Inference on Extreme Quantiles of Unobserved Individual Heterogeneity
Online Appendix

Vladislav Morozov*

June 16, 2023

This supplementary online appendix provides additional theoretical, numerical, and empirical results. Section OA.1 illustrates how the setting and the results of the paper can be extended to cover unbalanced data. In section OA.2 we compute rate conditions of theorems 3.1 and 3.3 for several general examples of distributions of θ and the estimation noise. These examples serve to further motivate the rate conditions of propositions 3.2 and 3.4. In section OA.3, we provide a deterministic version of conditions (2) and (3) of theorem 3.3. Additional simulation results are presented in section OA.4. We expand on the simulation study of section 5 by considering additional distributions for θ and the idiosyncratic disturbances. In all cases, the results closely match those reported in the main text. Additionally, we consider performance of several corrected estimators for extreme quantiles. We find that the median-unbiased estimator of example 4 with subsampled critical values offers improvements over the raw sample quantile. The extrapolation estimator also provides improved performance, but is less robust to sign of γ and the quantile considered. Finally, in section OA.5 we provide additional results for our empirical application. We discuss in detail estimation of the EV index γ and show robustness of our results by examining all the methods for constructing confidence intervals considered in the paper.

*vladislav.morozov@upf.edu; 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, André Souza, Andrea Sy, 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, the 2022 BSE Summer Forum, the 27th International Panel Data Conference, the 10th ICEEE, and the 3rd CESC for comments and discussions.

Contents

OA.1 Unbalanced Data 3
OA.2 Example Rate Conditions 3
  OA.2.1 Examples of Rate Conditions for EVT (Theorem 3.1) 4
  OA.2.2 Examples of Rate Conditions for IVT (Theorem 3.3) 5
  OA.2.3 Verification of Examples 5
OA.3 Deterministic Conditions For Intermediate Order Tail Equivalence Conditions 17
OA.4 Additional Simulation Results 19
  OA.4.1 Additional Simulations for Confidence Intervals 19
  OA.4.2 Corrected Estimators For Quantiles 39
OA.5 Supplemental Materials for the Empirical Application 49
  OA.5.1 Estimation of the EV Index 49
  OA.5.2 Equivalence of Tails up to Mean and Variance 51
  OA.5.3 Confidence Intervals for Extreme Quantiles of Productivity 56

OA.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

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

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 (OA.1.1), we represent eq. (OA.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. (OA.1.1).

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 3.2 and 3.4 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).

Remark 1. 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.

OA.2 Example Rate Conditions

To build intuition and motivate the form of the sufficient rate conditions of propositions 3.2 and 3.4, we 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 1 and 3 below only by a slowly varying function. Example 2 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, reflecting the assumptions of proposition 3.2.

We remark that we use the setting of example 1 in our simulation study in the main text. Similar simulation results for the settings of examples 2-3 are provided in section OA.4 of this online appendix.

OA.2.1 Examples of Rate Conditions for EVT (Theorem 3.1)

We consider two specifications for G_T. First, let $G_{\beta,T}$ be defined by its density $g_\beta(x) = (\beta/2)(1 + |x - \mu_T|)^{-\beta-1}, \beta > 0, x \in \mathbb{R}$ 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.

Example 1 ($\gamma > 0$)**.** 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 $\gamma = 1/\kappa > 0$. A convenient choice of a_N and b_N is $a_N = N^{1/\kappa} - 1, b_N = 0$. With this choice $\theta_{N,N}/(N^{1/\kappa} - 1)$ is asymptotically distributed as a Fréchet random variable.

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 $G_T = G_{normal,T}$, the condition trivializes to the requirement that $\sqrt{\log(N)}/(N^{1/\kappa}T^p) \to 0$, which always holds. We remark that the conditions hold independently of choice of (a_N, b_N) as long as $(\theta_{N,N} - b_N)/a_N$ converges to a non-degenerate random variable. This result follows as in the proof of proposition 3.2.

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 2 ($\gamma = 0$)**.** Let $F = F_{Gu,\lambda}$ be the exponential distribution with parameter λ. $F_{Gu,\lambda}$ satisfies assumption 2 with $\gamma = 0$. If $a_N = 1, b_N = \log N/\lambda$, then $(\theta_{N,N} - \log N/\lambda)$ 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 1, 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 3 ($\gamma < 0$)**.** 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 $\gamma = -1/\alpha < 0$. If $a_N = \theta_F/N^{1/\alpha}, b_N = \theta_F$, then $N^{1/\alpha}(\theta_{N,N} - \theta_F)/\theta_F$ 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 1 and 2. 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 1-3 can be written in common form using the EV index γ. If $G_T = G_{\beta,T}$, all conditions can be written as $N^{1/\beta-\gamma}(\log(T))^{1/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, the conditions can be equivalently written as $N^{-\gamma}\sqrt{\log(N)}/T^p \to 0$. Note the similarity to the rate conditions of proposition 3.2.

OA.2.2 Examples of Rate Conditions for IVT (Theorem 3.3)

The distributions of examples 1-3 satisfy assumption 4, and hence theorem 3.3 can be applied.

Example 5 ($\gamma > 0$, continued)**.** The normalizing constants of theorem 3.3 are $F^{-1}_{Fr,\kappa}(1 - k/N) = (1 - k/N)\,(N/k)^{1/\kappa} - 1$ and $c_N = \kappa^{-1}\,(N/k)^{1/\kappa}$. 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: $F^{-1}_{Gu,\lambda}\,(1 - k/N) = \log(N/k)/\lambda$ 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 $F^{-1}_{W,\alpha}\,(1 - k/N) = \theta_F - \theta_F\,(k/N)^{1/\alpha}$ and $c_N = (\theta_F/\alpha)\,(k/N)^{1/\alpha}$. 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 examples can be written in a common form:

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

where ν is any positive number. Again note the similar to the rate conditions of proposition 3.4.

OA.2.3 Verification of Examples

In this section we provide a detailed verification of the claims made in sections OA.2.1-OA.2.2.

Useful Expressions

For easy reference, we collect all the expressions that we use in verifying the examples.

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

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

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 (see eq. (A.1) in the proof appendix):

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

Verification of Conditions (TE-Inf) and (TE-Sup) of Theorem 3.1

Here we provide the details for examples 1-3 for conditions of theorem 3.1.

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

$$S_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}\left(1 - \frac{1}{N\tau} + u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right) + \frac{1}{T^p}G_T^{-1}(1-u)\right), \quad \text{(OA.2.1)}$$
$$s_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}\left(1 - \frac{1}{N\tau} - u\right) - F^{-1}\left(1 - \frac{1}{N\tau}\right) + \frac{1}{T^p}G_T^{-1}(u)\right).$$

We follow the same strategy for all three examples. The approach is similar to that of the proof of proposition 3.2. First, we construct a sequence $u_{S,\tau,N,T} \in [0, 1/N\tau]$ such that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$ under certain conditions on N and T. Since $\inf_{u\in[0,1/N\tau]} S_{\tau,N,T}(u) \leq S_{\tau,N,T}(u_{S,\tau,N,T})$, we obtain that

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

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

$$\liminf_{N,T\to\infty} \sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \geq 0.$$

Last, in all cases $a_N > 0$, hence by lemma A.3 it eventually holds that

$$\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \leq \sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \leq \inf_{u\in[0,1/N\tau]} S_{\tau,N,T}(u).$$

This implies that

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

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

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

The above holds for any $\tau \in (0, \infty)$, and so the conditions (TE-Inf) and (TE-Sup) of theorem 3.1 hold.

Example 1, page 4 Let $\theta \sim F_{Fr,\kappa}$. Take $a_N = N^{1/\kappa} - 1$, $b_N = 0$.

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

$$u_{S,\tau,N,T} = \frac{1}{N\tau}\frac{1}{\log(T)+1} \in \left[0, 1 - \frac{1}{N\tau}\right]. \quad \text{(OA.2.2)}$$

We show that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$, where $S_{\tau,N,T}$ is defined in eq. (OA.2.1).

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 OA.2.3, we obtain that

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

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 OA.2.3, we obtain that

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

Since μ_T is a bounded sequence, we conclude that for $\limsup_{N,T\to\infty}\inf_{u\in[0,\epsilon]}S_{\tau,N,T}(u)$ to be equal to zero, it is sufficient that

$$\frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p}\to 0. \tag{OA.2.3}$$

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

$$u_{s,\tau,N,T}=\frac{1}{N\tau}\frac{1}{\log(T)}. \tag{OA.2.4}$$

Observer that $u_{s,\tau,N,T}$ eventually lies in $[0,\epsilon]$ for any $\epsilon \in (0,1)$ 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

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

Exactly as above, we conclude that for $\liminf_{N,T\to\infty}\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u)$ 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

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

which again matches the condition derived for the infimum.

Finally, since $a_N > 0$, lemma A.3 implies that

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

If the above rate conditions on N and T hold, it holds that $\liminf_{N,T\to\infty}\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) = 0, \limsup_{N,T\to\infty}\inf_{u\in\left[0,\frac{1}{N\tau}\right]} S_{\tau,N,T}(u) = 0$. 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. (OA.2.2) and (OA.2.4)) 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 $u_{S,\tau,N,T} = c/(N\tau)$ where $c < 1$ is fixed constant. Then

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

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

Example 2, page 4 Let θ_i be exponential(λ). Let $a_N = 1$. Consider $S_{\tau,N,T}$ and pick $u_{S,\tau,N,T}$ as in eq. (OA.2.2). 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 3, page 4 Let $\theta \sim F_{W,\alpha}$ and let $a_N^{-1} = N^{1/\alpha}/\theta_F$. Consider $S_{\tau,N,T}$ and let $u_{S,\tau,N,T}$ be as in eq. (OA.2.2). First examine the $F^{-1}_{W,\alpha}$ terms. Using the expressions for inverses given in section

OA.2.3, we obtain

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

hence

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

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

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

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

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

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.

Intermediate Order Statistics, Examples 5, 6, 7, Page 5

Example cdfs satisfy assumption 4 First we establish that cdfs F of examples 5-7 satisfy assumption 4, hence theorem 3.3 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 for theorem 3.1. Define $\tilde{s}_{N,T}$ and $\tilde{S}_{N,T}$ similarly to eq. (OA.2.1):

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

where c_N is as in theorem 3.3.

We construct a sequence $\tilde{u}_{S,N,T} \in [0, U_{k,N},]$ such that $\tilde{S}_{N,T}(\tilde{u}_{S,N,T}) \to 0$ under certain conditions on N, T and k. As previously, since $\inf_{u\in[0,U_{k,N}]} \tilde{S}_{N,T}(u) \leq \tilde{S}_{N,T}(\tilde{u}_{S,N,T})$, we obtain that $\limsup_{N,T\to\infty} \inf_{u\in[0,U_{k,N}]} \tilde{S}_{N,T}(u) \leq 0$. A similar argument can be applied to $\tilde{s}_{N,T}$ by picking a point $\tilde{u}_{s,N,T}$ that lies in $[0,\epsilon]$ with probability approaching 1 (wpa1) to conclude that wpa1 $\liminf_{N,T\to\infty} \sup_{u\in[0,\epsilon]} \tilde{s}_{N,T}(u) \geq 0$. Proceeding as above, we conclude that conditions (2) and (3) of theorem 3.3 hold.

First, we establish the following elementary lemma.

Lemma OA.2.1. *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 $f'(x) = \rho\gamma \left(1 + x^{\rho}\right)^{\gamma-1} x^{\rho-1}$, 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 5 Let $F = F_{Fr,\kappa}$. Compute the normalizing functions of theorem 3.3.

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

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

$$\tilde{u}_{S,N,T} = U_{k,N} \frac{1}{N^{\rho} + 1} \in \left[0, U_{k,N}\right]. \tag{OA.2.5}$$

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 OA.2.3.

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) $(N/k)U_{k,N} \xrightarrow{p} 1$.[1]. Then the $F^{-1}_{Fr,\kappa}$ terms decay if k is chosen such that

$$\sqrt{k}\left[\left(\frac{N^{\rho}+1}{N^{\rho}}\right)^{1/\kappa} - 1\right] \to 0. \tag{OA.2.6}$$

Also observe that the $((N^{\rho}+1)/N^{\rho})^{1/\kappa} - 1 = O(N^{-\rho})$ by lemma OA.2.1. 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

$$\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1 - U_{k,N}\frac{1}{N^{\rho}+1}\right)$$

[1]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 $\{Y_1, \ldots, Y_N\}$, and $P(Y_i \leq y) = 1 - 1/y$ if $y \geq 1$ and zero otherwise. Observe that $U_{k,N} \stackrel{d}{=} 1/Y_{N-k,N}$, the result then follows. See also the proof of theorem 3.3

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

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

$$\frac{k^{1/2+1/\kappa-1/\beta}}{N^{1/\kappa-1/\beta-\rho/\beta} T^p} \to 0. \tag{OA.2.7}$$

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

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

Thus, the scaled G term decays if

$$\frac{k^{1/2+1/\kappa} \sqrt{\log(N)}}{N^{1/\kappa}} \frac{1}{T^p} \to 0. \tag{OA.2.8}$$

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 $\tilde{u}_{s,N,T} = U_{k,N}/N^\rho$. Then conditions (2) and (3) hold by the same argument as above.

