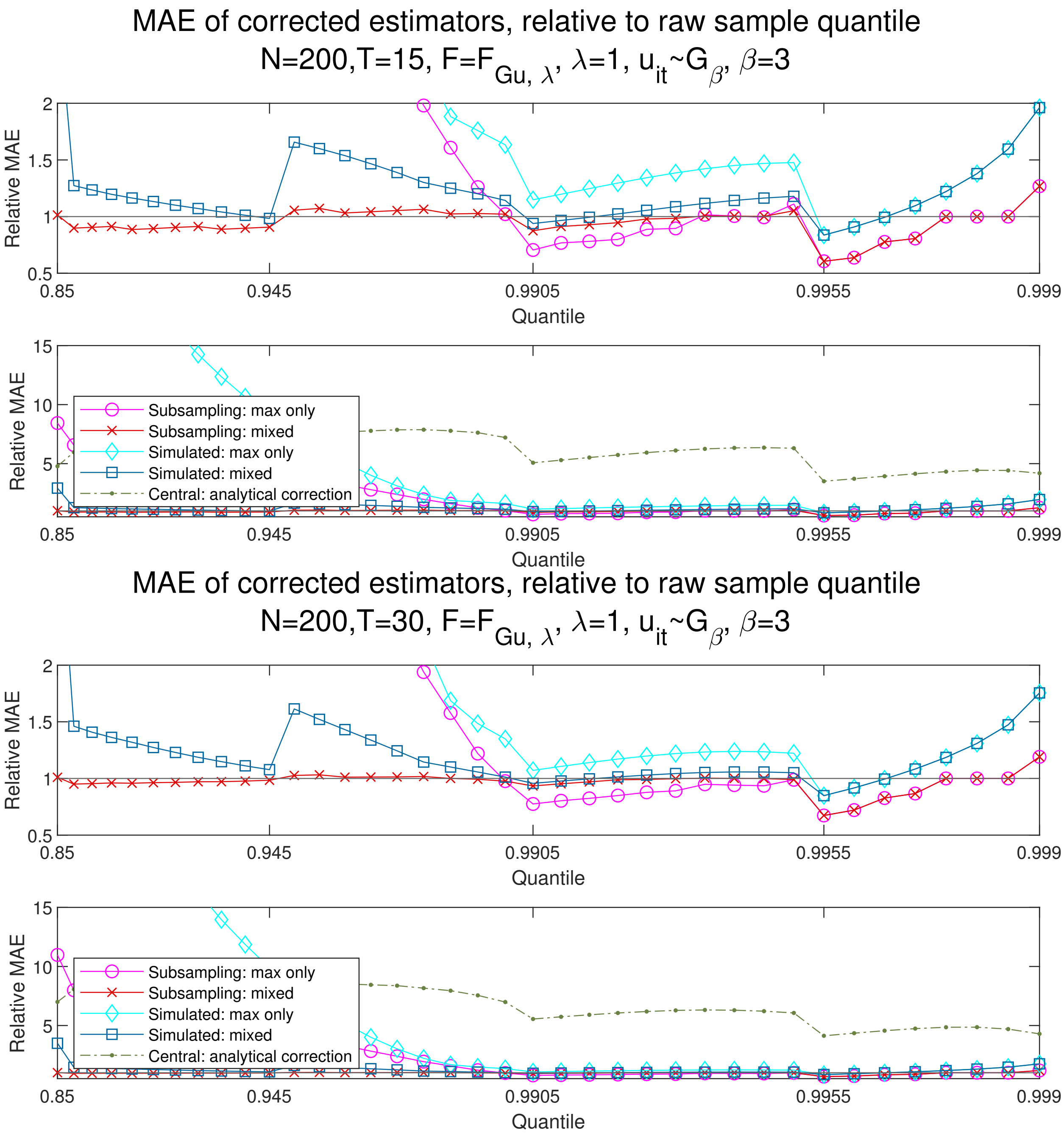
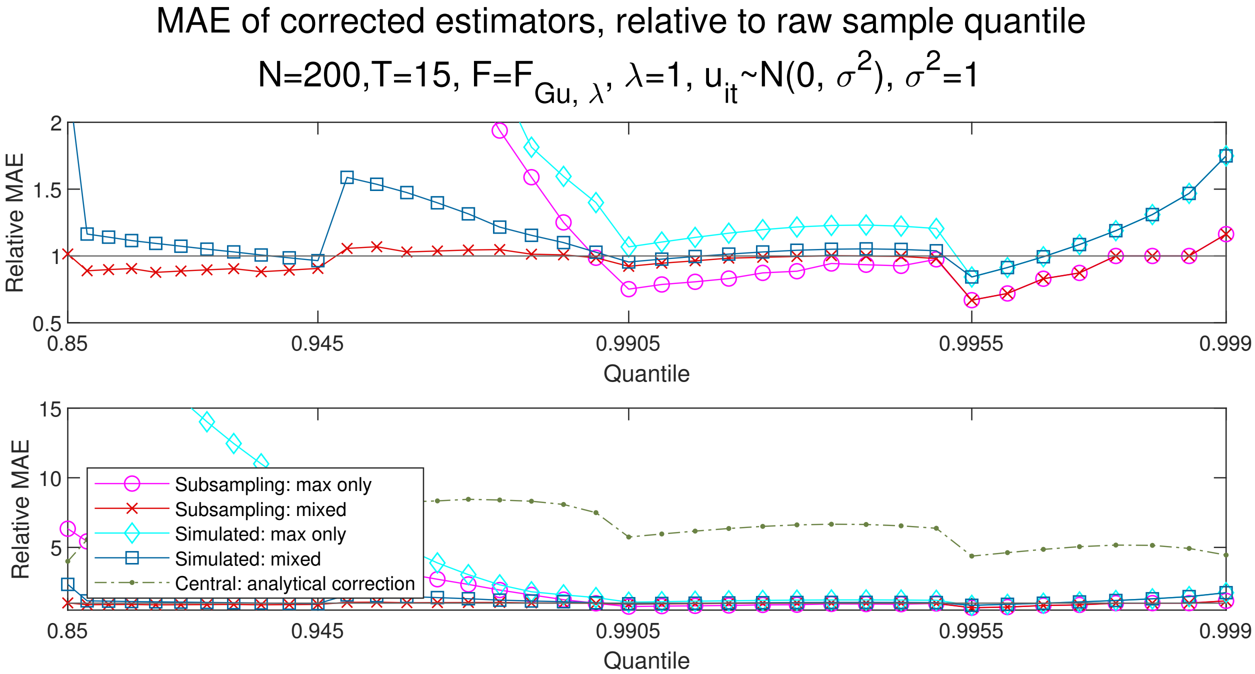


Figure 6: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 3$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.