Rate conditions if $k = N^\delta$ Conditions (OA.2.6), (OA.2.7), and (OA.2.8) 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 (OA.2.6) transforms to the requirement that $N^{\delta/2-\rho} \to 0$, which holds if $\rho > \delta/2$. Condition (OA.2.7) 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 (OA.2.8) 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

$$\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta-1/\kappa)+\nu/\beta}}{T^p} \to 0. \tag{OA.2.9}$$

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 (OA.2.6). If we instead use $\tilde{u}_{S,N,T} = U_{k,N}/(\log(T)+1)$ (similarly to section (OA.2.3)), then (OA.2.6) 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. (OA.2.5) 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. (OA.2.9) to eq. (OA.2.3)). 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 (OA.2.9) is to condition (OA.2.3). 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 5 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 3.3

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

Pick $\tilde{u}_{S,N,T}$ as in eq. (OA.2.5):

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

Then

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

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

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

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

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

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

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

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

Example 7, page 5 Finally, let $F = F_{W,\alpha}$. We proceed as in the previous two examples.

First compute the normalizing functions of theorem 3.3.

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

Pick $\tilde{u}_{S,N,T}$ as in eq. (OA.2.5):

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

Then

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

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

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

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

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

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

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

OA.3 Deterministic Conditions For Intermediate Order Tail Equivalence Conditions

For completeness, we provide a deterministic sufficient condition for conditions (2) and (3) of theorem 3.3.

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

$$P\left(\left|\frac{N}{k}U_{k,N} - 1\right| \geq \delta_{k,N,T}\right) \to 0. \tag{OA.3.1}$$

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

$$s_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1 - \frac{k}{N}(1+\delta) - u\right) - F^{-1}\left(1 - \frac{k}{N}(1-\delta)\right) + \frac{1}{T^p}G_T^{-1}(u)\right),$$
$$S_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1 - \frac{k}{N}(1-\delta) + u\right) - F^{-1}\left(1 - \frac{k}{N}(1+\delta)\right) + \frac{1}{T^p}G_T^{-1}(1-u)\right) \xrightarrow{p} 0.$$

for all δ, u such that the above functions are well-defined. If for some $\epsilon \in (0,1)$

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

then

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

A sequence $\delta_{k,N,T}$ of eq. (OA.3.1) 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

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

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

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

$\to 0.$

To obtain the above inequality, we decrease the choice set for the inf. In the first F^{-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

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

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.3 also show that for the original random supremum and infimum with probability approaching 1

$$\sup_{u\in[0,\epsilon]}\{\dots\} \leq \sup_{u\in\left[0,1-U_{k,N}\right]}\{\dots\} \leq \inf_{u\in\left[0,U_{k,N}\right]}\{\dots\}.$$

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. □

OA.4 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 OA.4.1 we report the results of an expanded version of the simulation study in the main text, covering more confidence intervals and additional distributions for θ_i and u_{it}. Second, in section OA.4.2 we consider the performance of different estimators for quantiles proposed in section 4.

OA.4.1 Additional Simulations for Confidence Intervals

In this section we report the results of an expanded simulation study for performance of confidence intervals. The setup as in in section 5 of the main text. 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 section OA.2.1). We consider three distributions for parameter of interest θ_i, corresponding to the three distributions considered in examples 1-3 above. As in the main text, we consider two cross-sectional sample sizes: $N=200$ and $N=2000$. For the heavy-tailed distribution $F_{Fr,\kappa}$ for θ we provide results for $T=10$, in line with the main text and the empirical application. For the light-tailed distribution $F_{Gu,\lambda}$ and the finite-tailed distribution $F_{W,\alpha}$ we provide results for $T = 15$.

In addition to the confidence intervals considered in the main text, we also consider several additional intervals based on the feasible EVT (theorem 4.3) in the main text. the required critical values can be obtained by simulating from the limit distribution of theorem 4.3 after plugging in a consistent estimator for γ. We use the estimators of remark 9. For $F_{Fr,\kappa}$ both the Hill and the PWM estimator are consistent. For $F_{Gu,\lambda}$ and $F_{W,\alpha}$ we only consider the PWM estimator.

The full results are plotted on figures 1-18. The results broadly match those presented in the main text, and 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 if the sign of γ is not known. 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. Using simulated critical values based on the PWM estimator is viable for $N = 2000$ if the distribution has an infinite right tail ($\gamma \geq 0$). In this case the corresponding CIs have good coverage properties and are somewhat shorter that the intervals with subsampling-based critical values. If instead $\gamma < 0$, CIs with simulated critical values should not be used; these intervals are uniformly dominated in terms of coverage by those based on subsampling. As in the main text, intermediate order approximations are a viable option in the regions where the corresponding rate conditions hold and the statistic is stable. We recommend subsampling to obtain the required critical values.

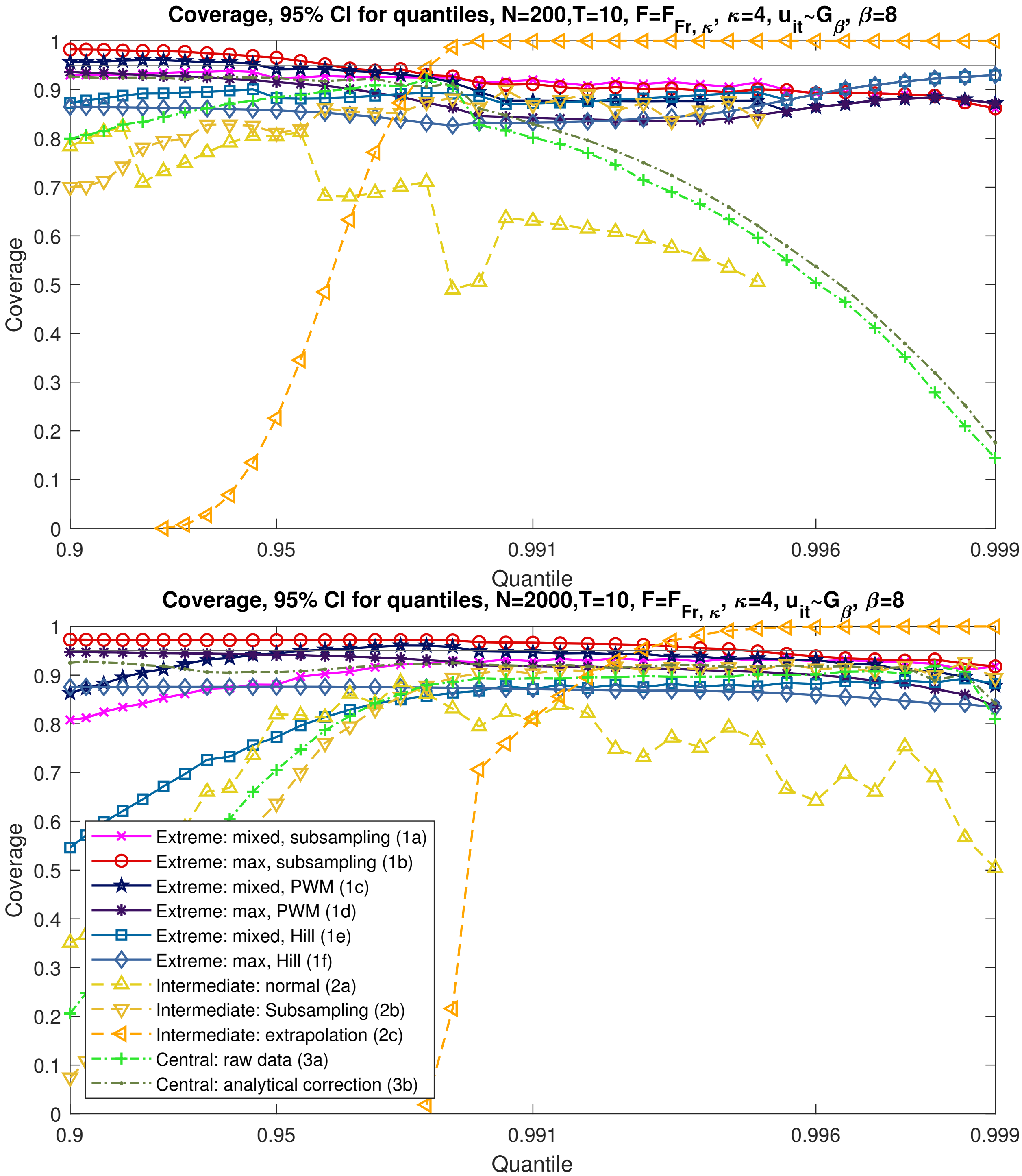


Figure 1: Coverages for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 8$

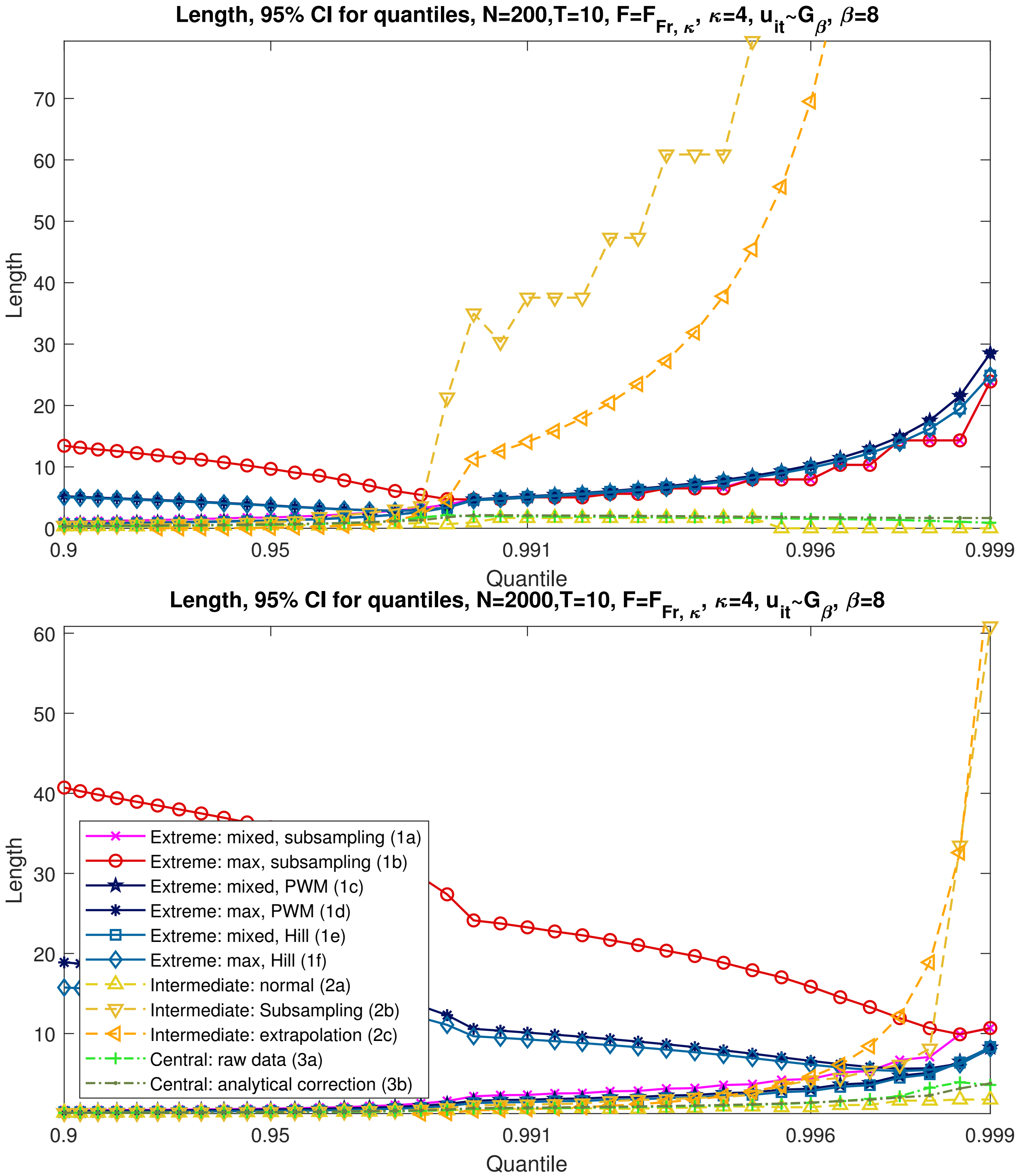


Figure 2: Confidence interval length for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 8$