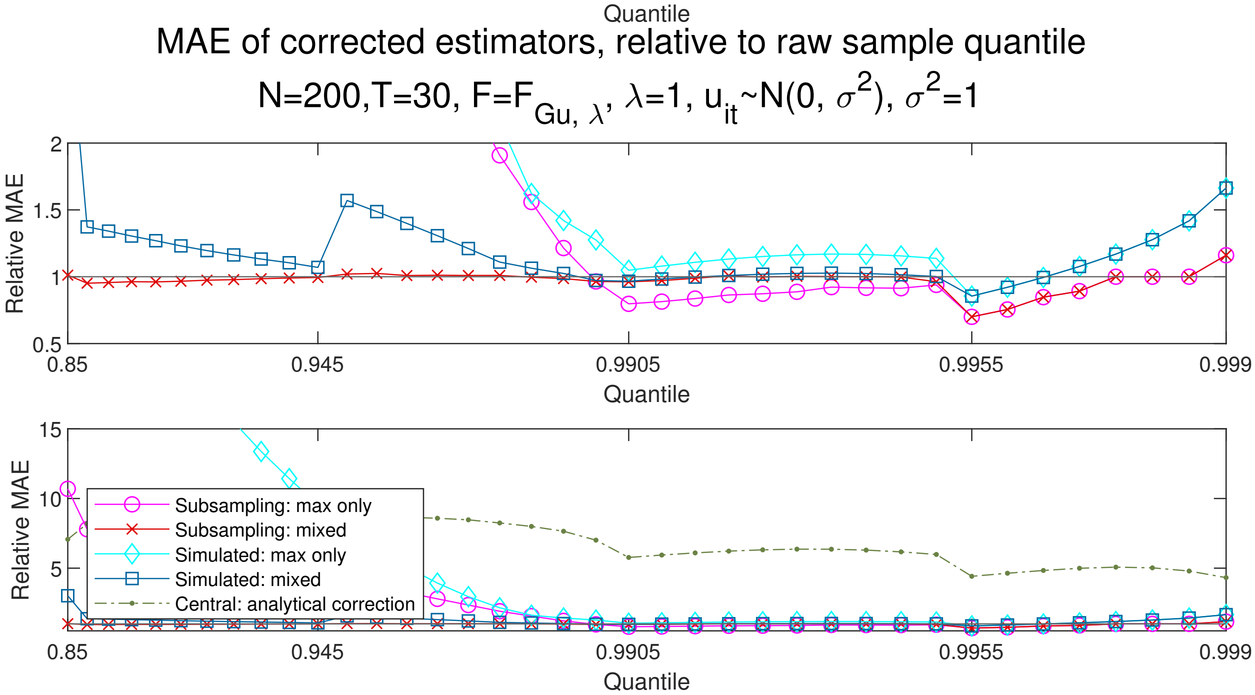


Figure 7: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

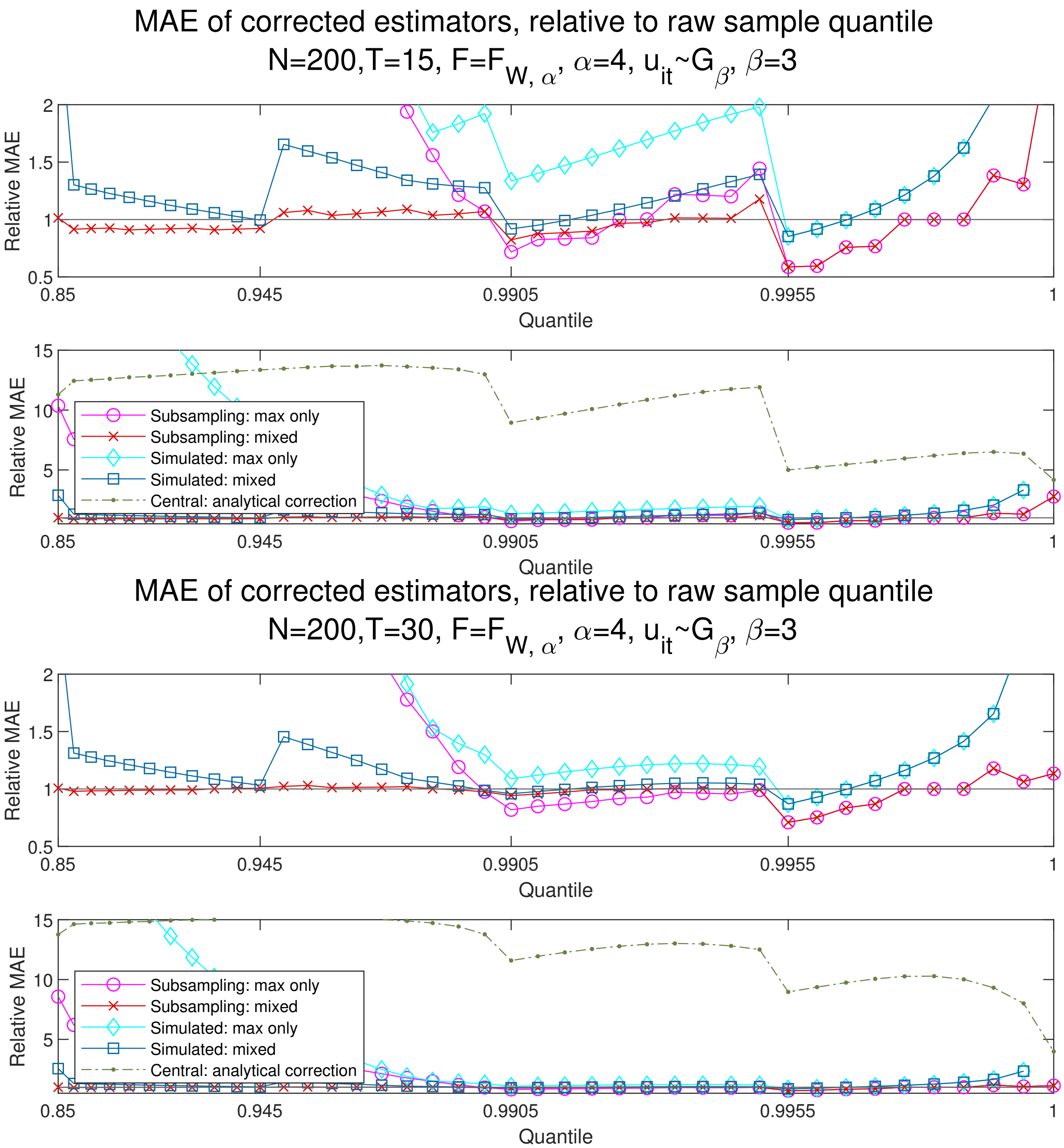


Figure 8: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_\beta, \beta = 3$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

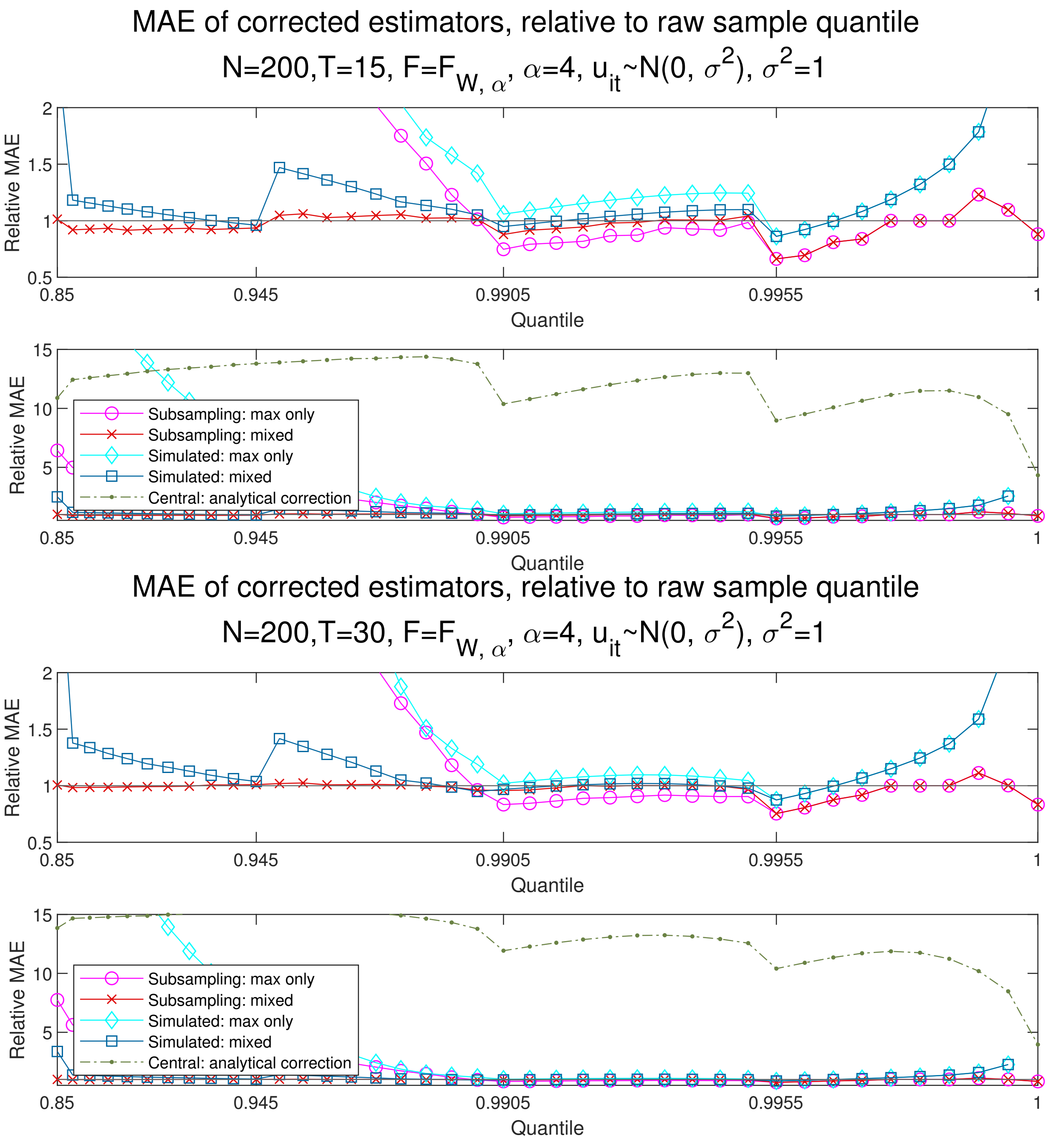


Figure 9: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

B.3.2 Full Results for Confidence Intervals

In this section we report the full results of the simulation study described in section 5 of the main text. As in the main text, we consider two sample sizes $(N = 200, T = 15)$ and $(N = 200, T = 30)$. We report results for two distributions of u_{it}: $u_{it} \sim G_\beta$ where $G_\beta = G_{\beta,T}$ with $\mu_T = 0$, $\beta = 3$, and $u_{it} \sim N(0, 1)$ (see example 5). We consider three distributions for parameter of interest θ_i, corresponding to the three distributions considered in examples 5-7.

We report coverages and lengths for the following confidence intervals:

(1) Subsampled: based on the feasible EVT 4.3. Quantiles of the limit distribution are estimated using subsampling as in theorem 4.4. We report the two CIs proposed in example 10. The "max only" CI is centered at the maximum for all quantiles, and only the critical values track the target quantile. The "mixed" CI is instead centered at the corresponding sample quantile for each quantile of interest. We pick $q = 2$.

(2) Simulated: based on the feasible EVT 4.3. Quantiles of the limit distribution are estimated by simulating from eq. (10) with γ estimated as in theorem 4.5. We report three confidence intervals: the "max only" CI and the "mixed" CI are defined as before. In addition, we report the estimator where as the estimator for γ we use the truncated Pickands estimator defined as

$$\tilde{\gamma} = \begin{cases} \hat{\gamma}_P, & \operatorname{sgn}(\hat{\gamma}_P) = \operatorname{sgn} \gamma, \\ 0 & \text{else.} \end{cases}$$

(3) Intermediate: based on the normal approximation of theorem 4.6. We report two intervals: one using asymptotic quantiles from a $N(0, 1)$ distribution, and one using subsampling to estimate quantiles.

(4) Central: raw data. We construct the CI using the raw ϑ_i using a binomial CI.

(5) Central with the analytical correction of Jochmans and Weidner (2022). Based on theorem 4.7: we report the CI centered $\vartheta_{\lfloor N\hat{\tau}^*\rfloor,N,T}$. As advocated by Jochmans and Weidner (2022), we report bootstrap confidence intervals with 1000 bootstrap samples.

The full results are plotted on figures 10-21. We refer to the main text for a full discussion.

For inference on extreme quantiles, we recommend using the mixed CI based on an extreme approximation, using subsampled critical values. For all distributions, it combines good coverage with favorable length properties. While the max-only CI with subsampled critical values has slightly better coverage properties, that comes at the price of a significantly longer interval. Truncation in $\hat{\gamma}_P$ appears to generally worsen coverage, and we recommended against it. As in the main text, we also advise against all options based on intermediate approximations for the sample size considered. While using subsampled critical values in place of asymptotic ones improves coverage, both intervals are dominated by CIs based on extreme approximations.

B.3.3 Results for Noiseless Data

In this section we repeat our simulation exercise with data not subject to estimation noise. This effectively corresponds to setting $T = \infty$ and $\vartheta_{i,T} = \theta_i$. Such simulations allow us to separate the impact of estimation noise (small T effect) from the impact of small cross-sectional size (small N effect). This is achieved by comparing the results of this section to those of sections B.3.1 and B.3.2.

We consider the same setup as in the main text. N is equal to 200. The corrected estimators for quantiles are as in section B.3.1. The confidence intervals considered are as in section B.3.2. We remark that the two central approximations are identical without estimation noise; consequently we plot a single line for the central approximation. Our results are reported on figs. 22-24.

As in the main text, we find that the rate conditions of section 4 are important for inference. To see this, let θ_i be drawn from a $F_{W,\alpha}$ distribution and consider figs. 18 and 20. Our rate conditions hold for $T = 30$, but not for $T = 15$. Consequently, for $T = 30$ the confidence intervals have the same coverages as in the noiseless case, and the small T effect is practically eliminated. This result can be seen by comparing the results for $T = 30$ with fig. 24. All the size distortions that remain in figs. 18 and 20 for $T = 30$ are also present in fig. 24; thus these distortions are due to the small N effect. Similarly, the differences in coverages between figs. 10, 12, and 22 are minor, as rate conditions hold both for $T = 15$ and $T = 30$ on figs. 10 and 12.

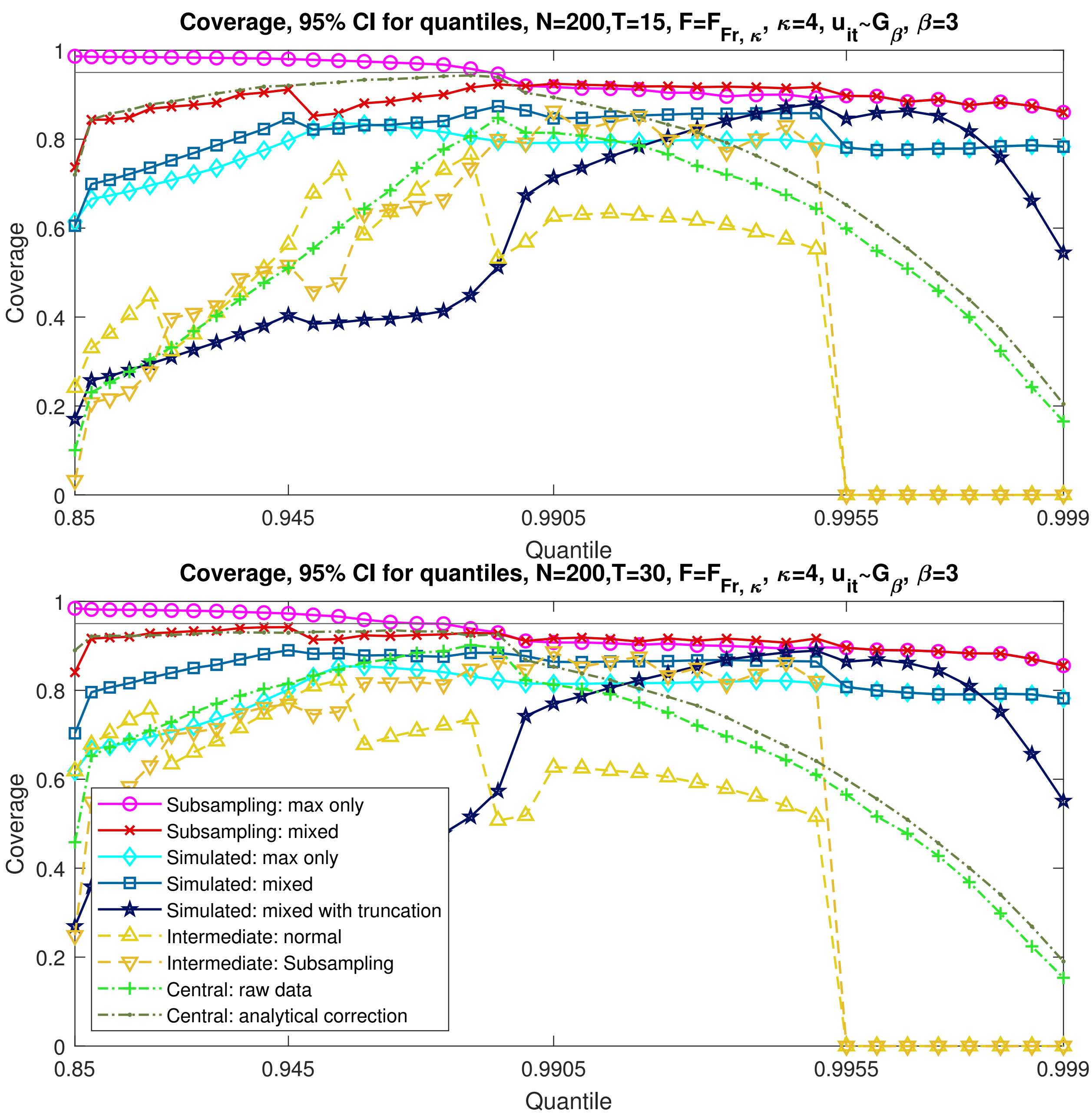


Figure 10: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 3$

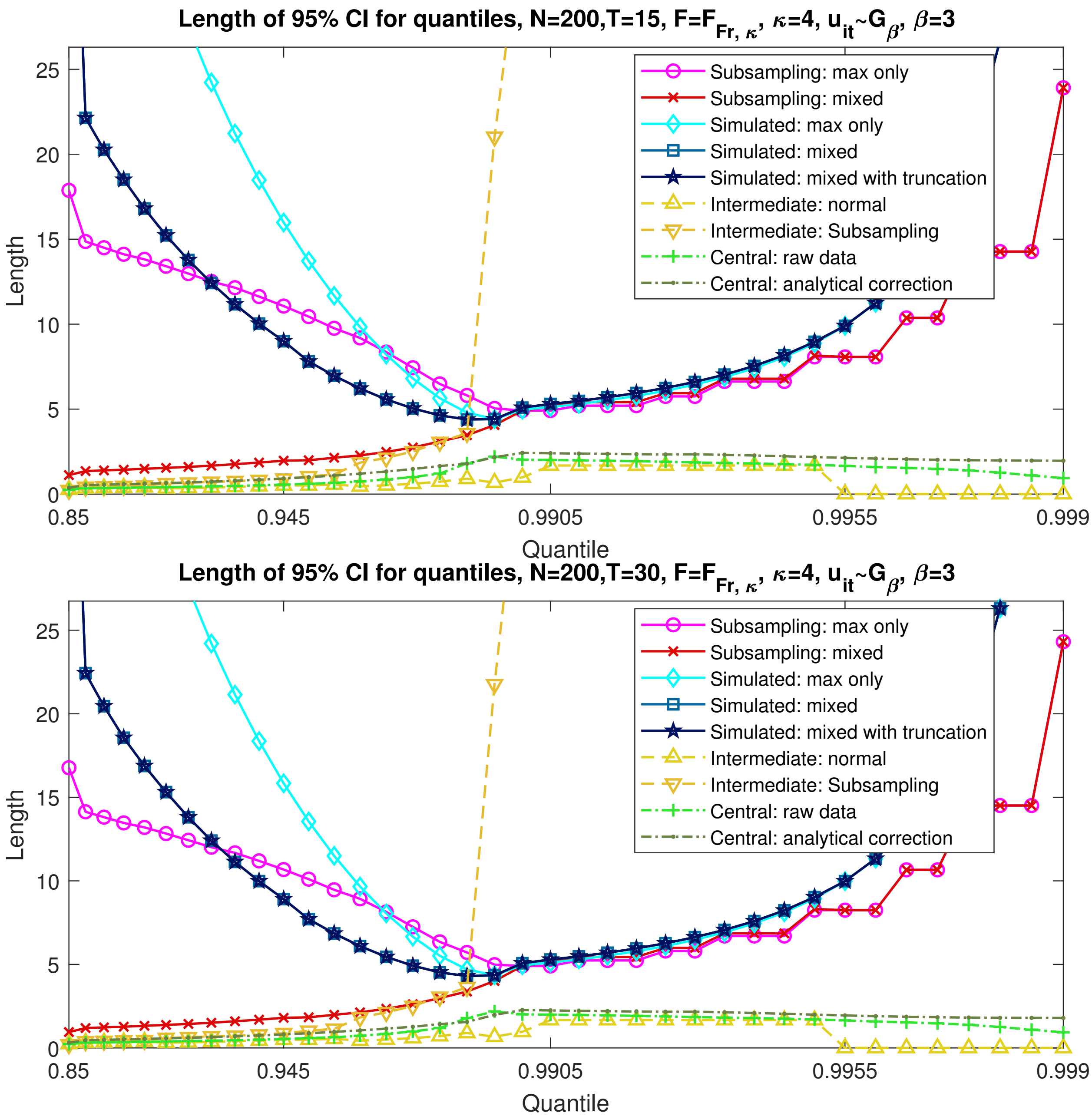


Figure 11: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 3$

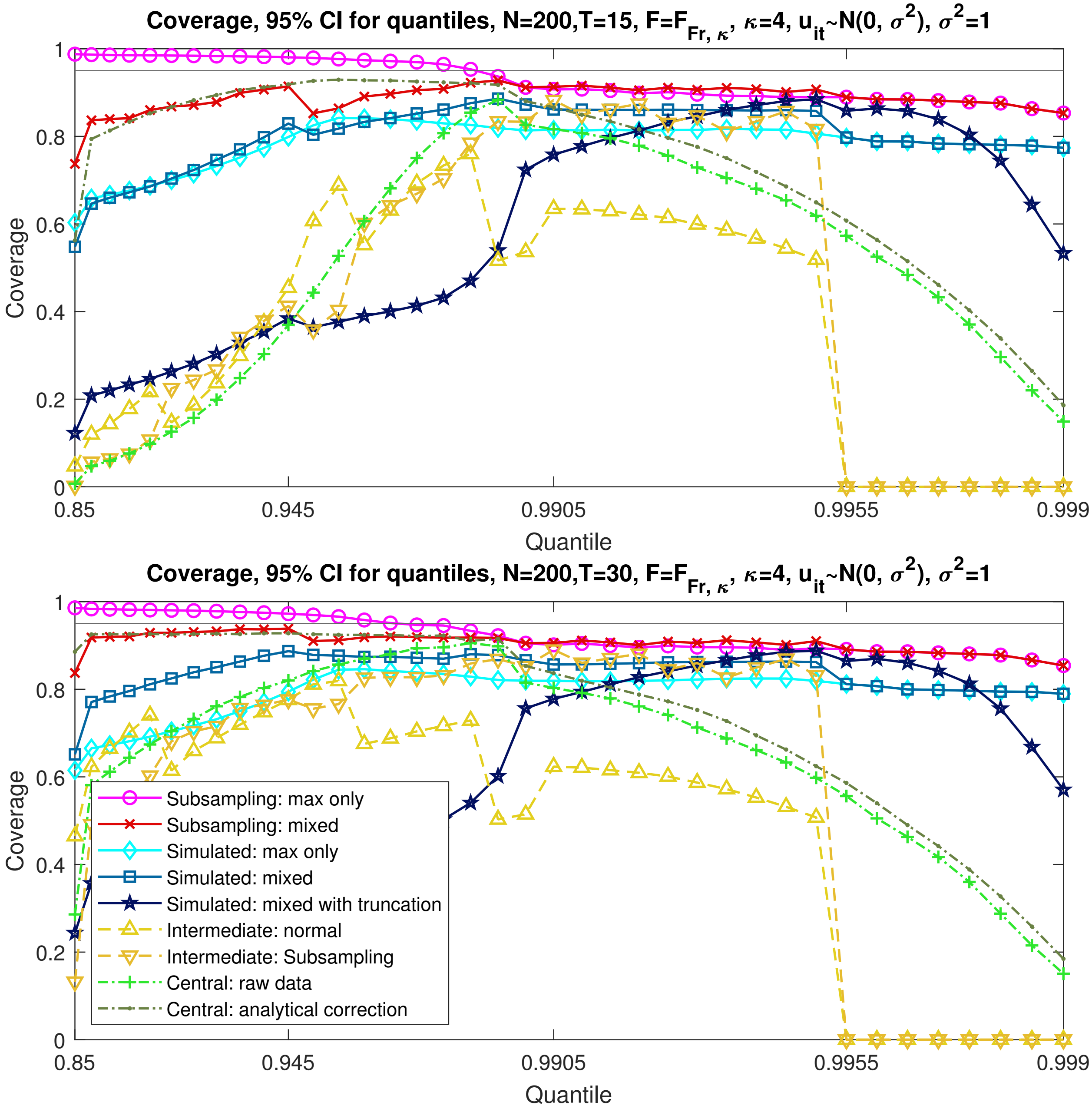


Figure 12: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$

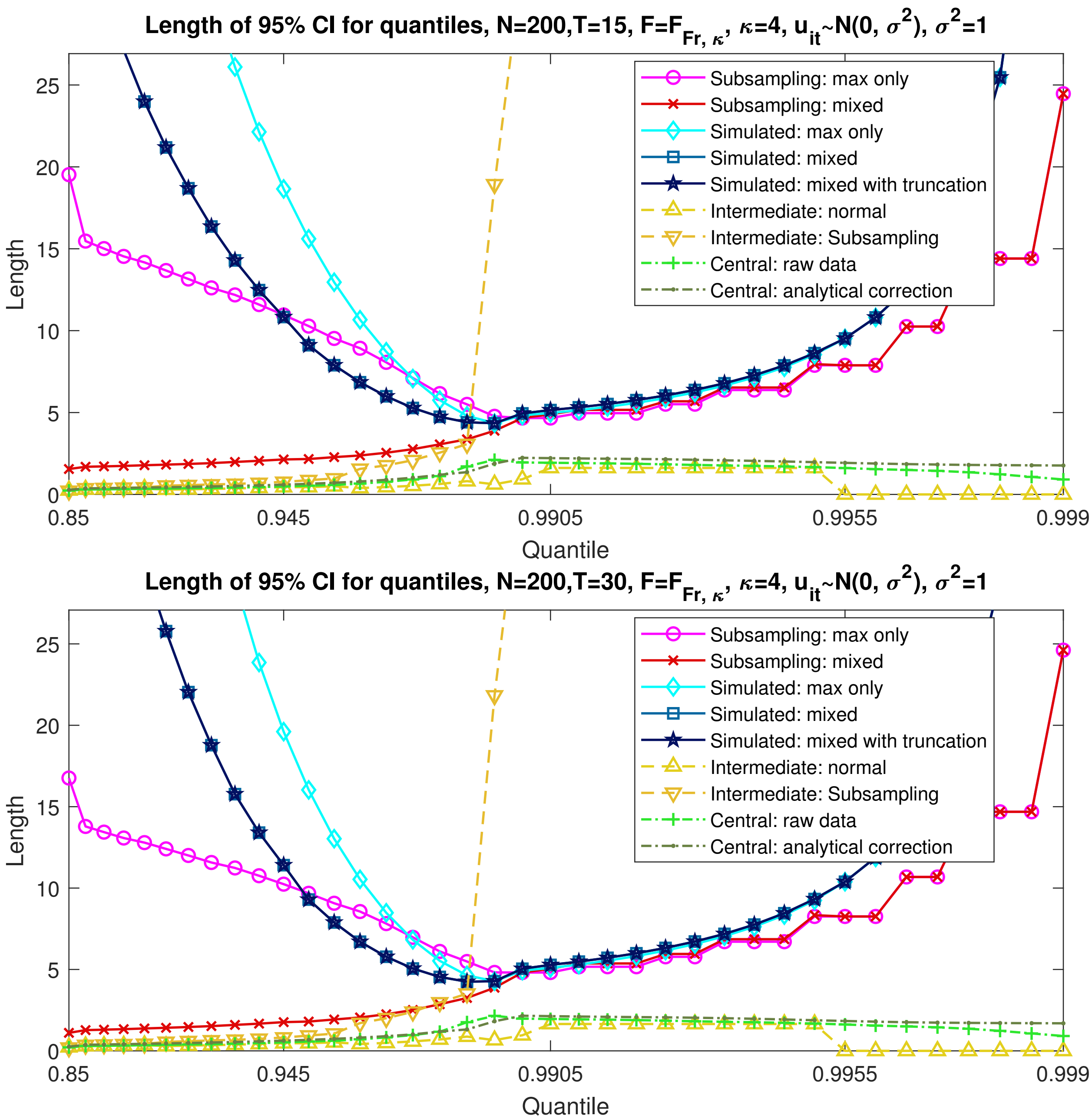


Figure 13: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$

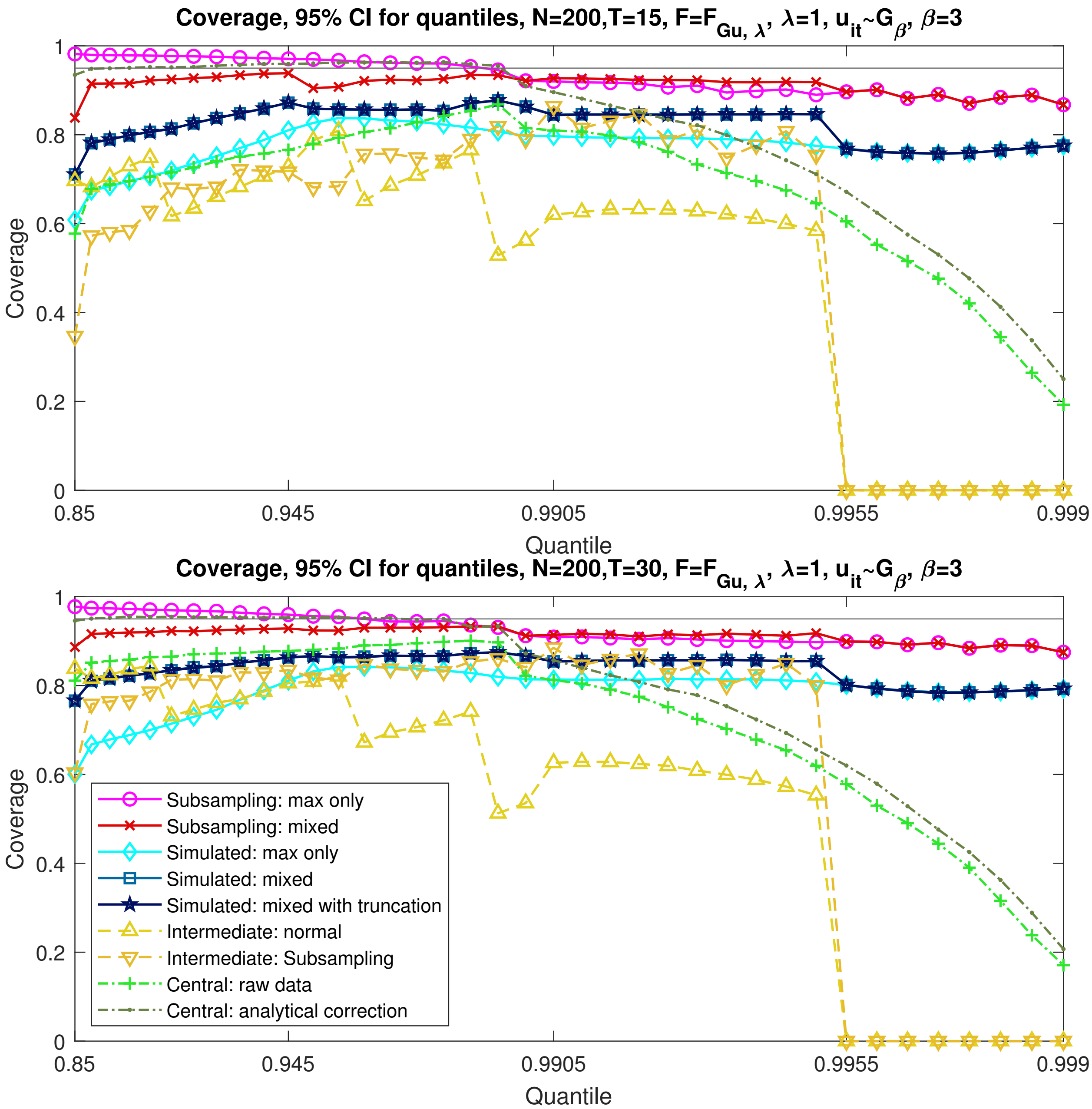


Figure 14: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 3$

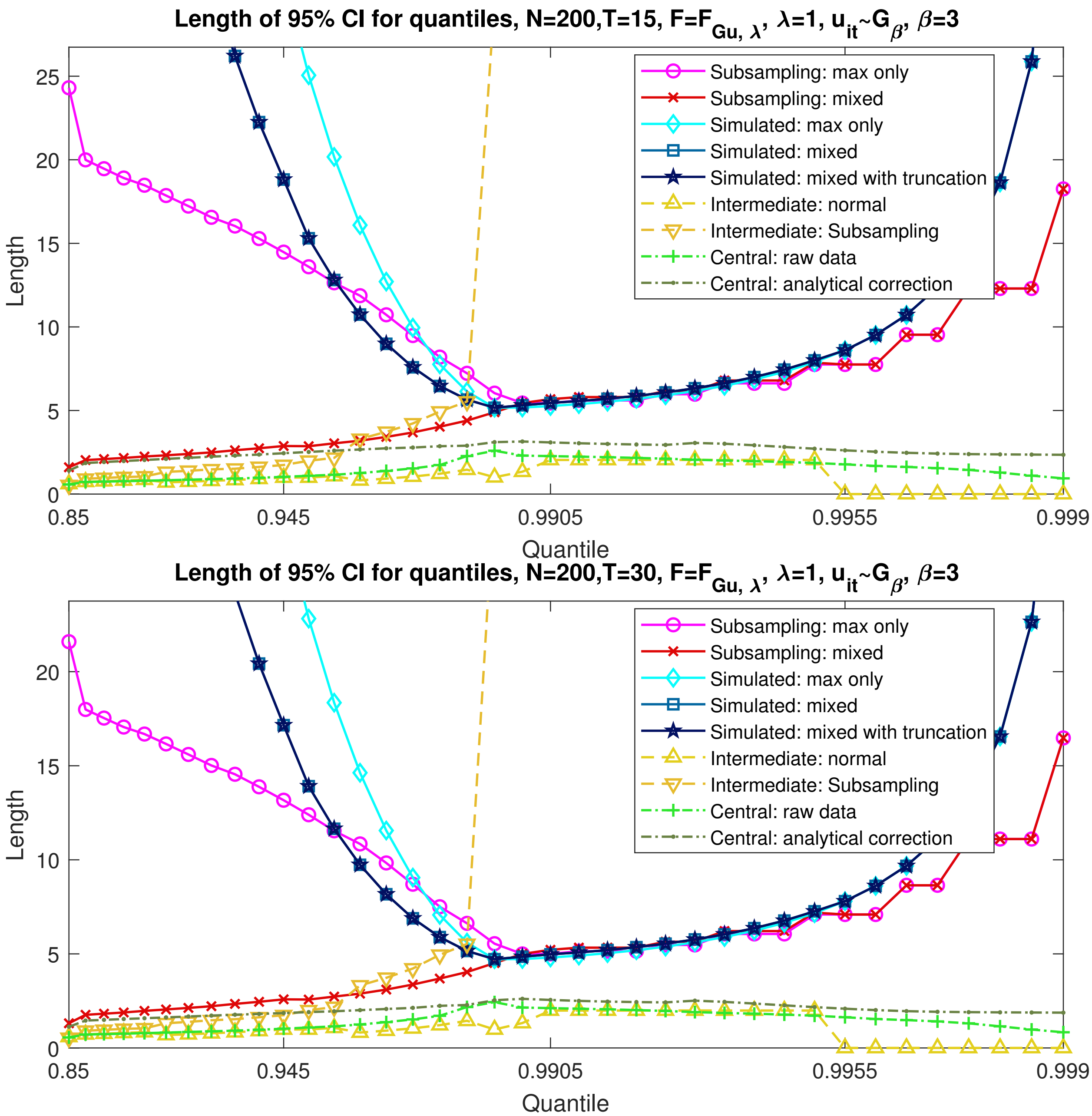