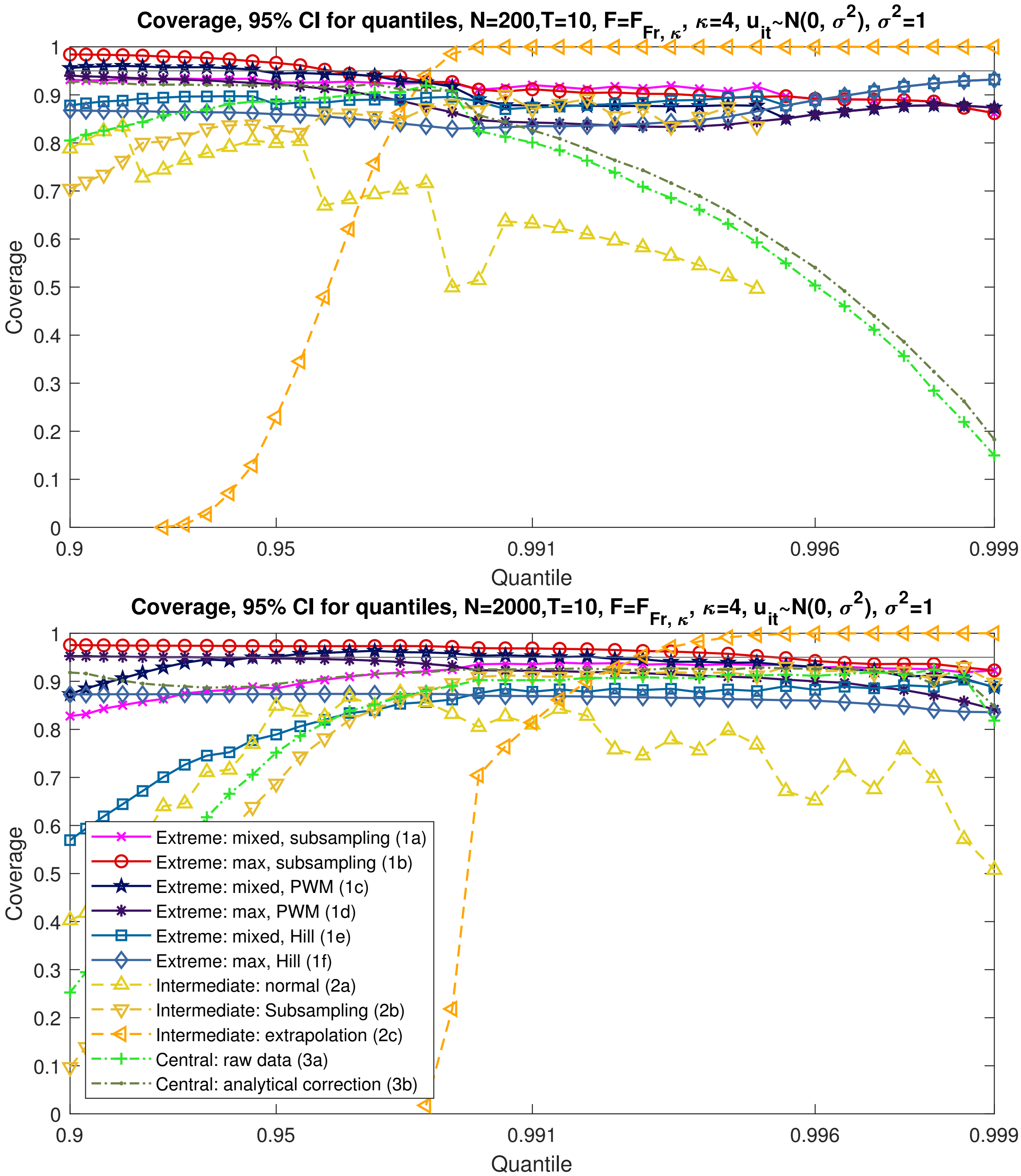


Figure 3: Coverages for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$

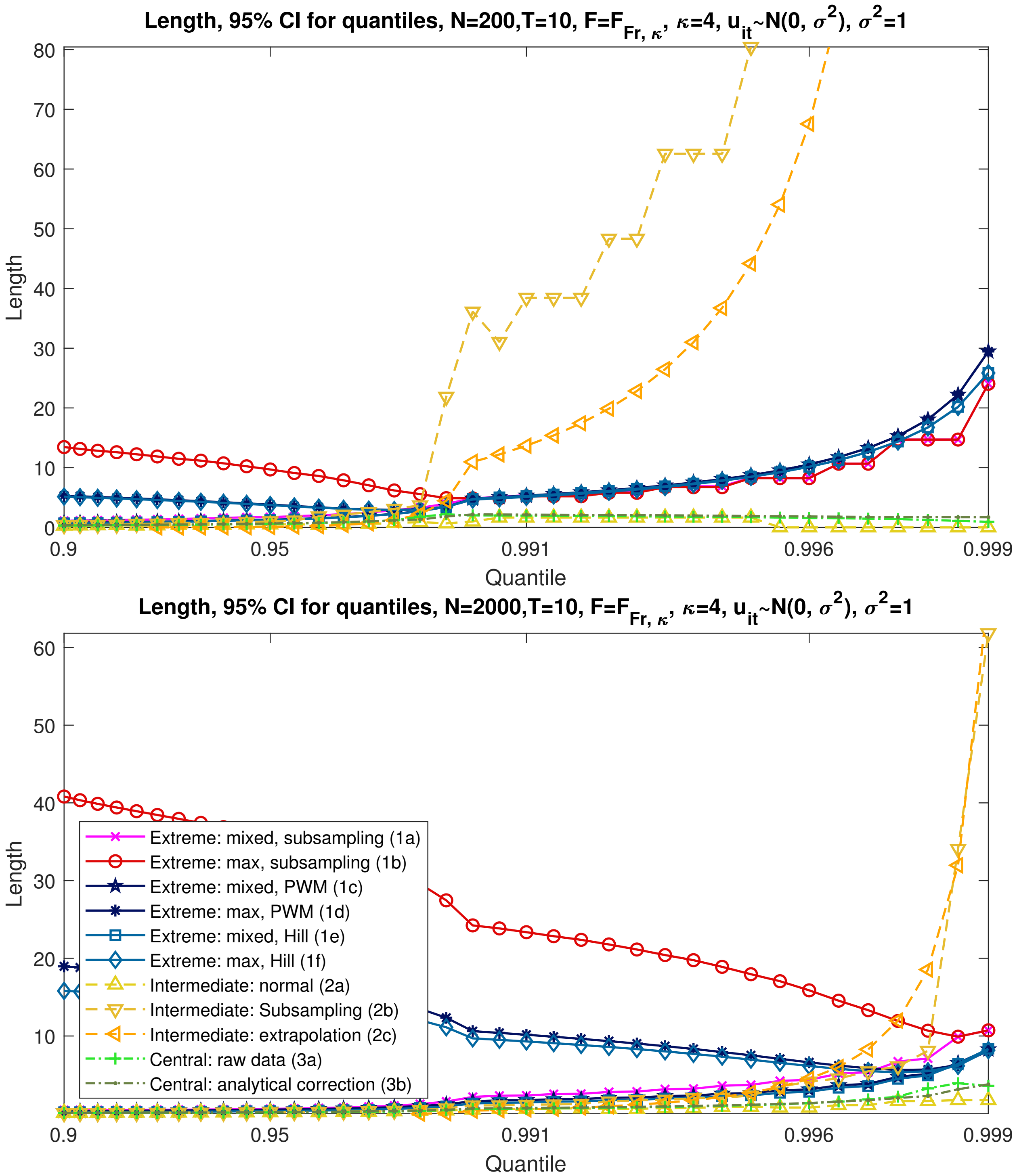


Figure 4: Confidence interval length for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$

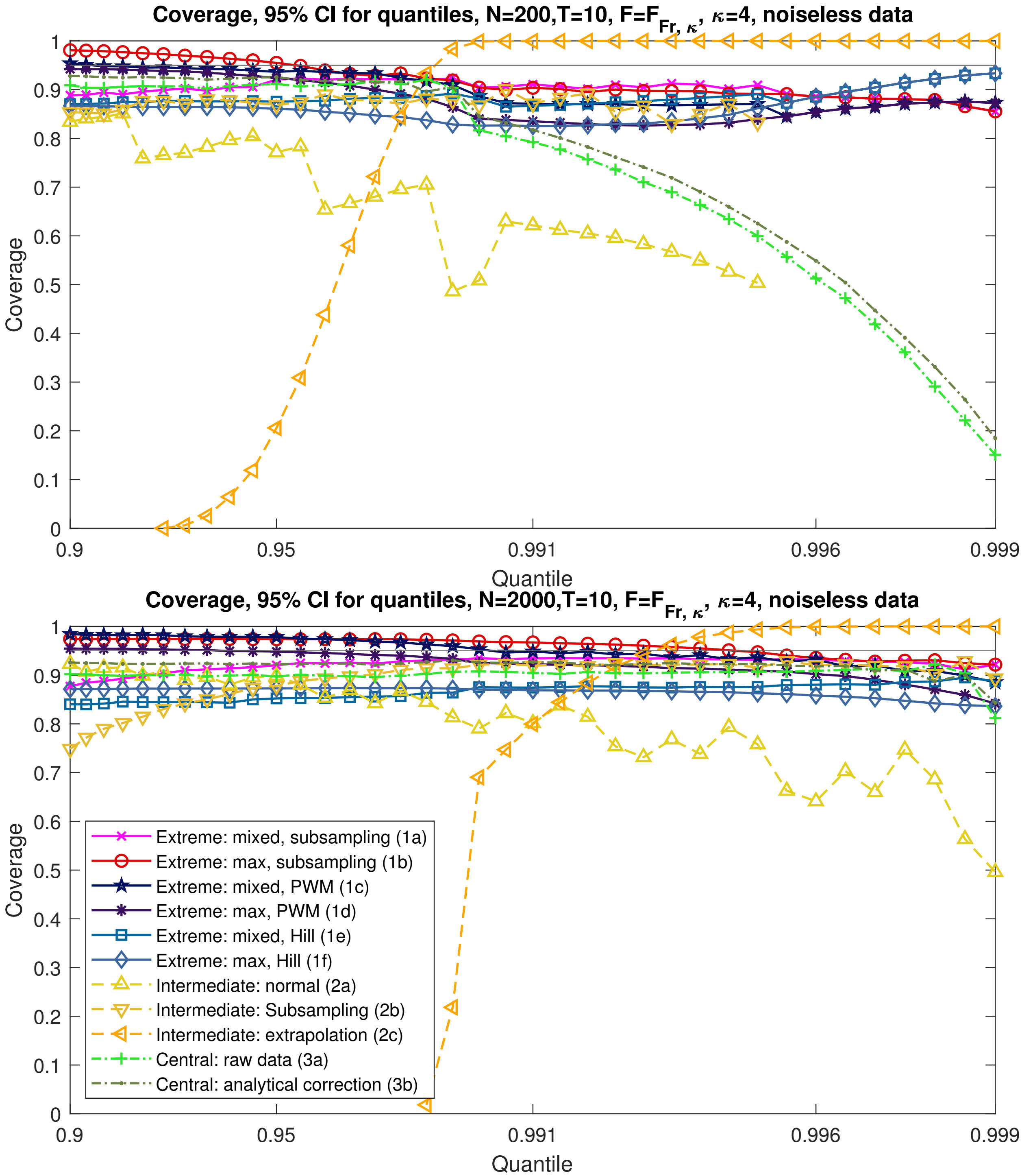


Figure 5: Coverages for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, noiseless data

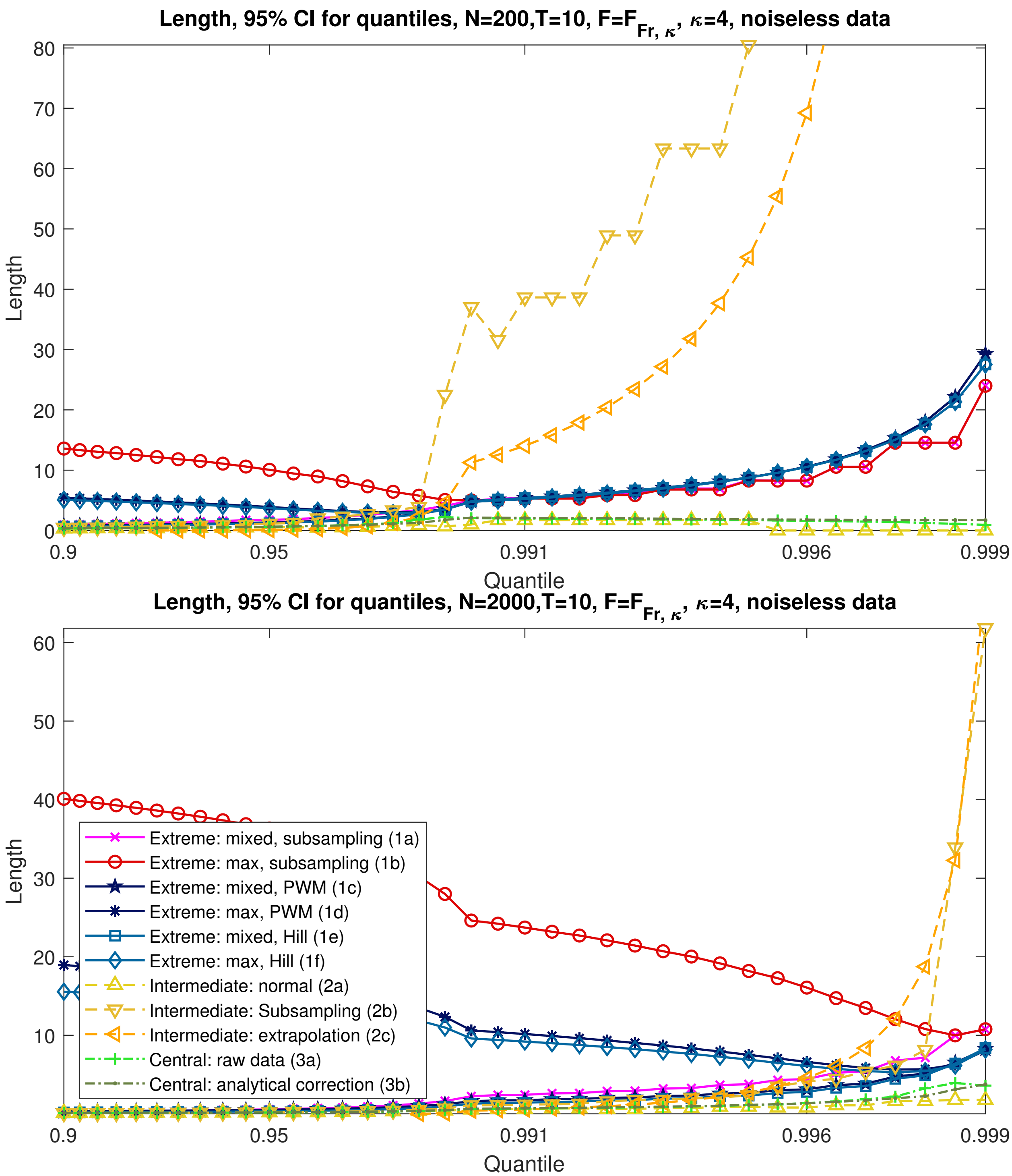


Figure 6: Confidence interval length for different approximations, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, noiseless data