Figure 15: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 3$

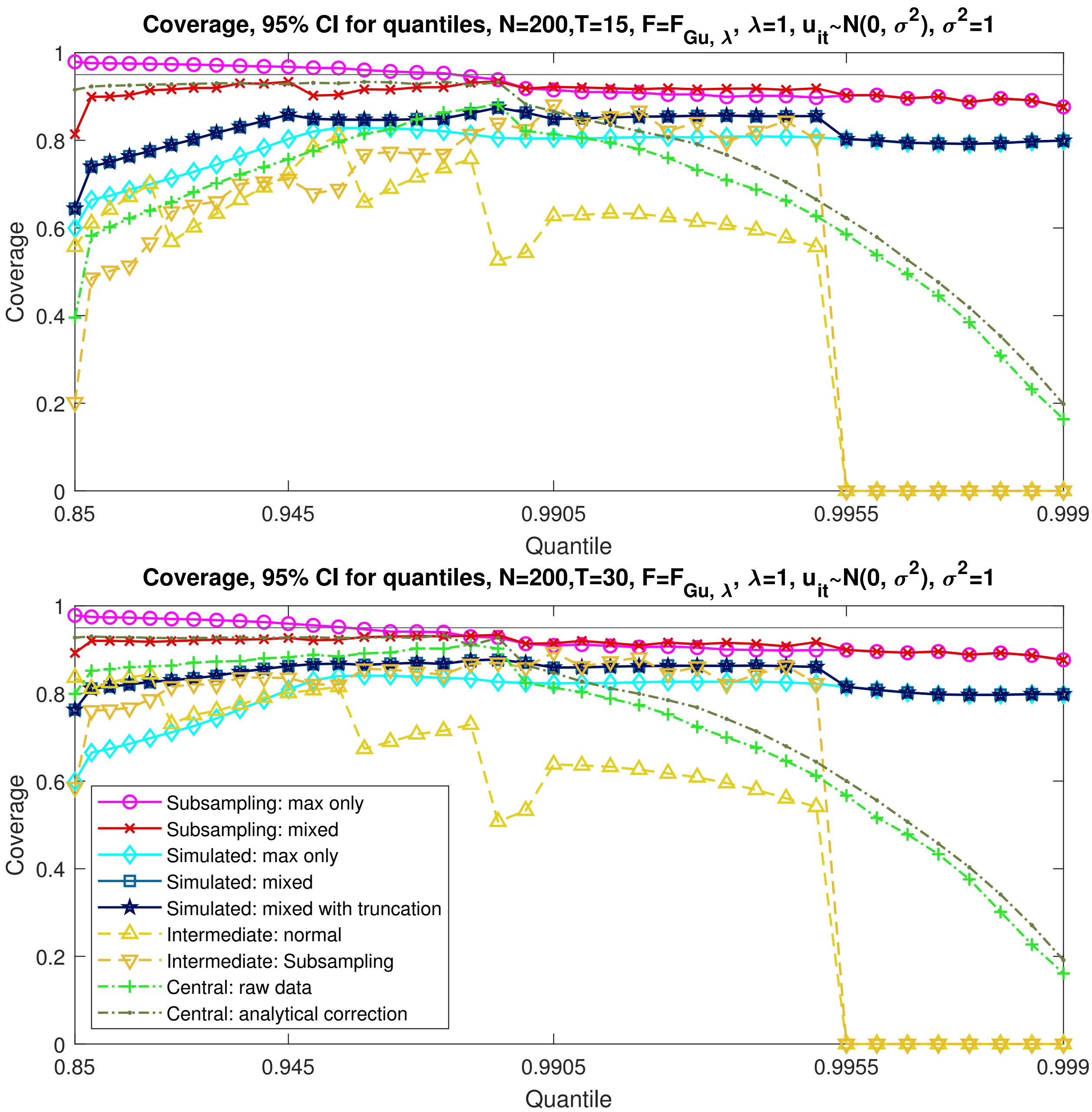


Figure 16: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim N(0, 1)$

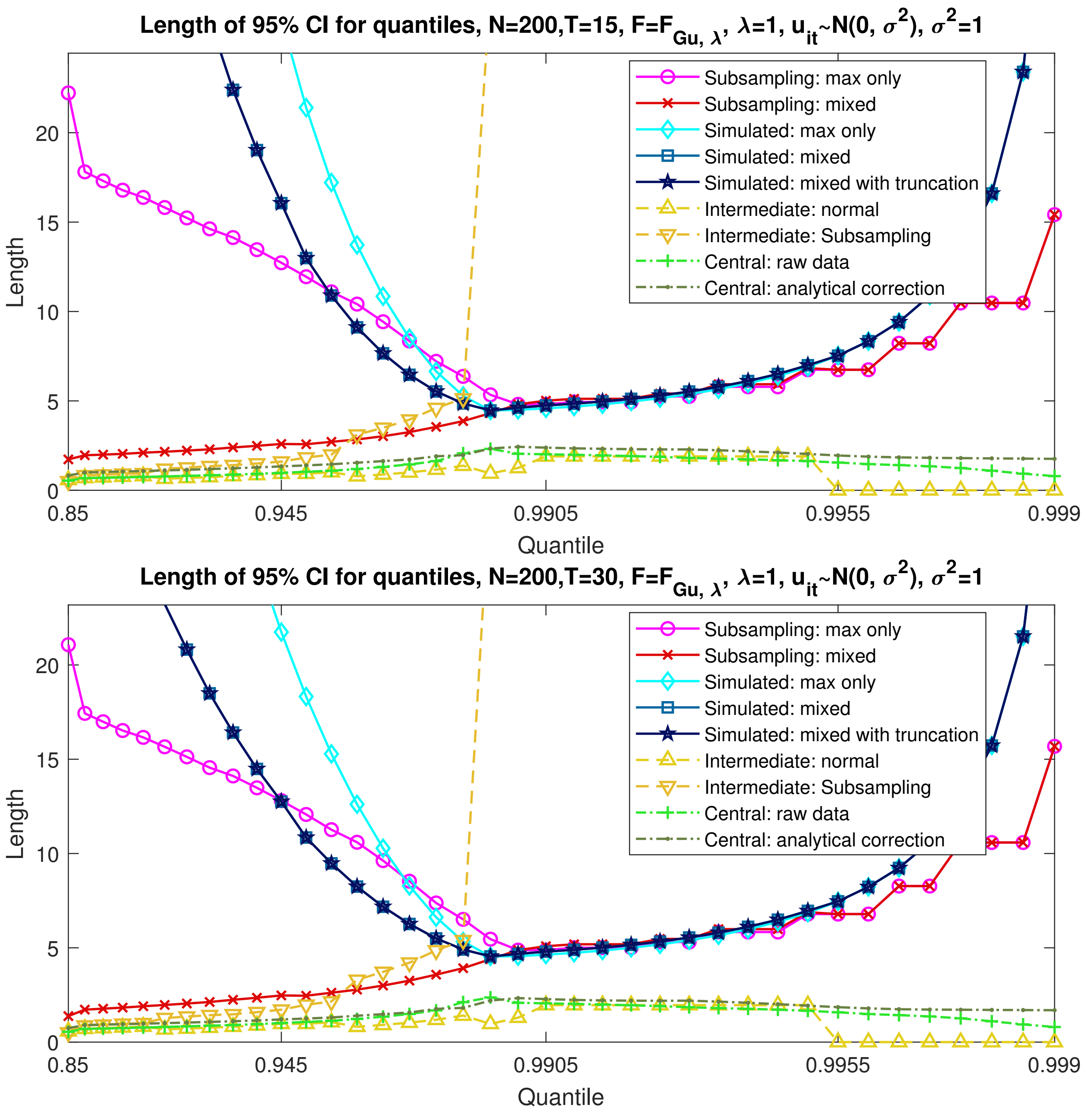


Figure 17: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 3$

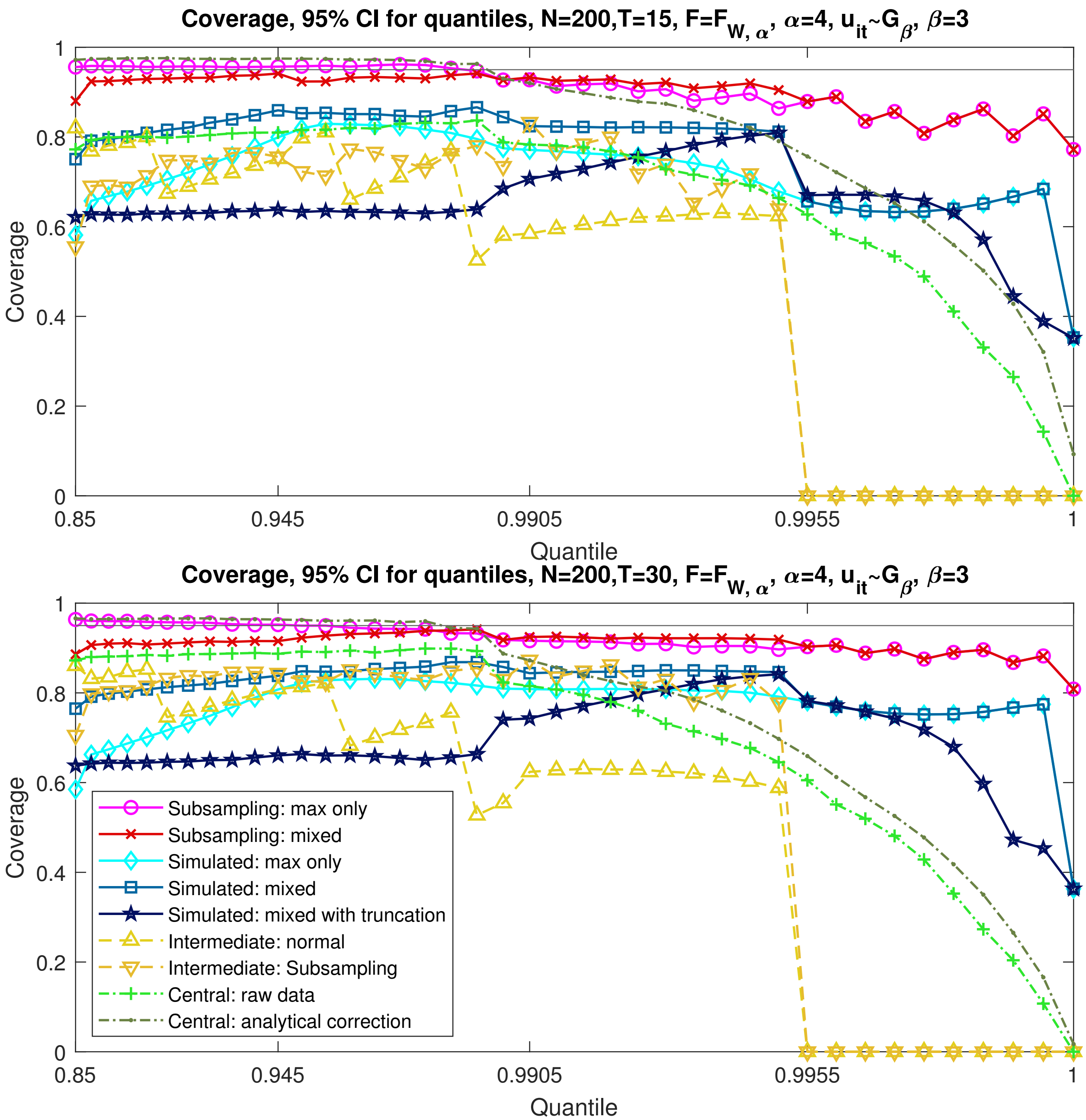
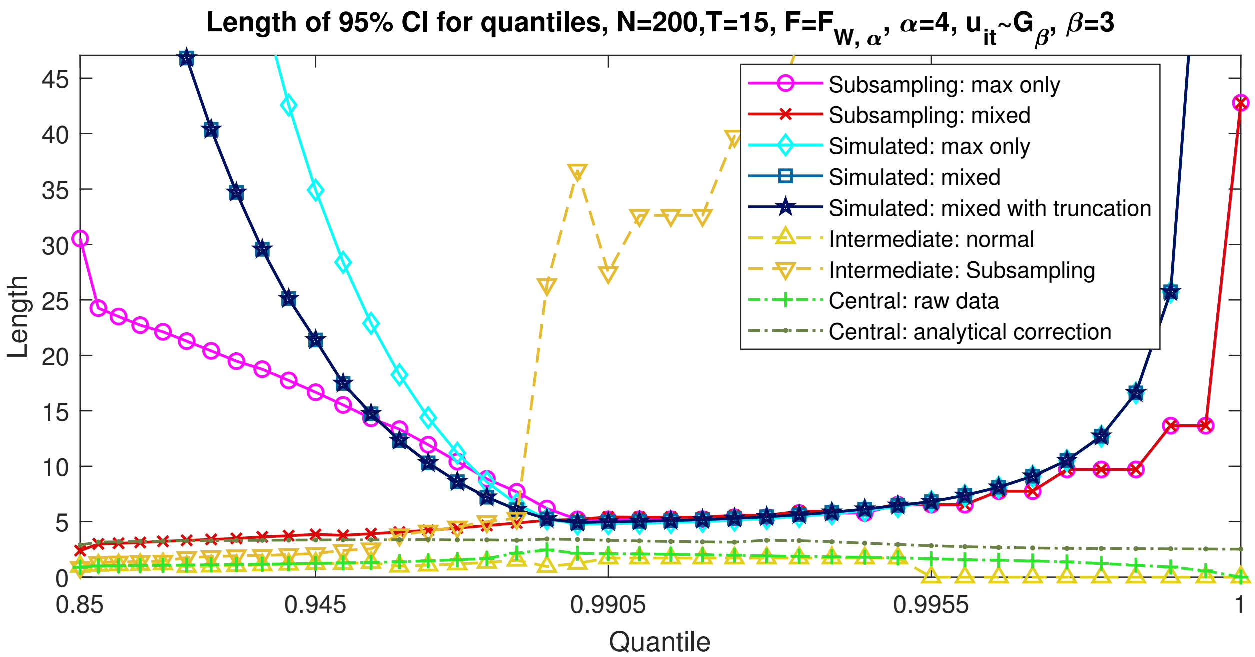


Figure 18: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_\beta, \beta = 3$



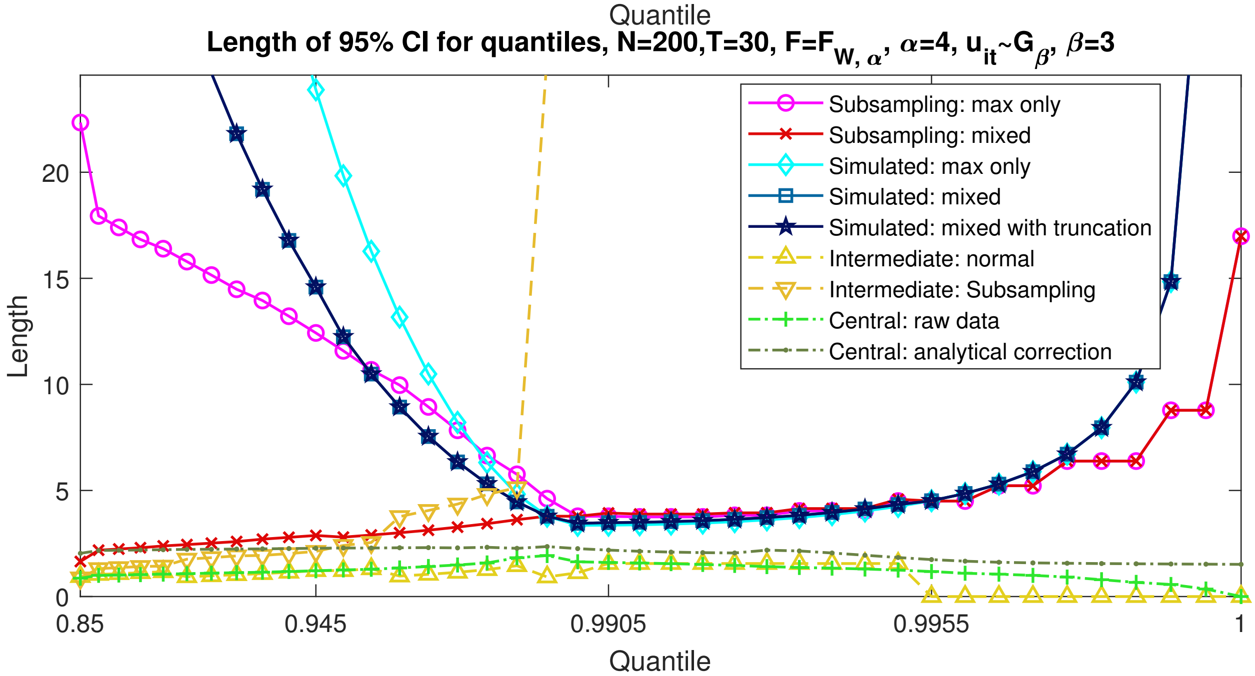


Figure 19: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_\beta, \beta = 3$