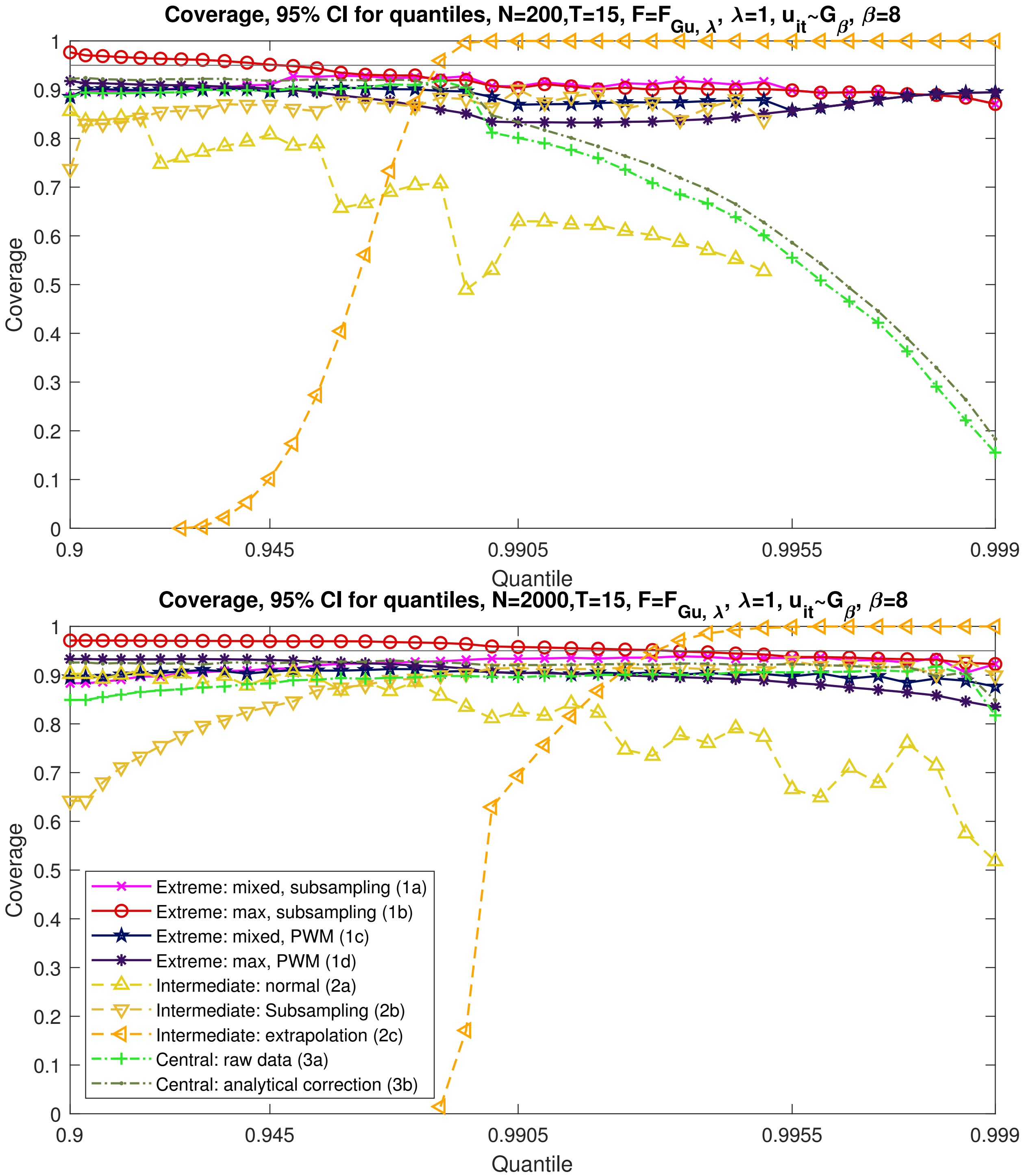


Figure 7: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 8$

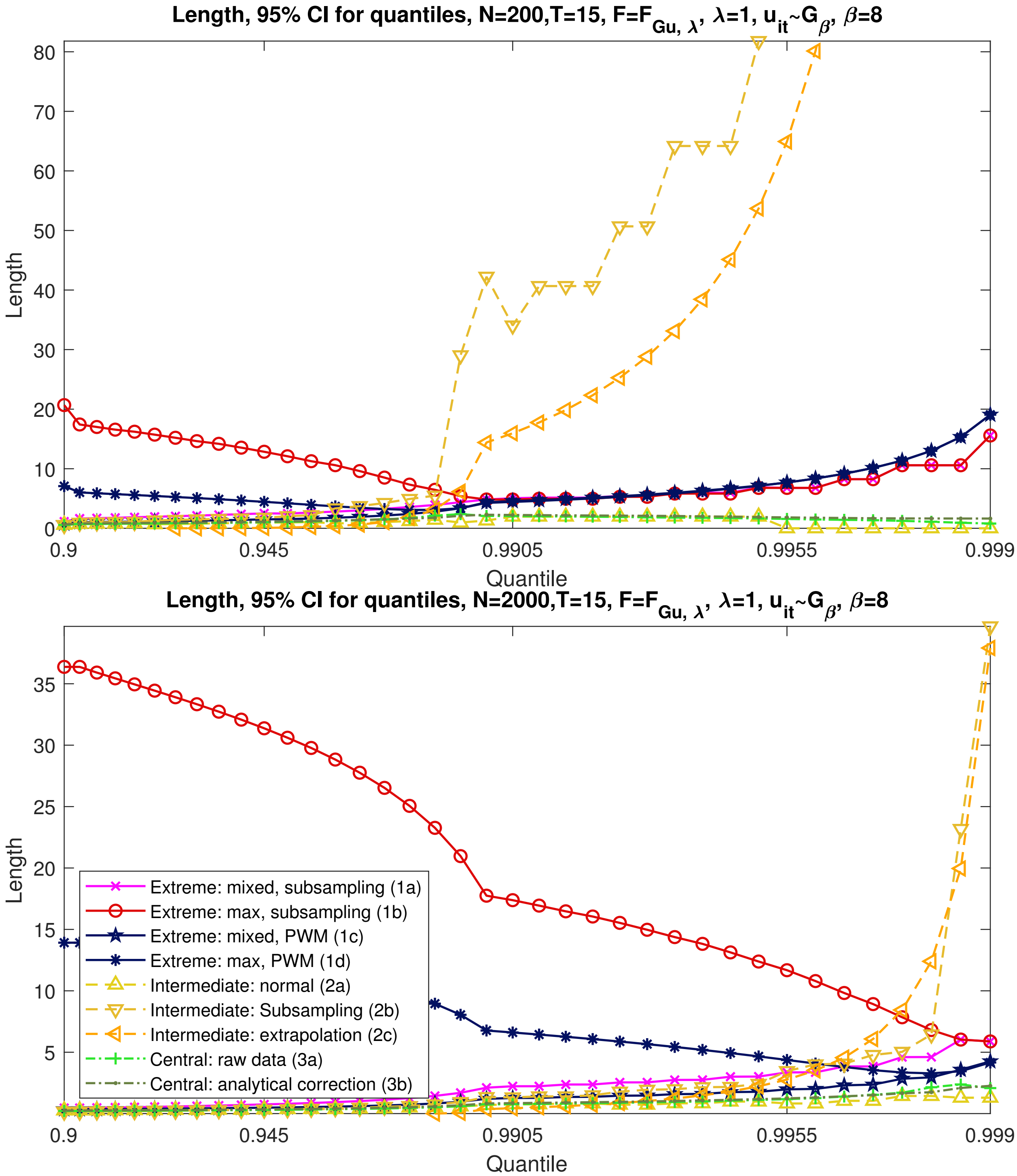


Figure 8: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 8$

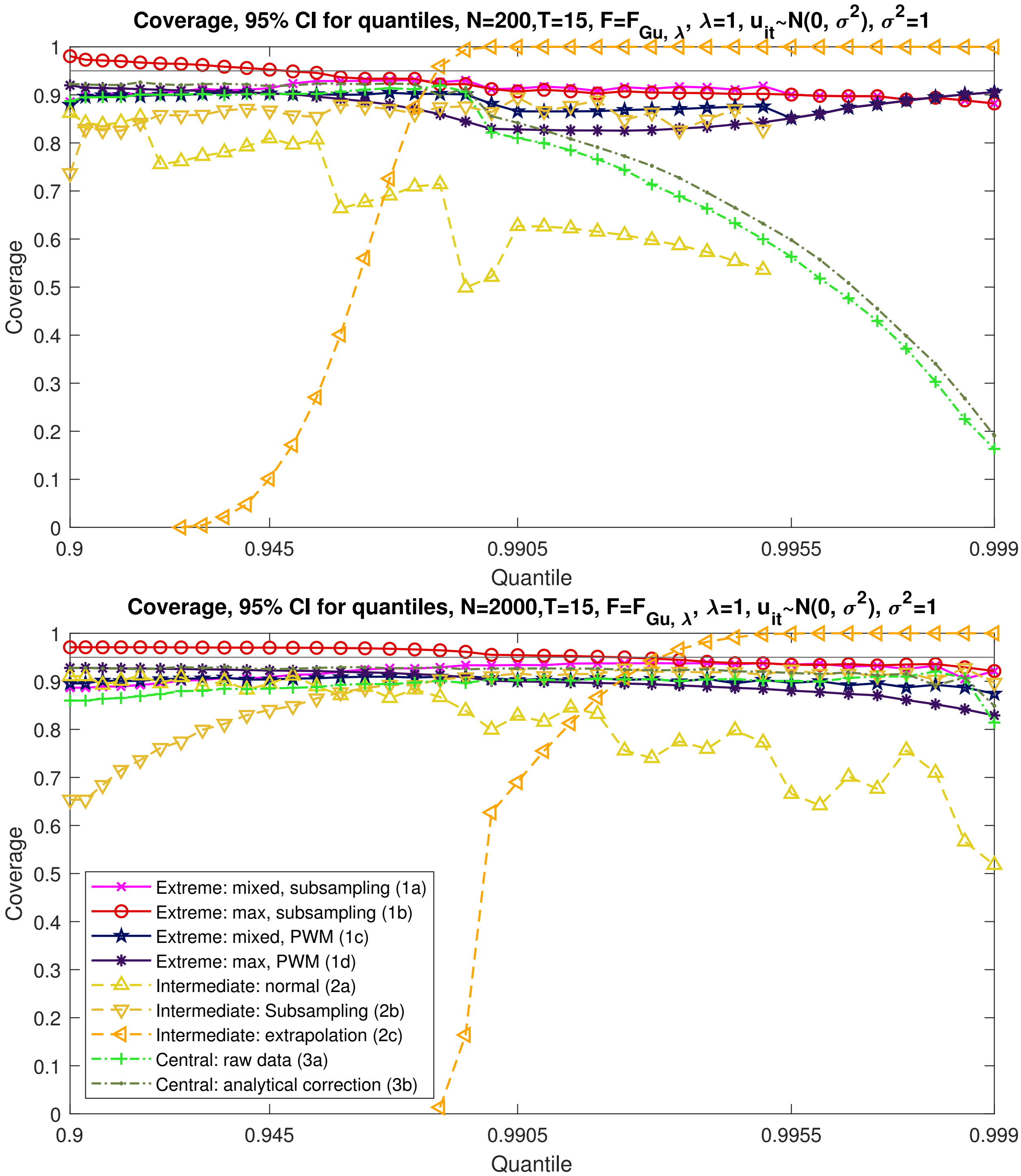


Figure 9: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim N(0, 1)$

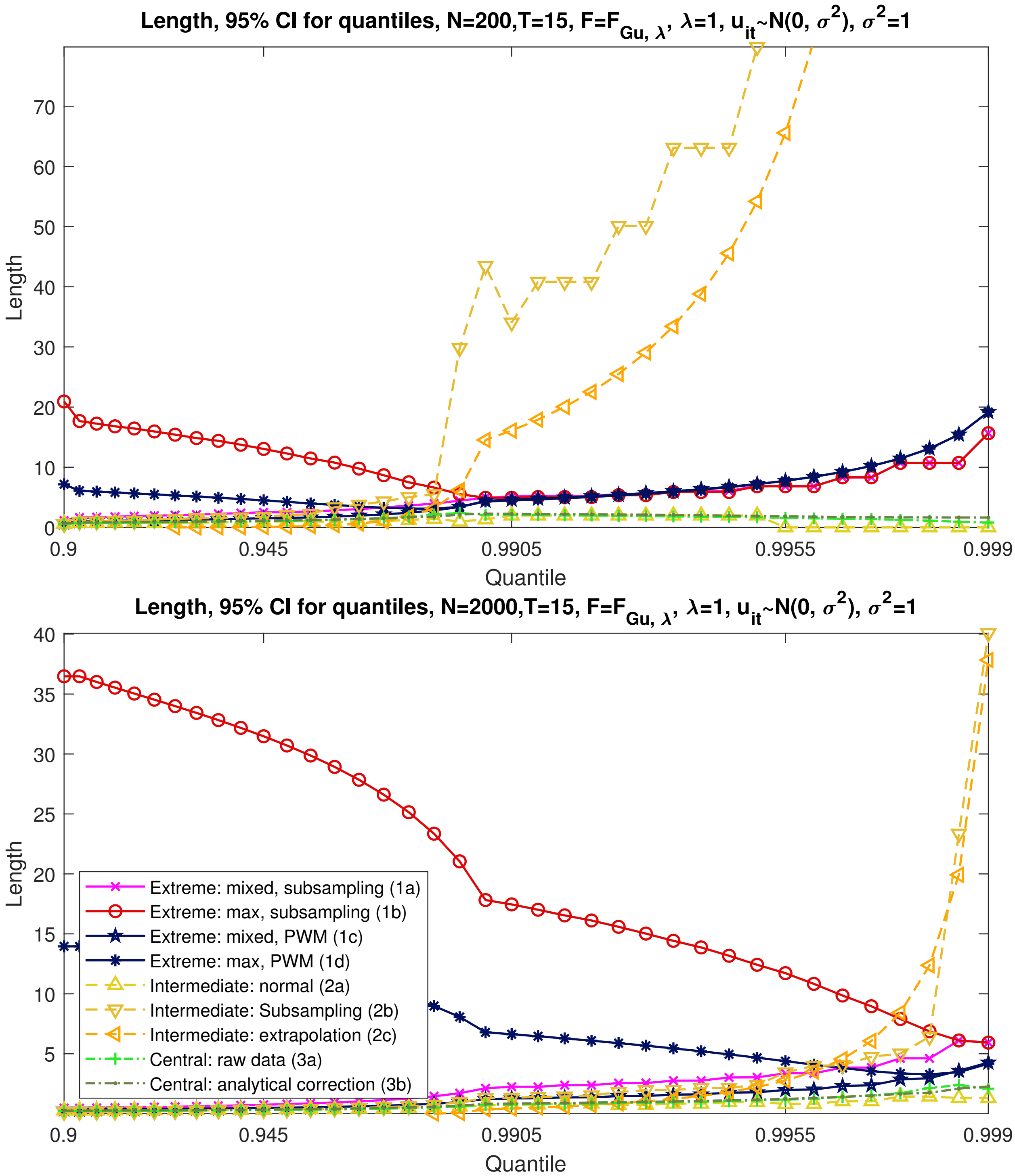


Figure 10: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim N(0, 1)$

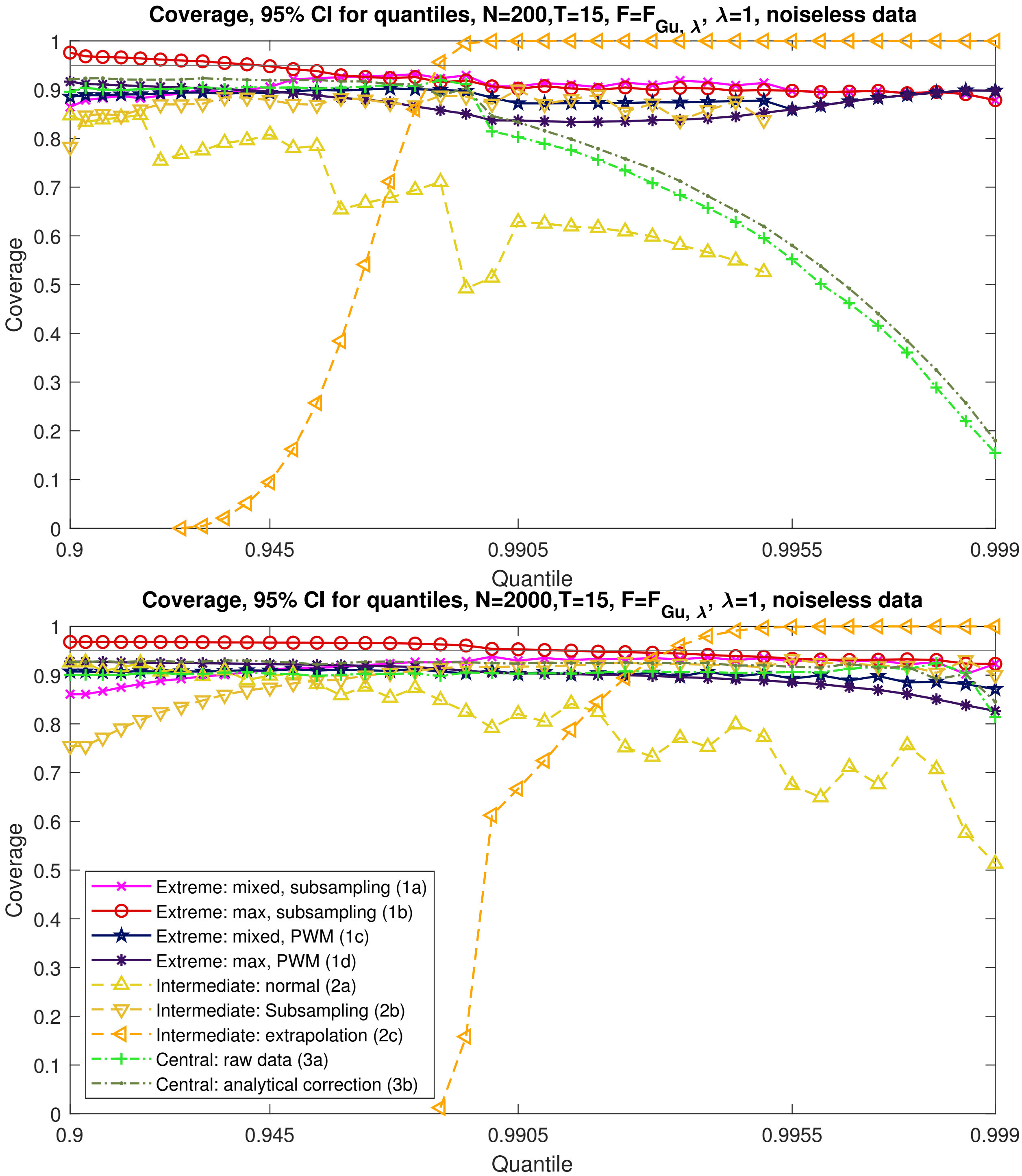


Figure 11: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, noiseless data

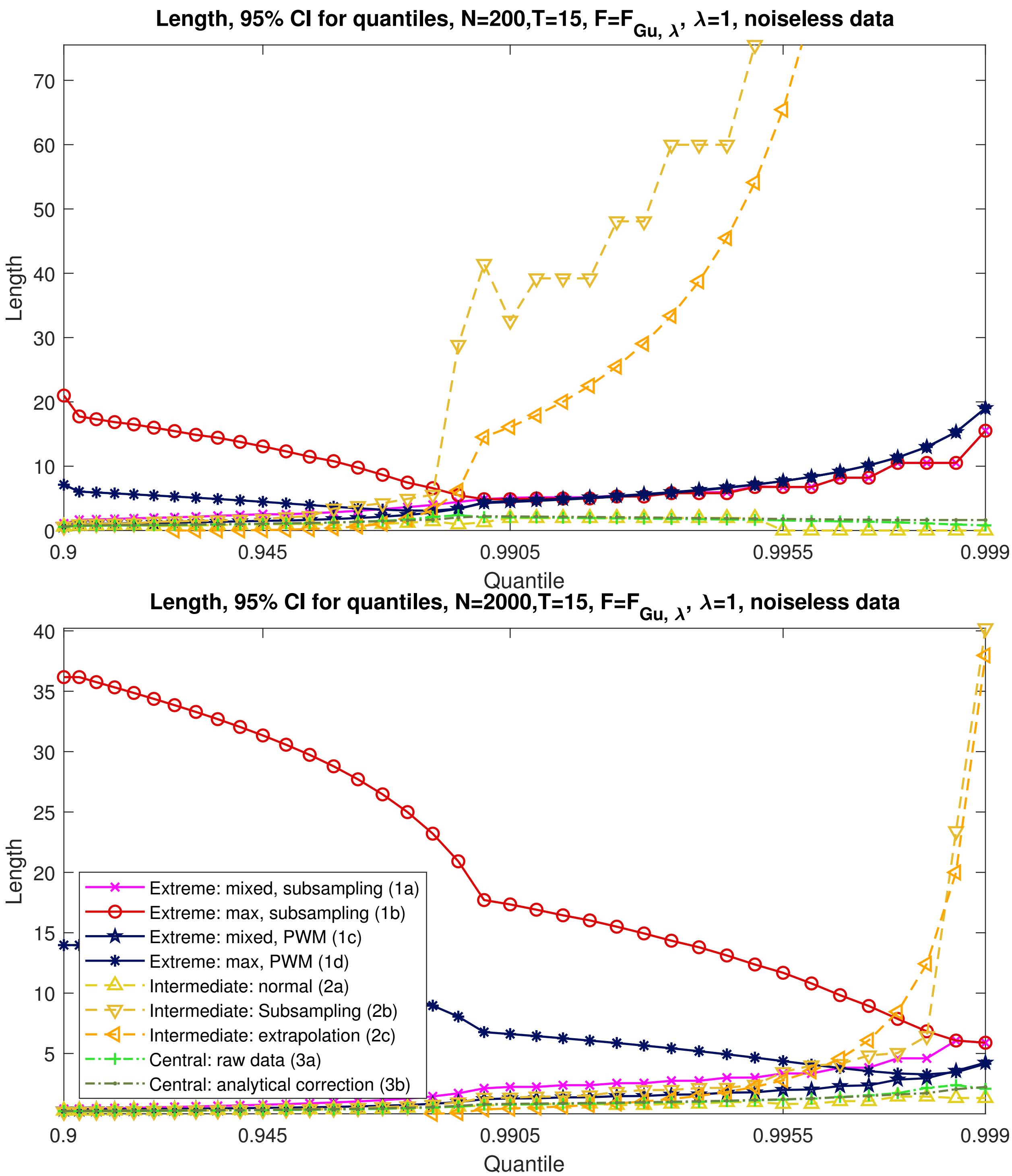
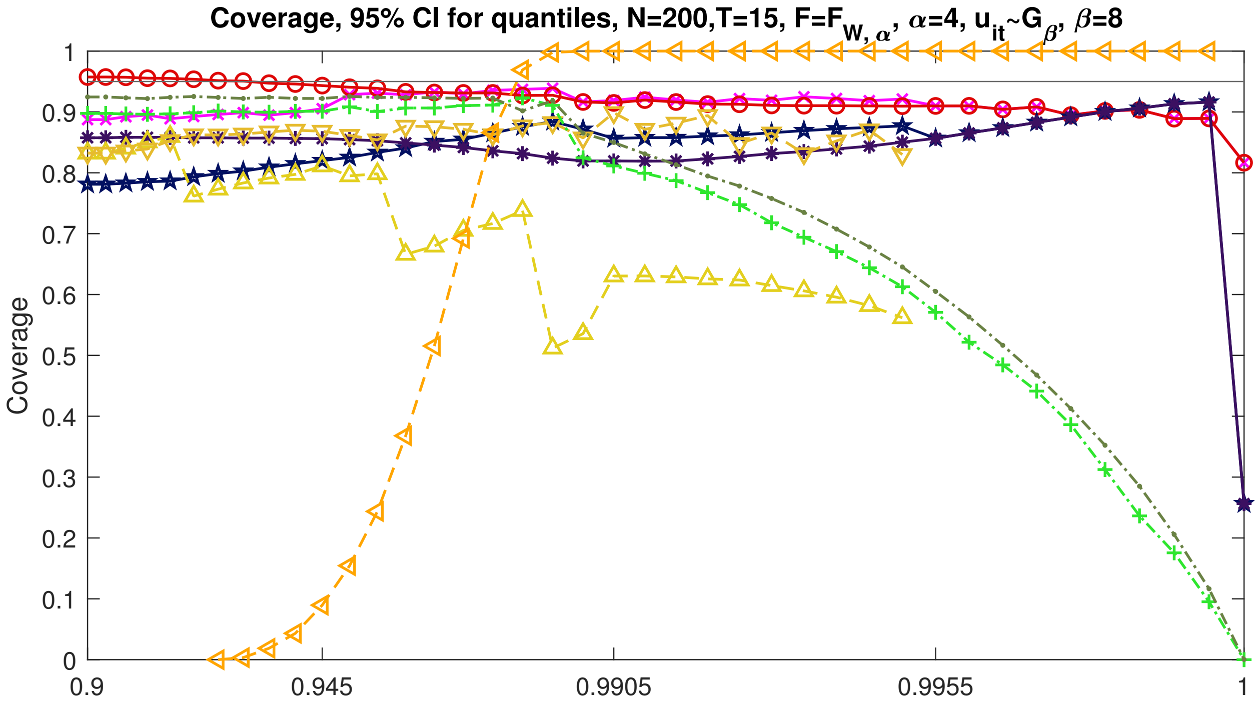


Figure 12: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, noiseless data



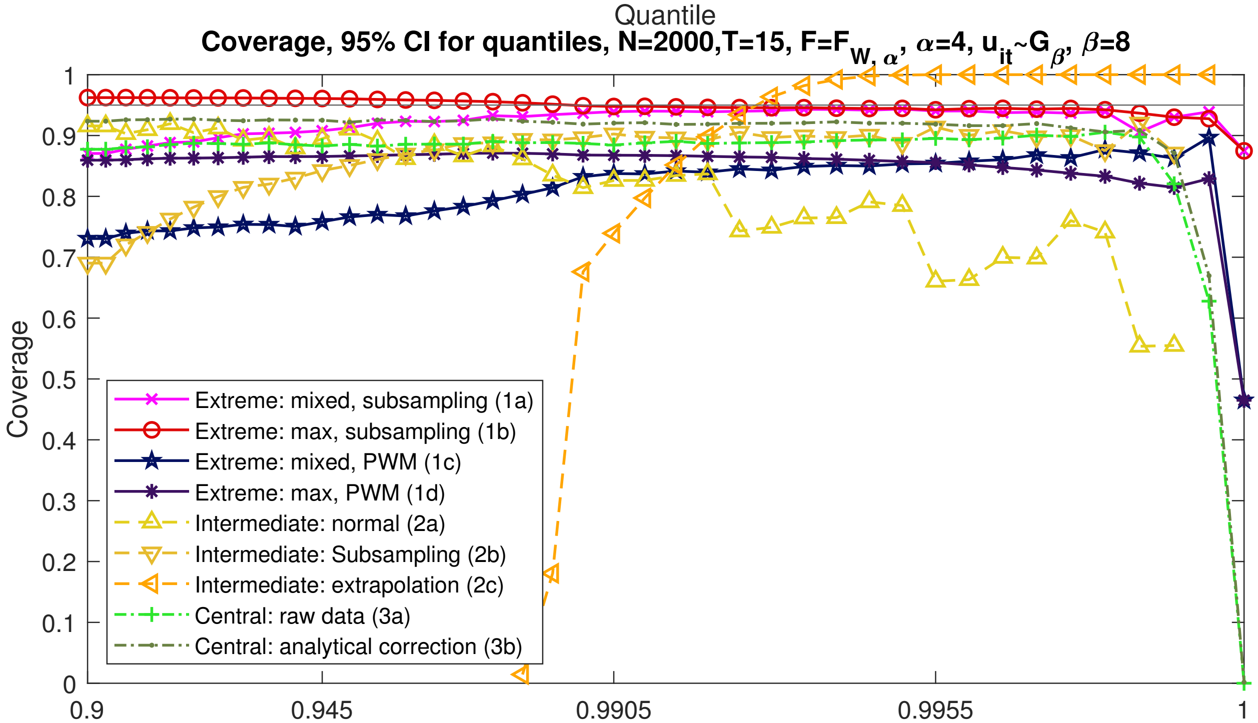


Figure 13: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_{\beta}, \beta = 8$

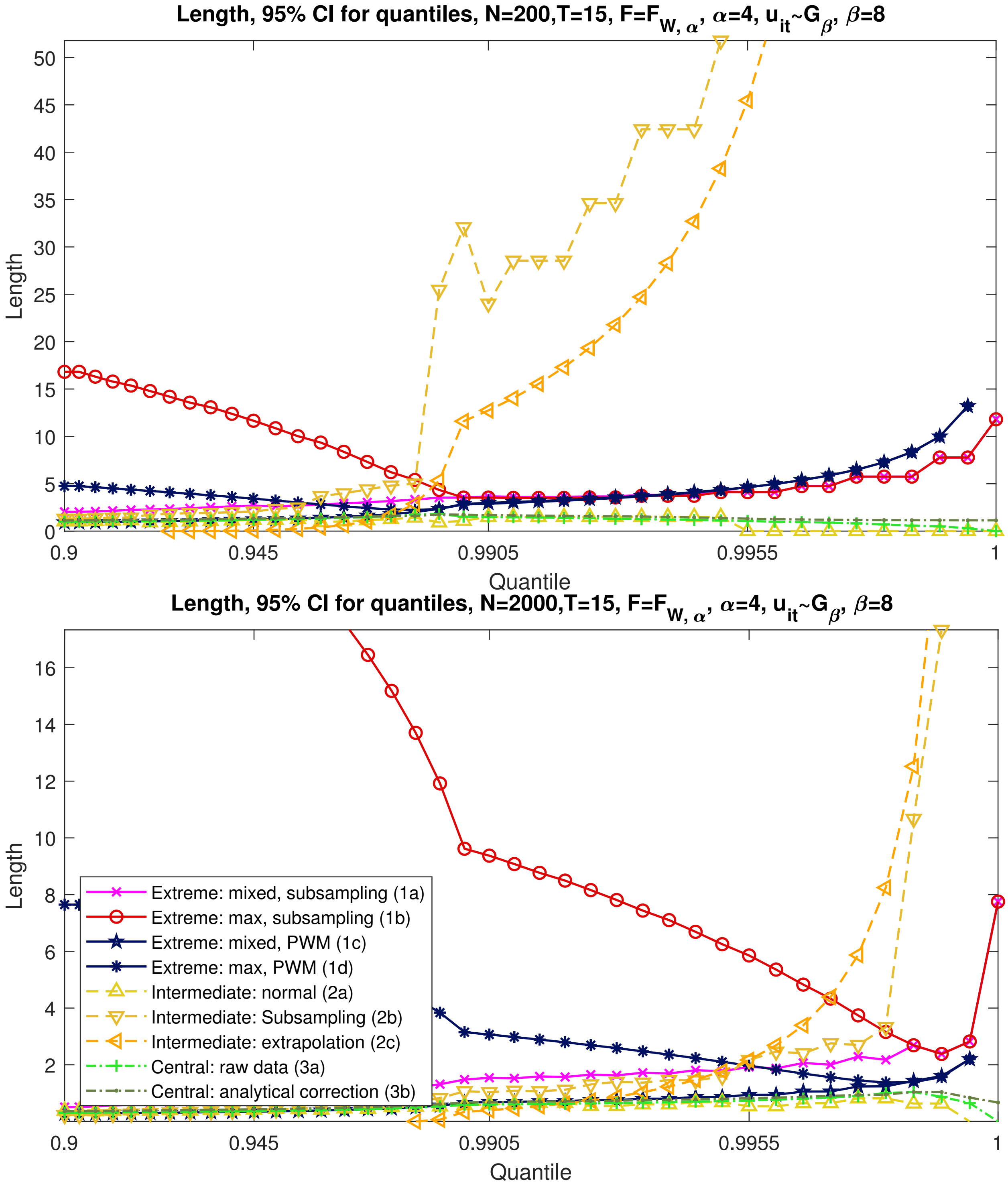


Figure 14: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_\beta, \beta = 8$