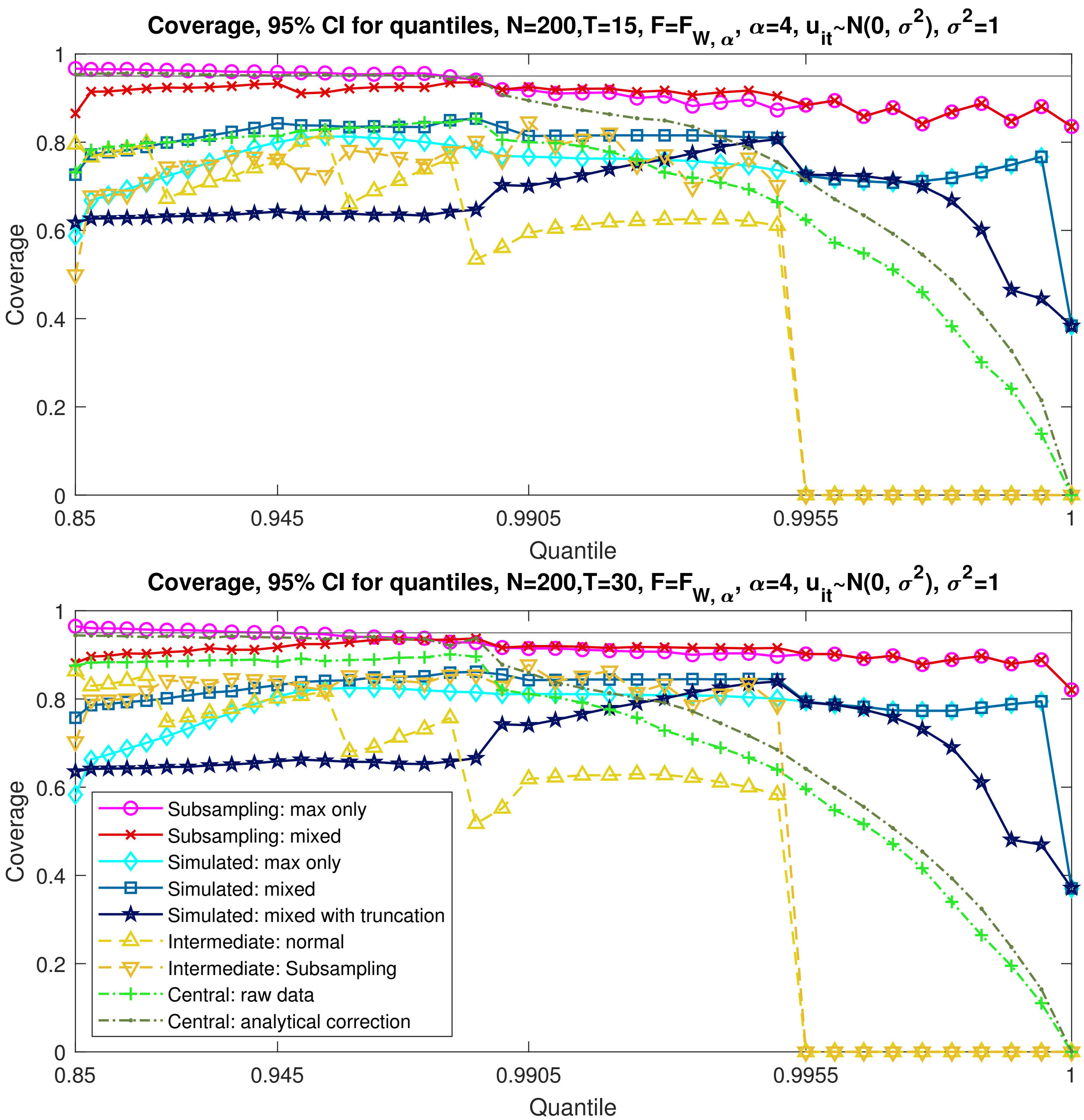


Figure 20: Coverages for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$

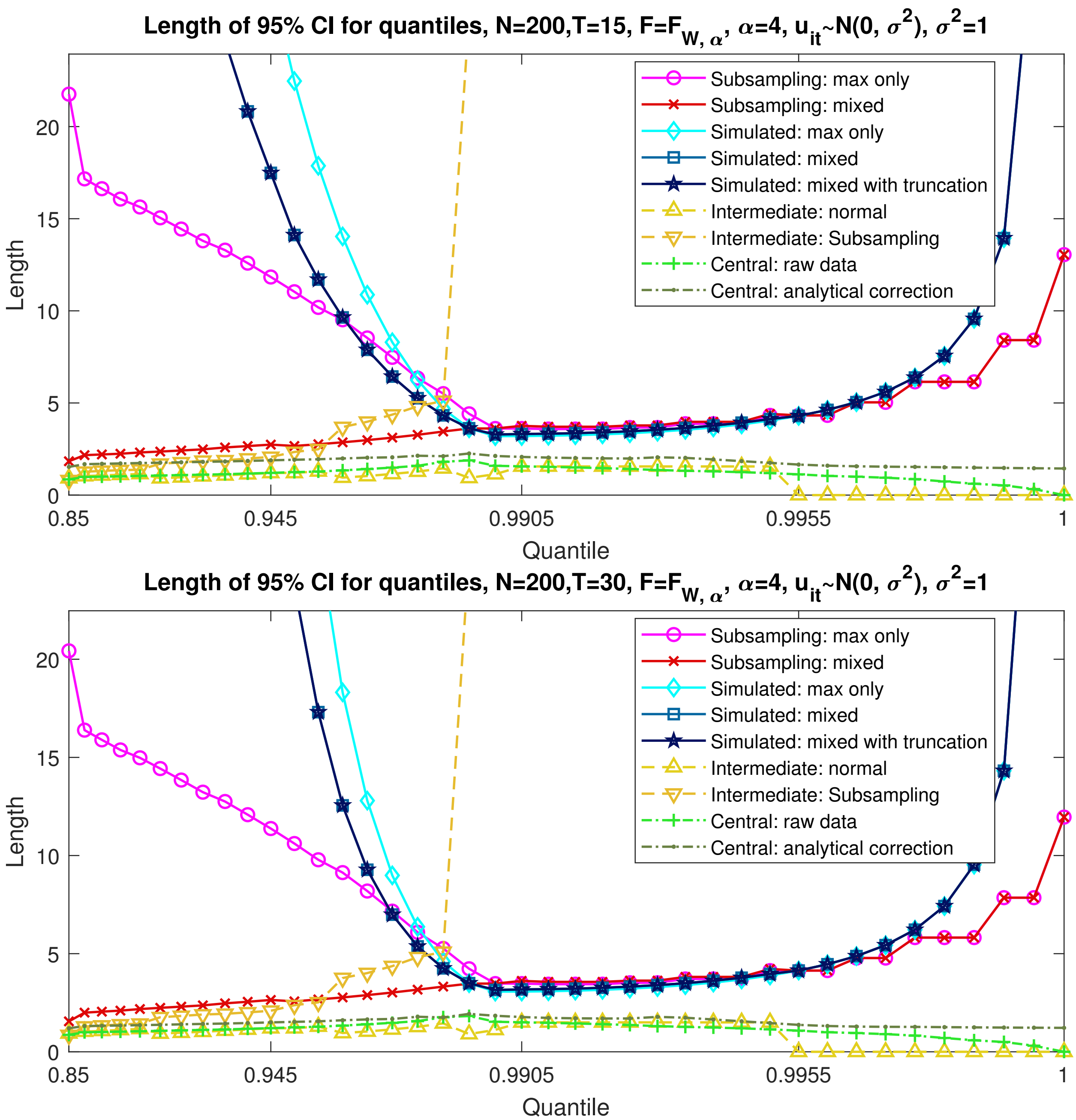


Figure 21: Confidence interval length for different approximations, $N = 200, T = 15, 30$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$