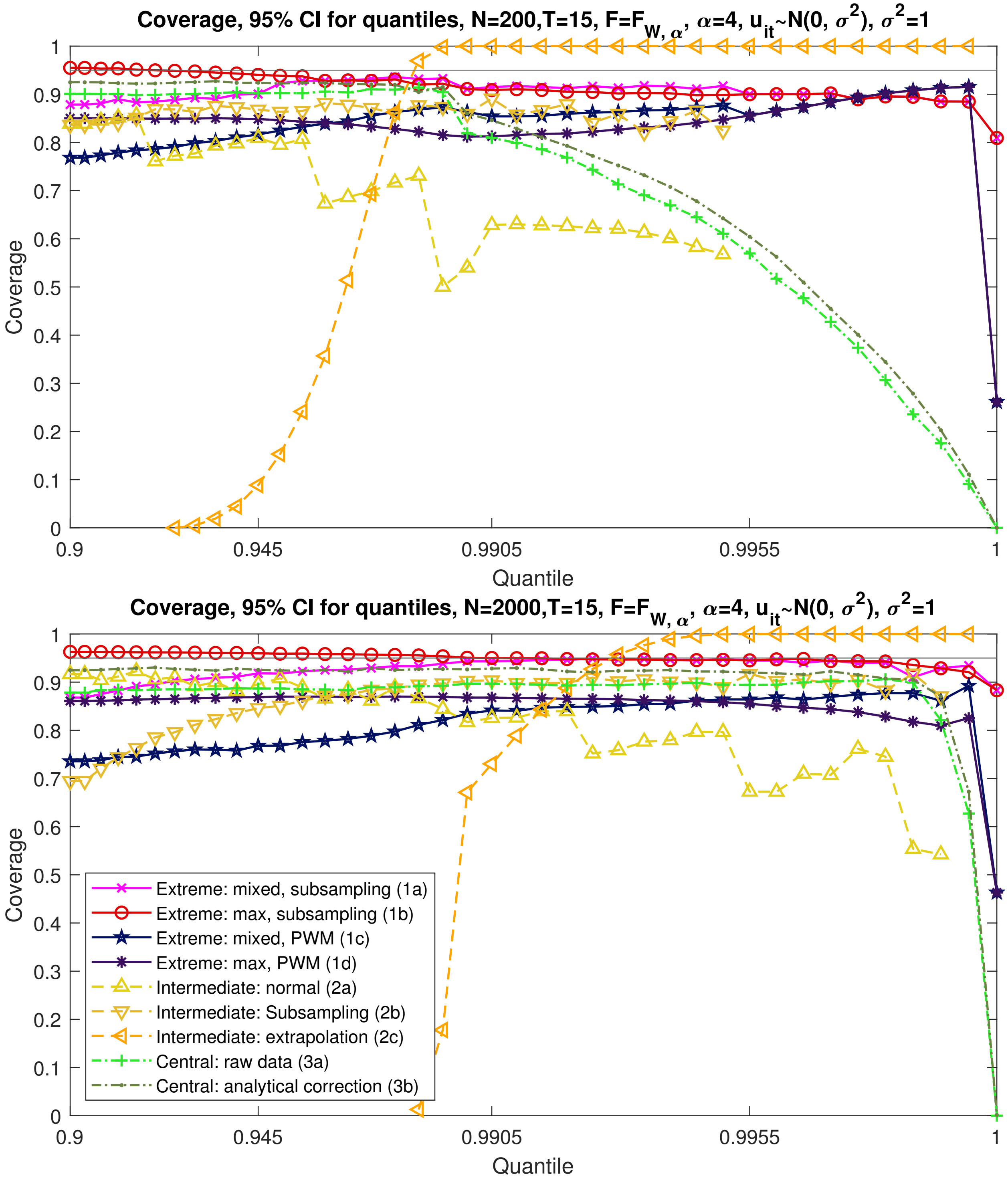


Figure 15: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$

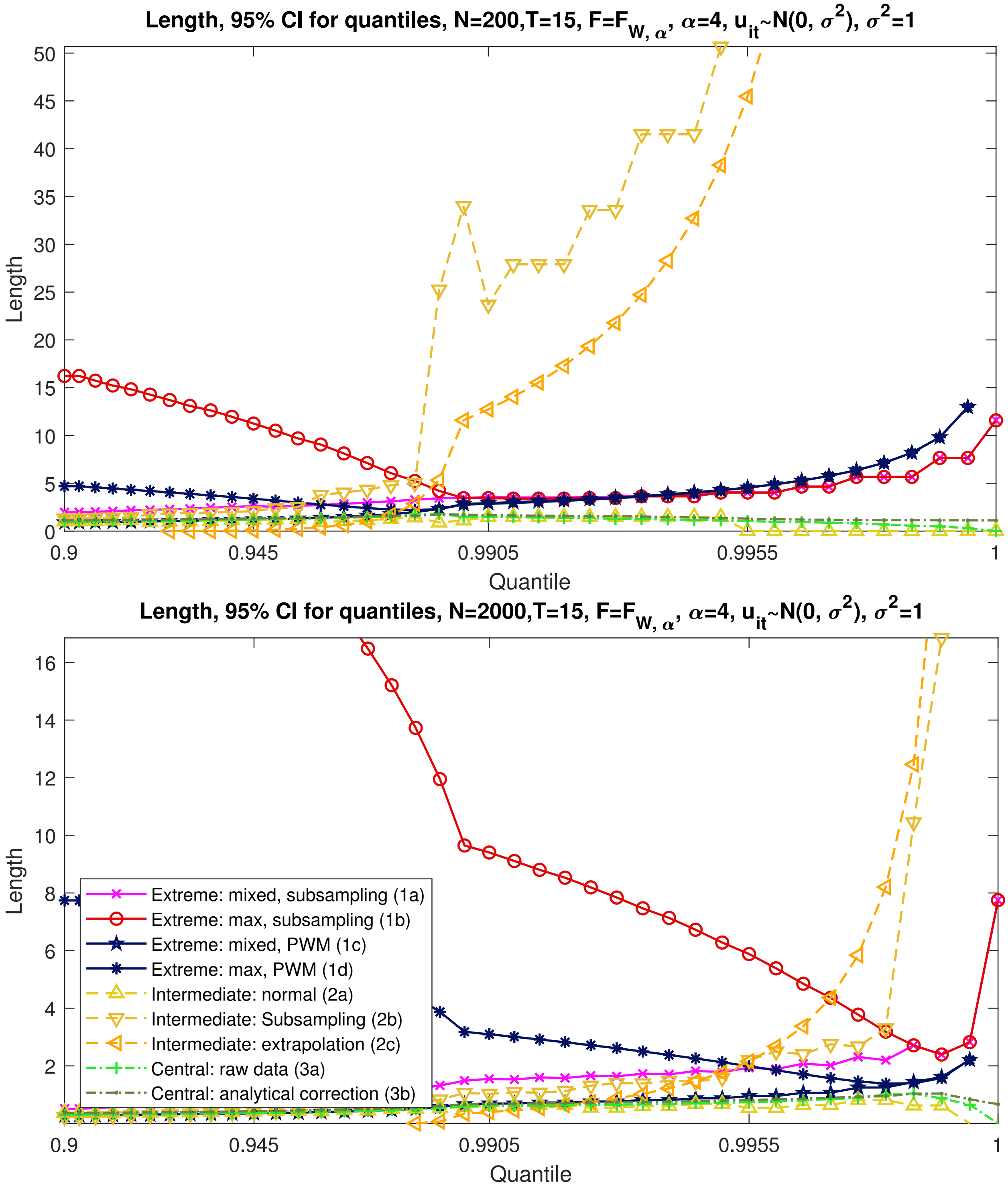


Figure 16: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$

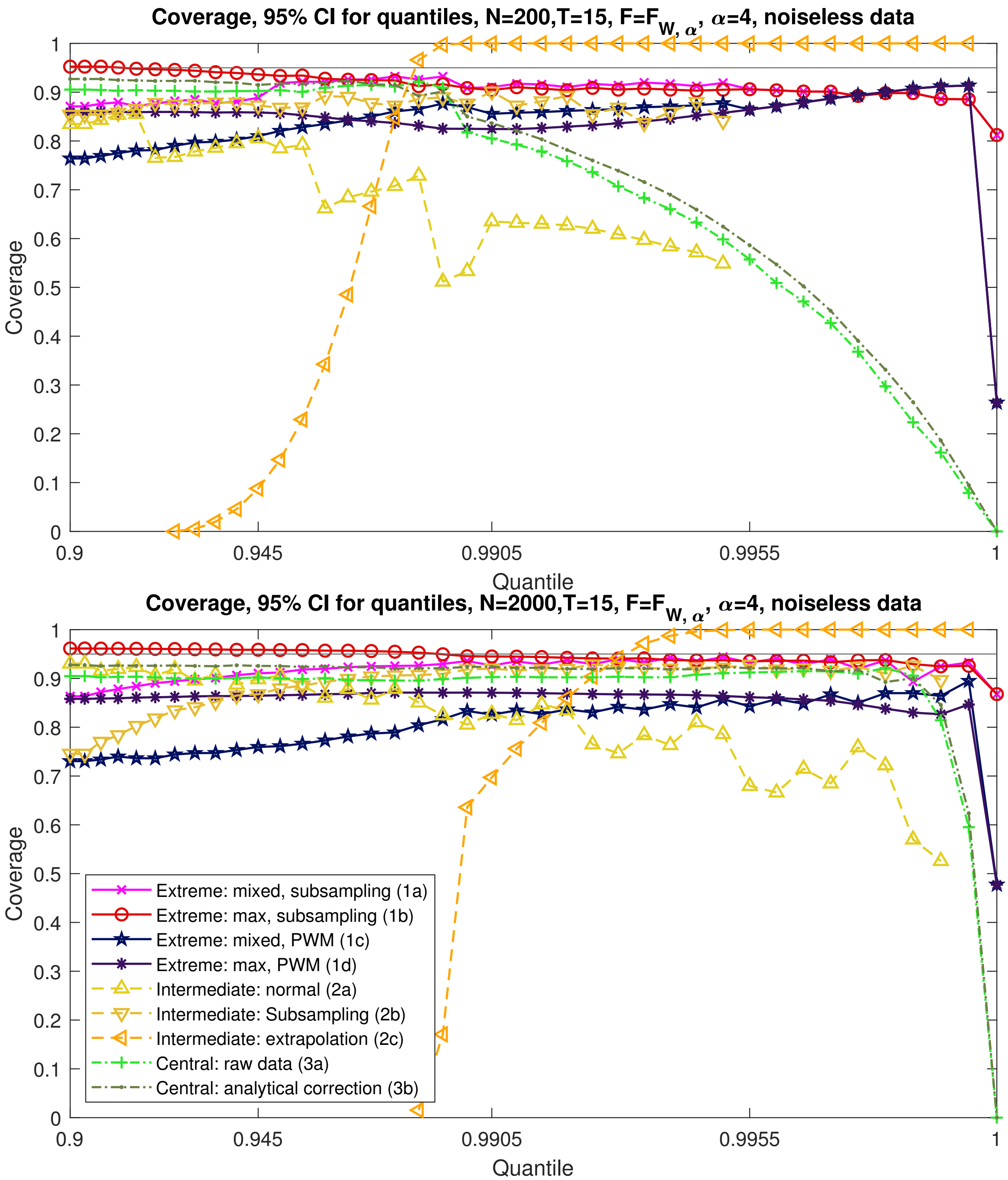


Figure 17: Coverages for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, noiseless data

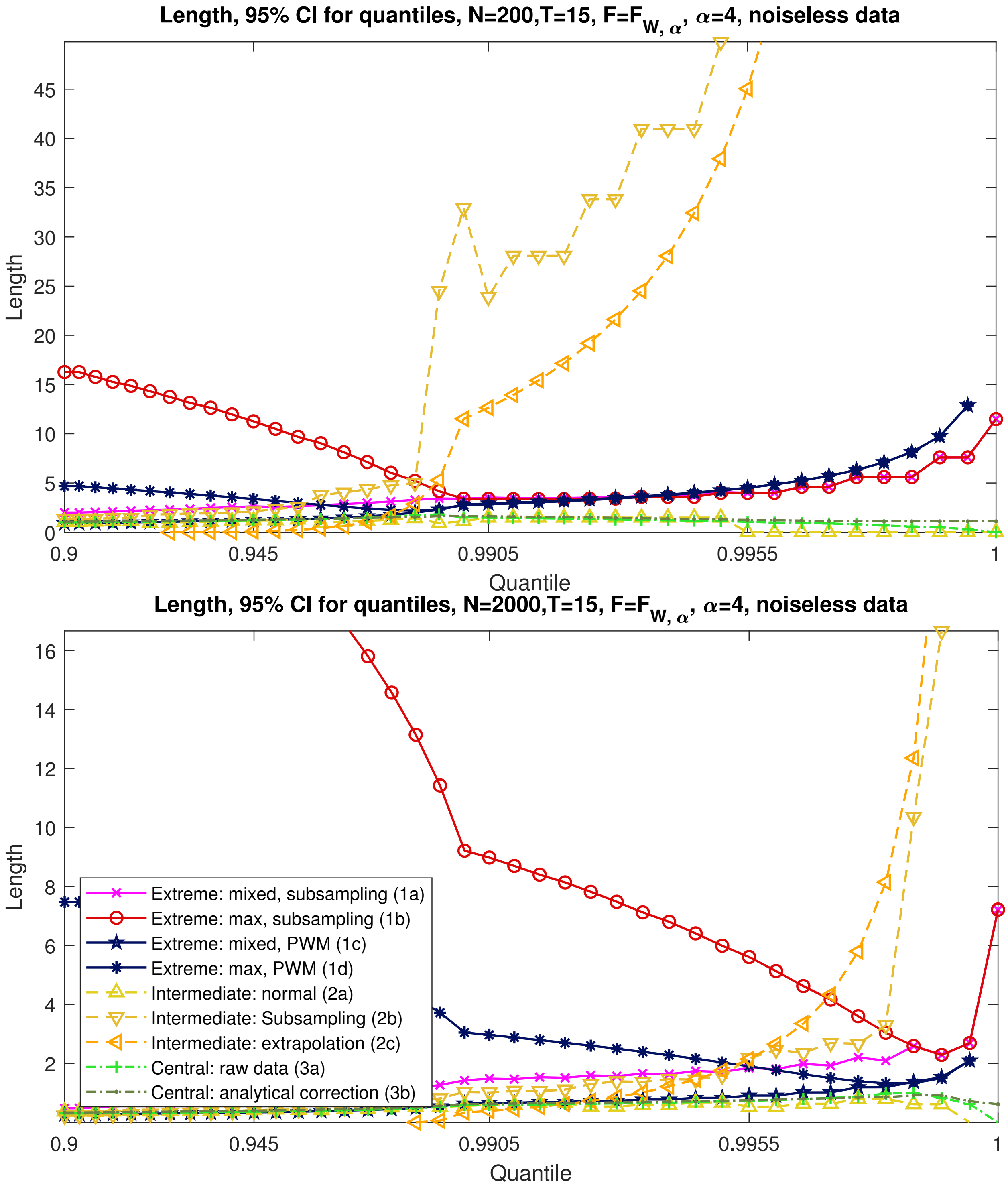


Figure 18: Confidence interval length for different approximations, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, noiseless data

OA.4.2 Corrected Estimators For Quantiles

In this section we assess the performance of adjusted estimators for quantiles proposed in example 4 and in section 4.3 of the main text. Performance is compared in terms of mean absolute error. We work in the setup of section 5 in the main text and section OA.4.

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 $r = \lfloor l \rfloor$. "Mixed" estimators are based on adjusting the sample quantile. Let $\hat{c}_\alpha$ be a consistent estimator of the αth quantile of $[(E_1^* + \cdots + E_{r+1})^{-\gamma} - l^{-\gamma}] / [\left(E_1^* + \cdots + E_{q+1}^*\right)^{-\gamma} - (E_1^*)^{-\gamma}]$.

$$\mathcal{M}_{N,T}^{mixed} = \vartheta_{N-r,N,T} - \hat{c}_{1/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right) \tag{OA.4.1}$$

"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 $[(E_1^*)^{-\gamma} - l^{-\gamma}] / [\left(E_1^* + \cdots + E_{q+1}^*\right)^{-\gamma} - (E_1^*)^{-\gamma}]$. The estimator is then defined as

$$\mathcal{M}_{N,T}^{Max\ only} = \vartheta_{N,N,T} - \tilde{c}_{1/2} \left(\vartheta_{N-q,N,T} - \vartheta_{N,N,T}\right)$$

$\hat{c}_\alpha$ and $\tilde{c}_\alpha$ are consistently estimated by subsampling and by simulation, as described in section 4.1. We also consider the extrapolation estimator of theorem 4.3.1 of de Haan and Ferreira (2006) and 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 4.3.