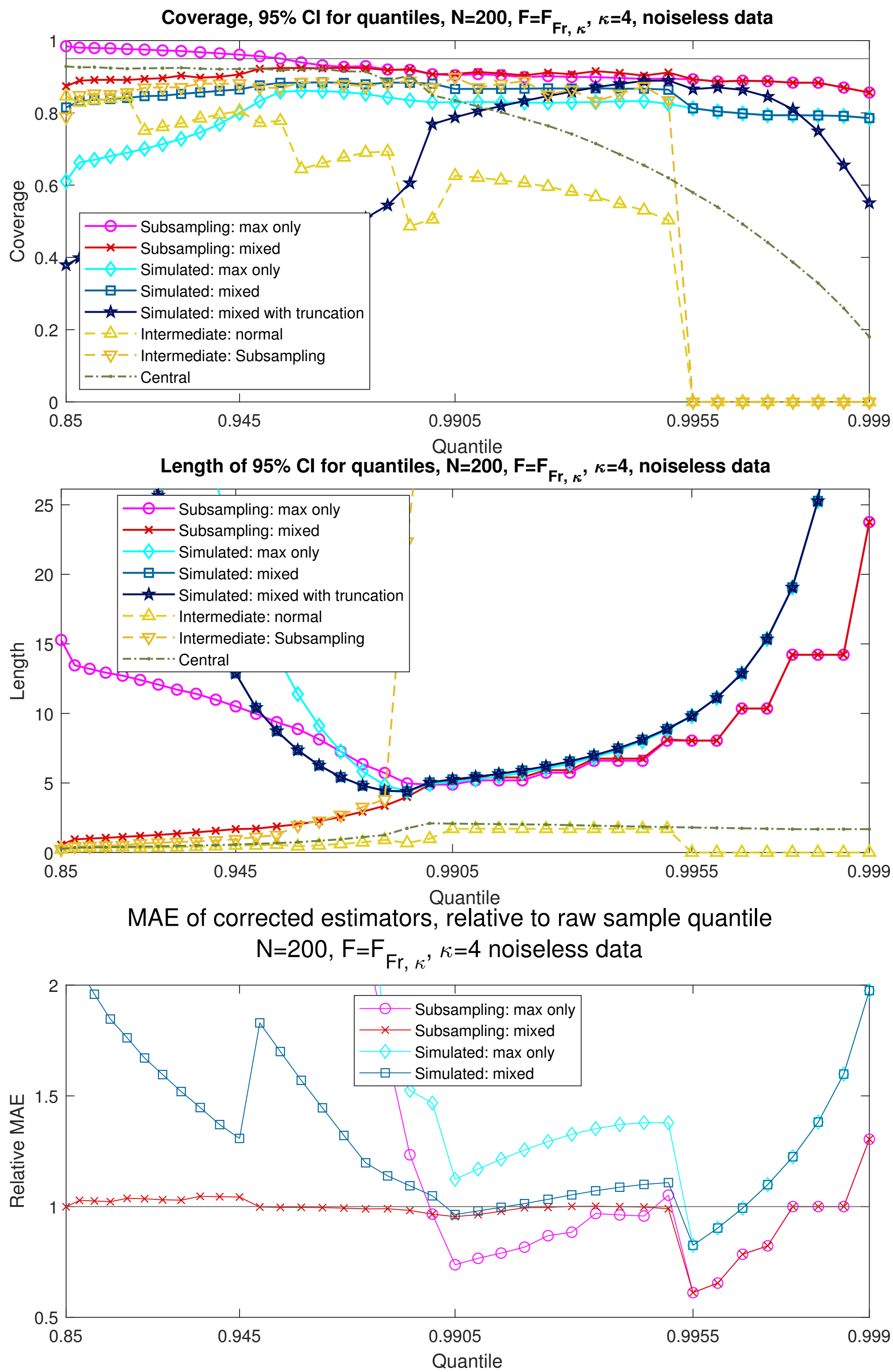


Figure 22: Noiseless data: confidence interval coverage, length, and performance of corrected quantile estimators for different approximations, $N = 200$, $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$.

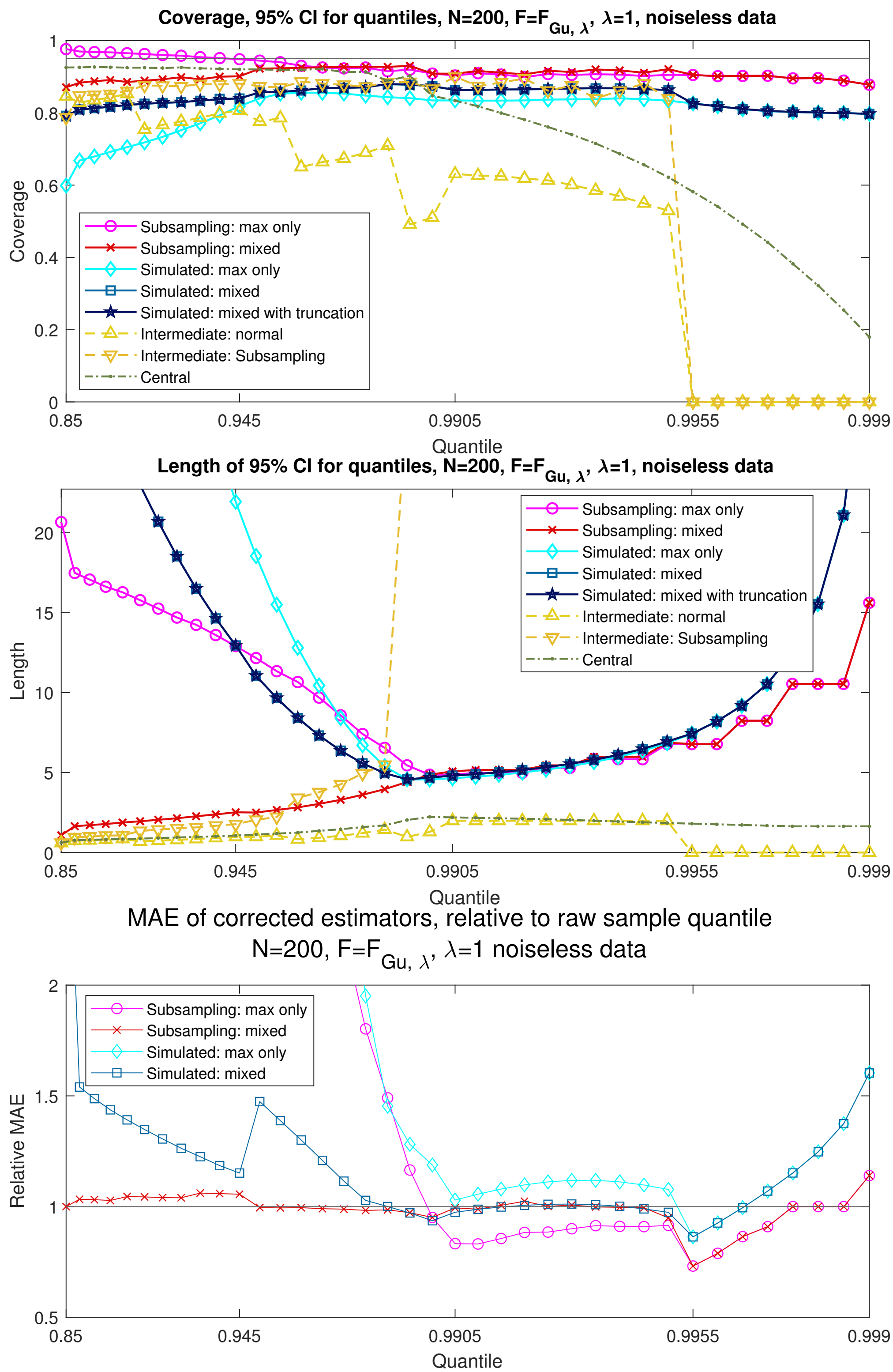


Figure 23: Noiseless data: confidence interval coverage, length, and performance of corrected quantile estimators for different approximations, $N = 200$, $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$.

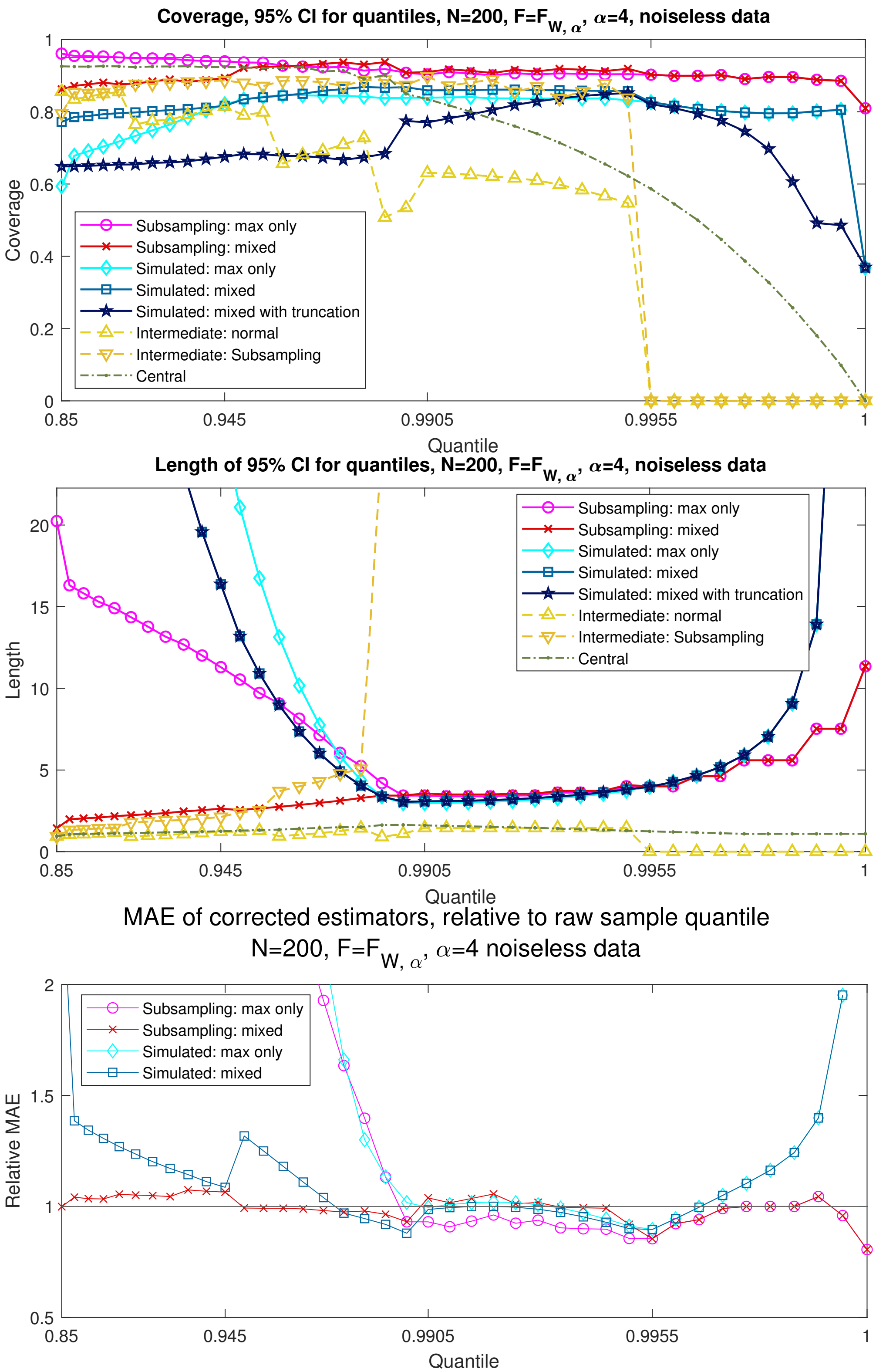


Figure 24: Noiseless data: confidence interval coverage, length, and performance of corrected quantile estimators for different approximations, $N = 200$, $\theta \sim F_{W,\alpha}$, $\alpha = 4$.