We plot the results graphically on figs. 19-27 . 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 2; 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.

There are two estimators that offer improvements over the unadjusted sample quantile – estimator (OA.4.1) with quantiles $\hat{c}_\alpha$ estimated by subsampling and the extrapolation estimator. The first estimator is more robust, while the second one can potentially yield larger improvements. Estimator (OA.4.1) has relative MAE <1 for a slight majority of quantiles considered and does not suffer from significant breakdowns. The extrapolation estimator offers stronger improvements for higher quantiles than estimator (OA.4.1) for distributions with $\gamma \geq 0$. However, it appears to perform less favorably in the case of $\gamma < 0$ ($F_{W,\alpha}$). We recommend against all other correction methods and against using quantiles estimated by simulation.

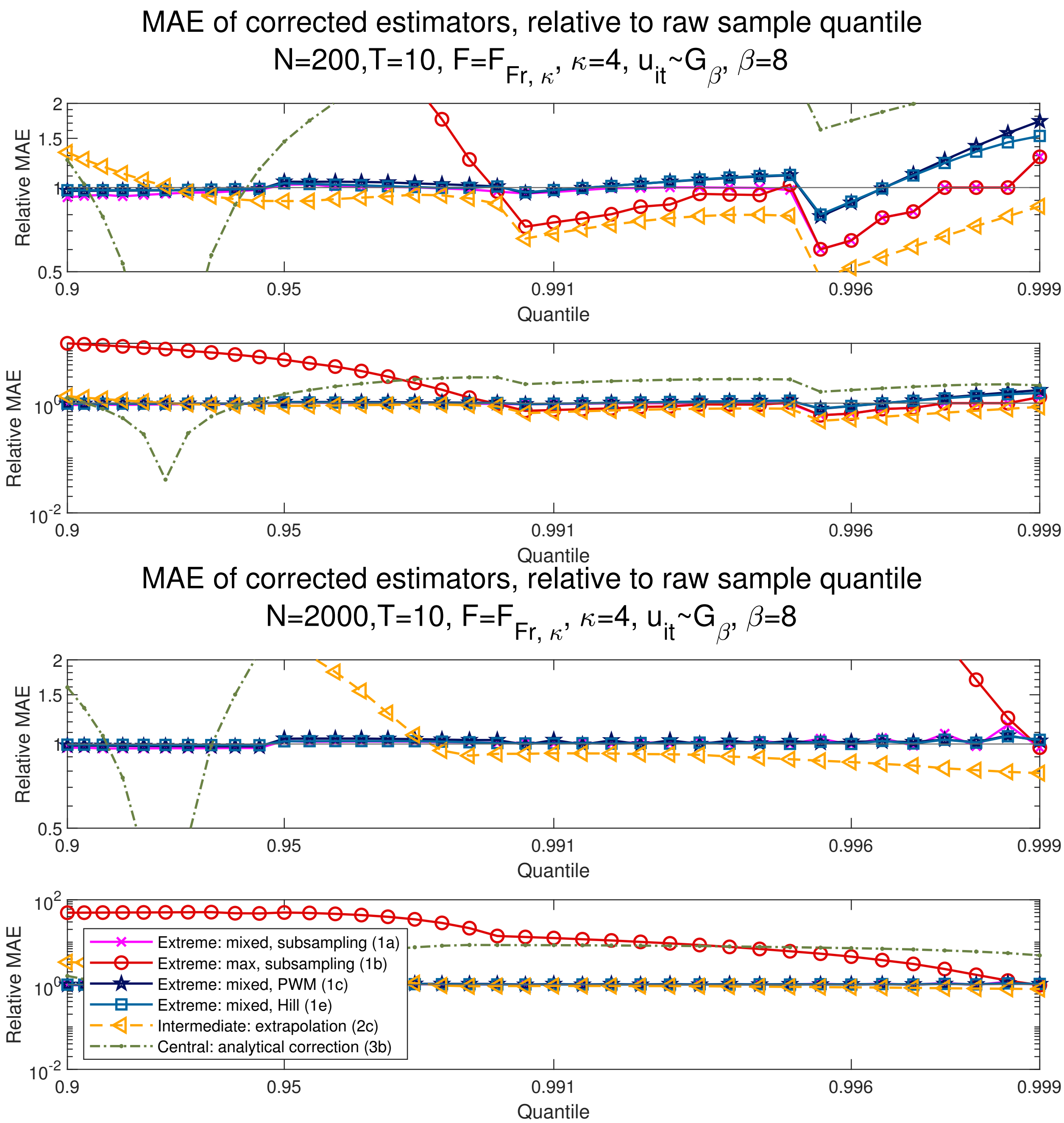


Figure 19: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim G_\beta, \beta = 8$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

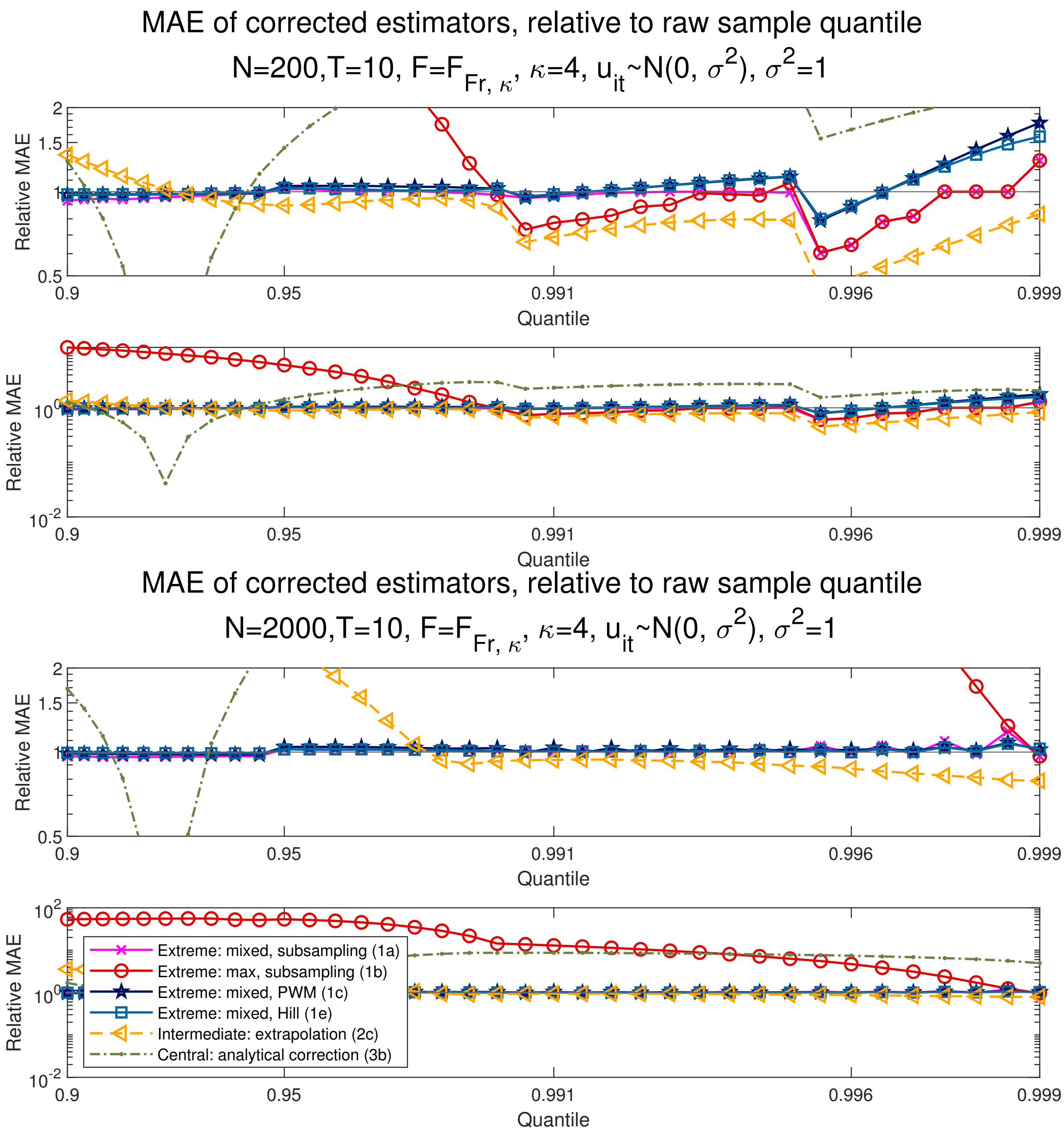


Figure 20: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

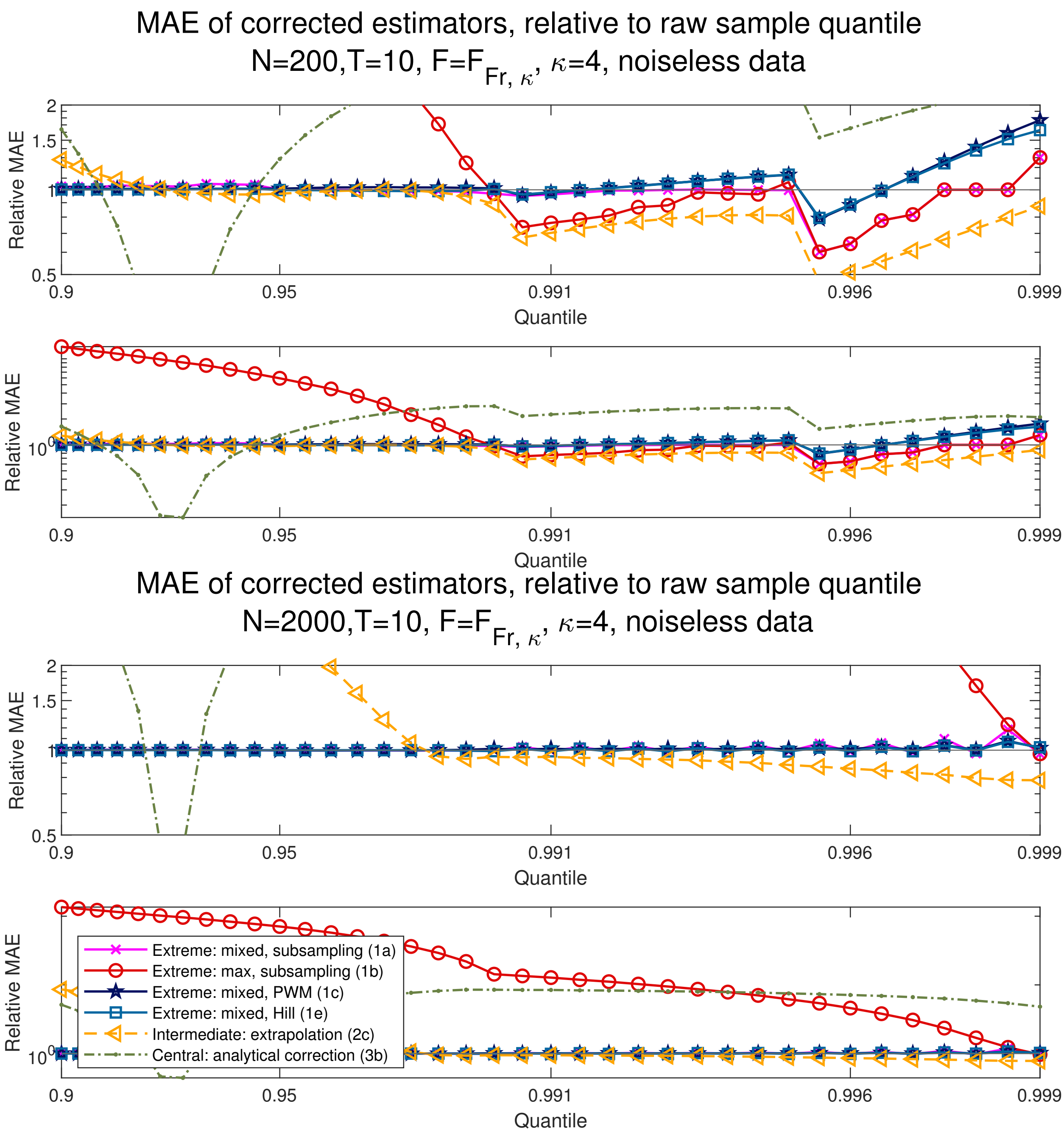


Figure 21: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 10$. $\theta \sim F_{Fr,\kappa}$, $\kappa = 4$, noiseless data. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

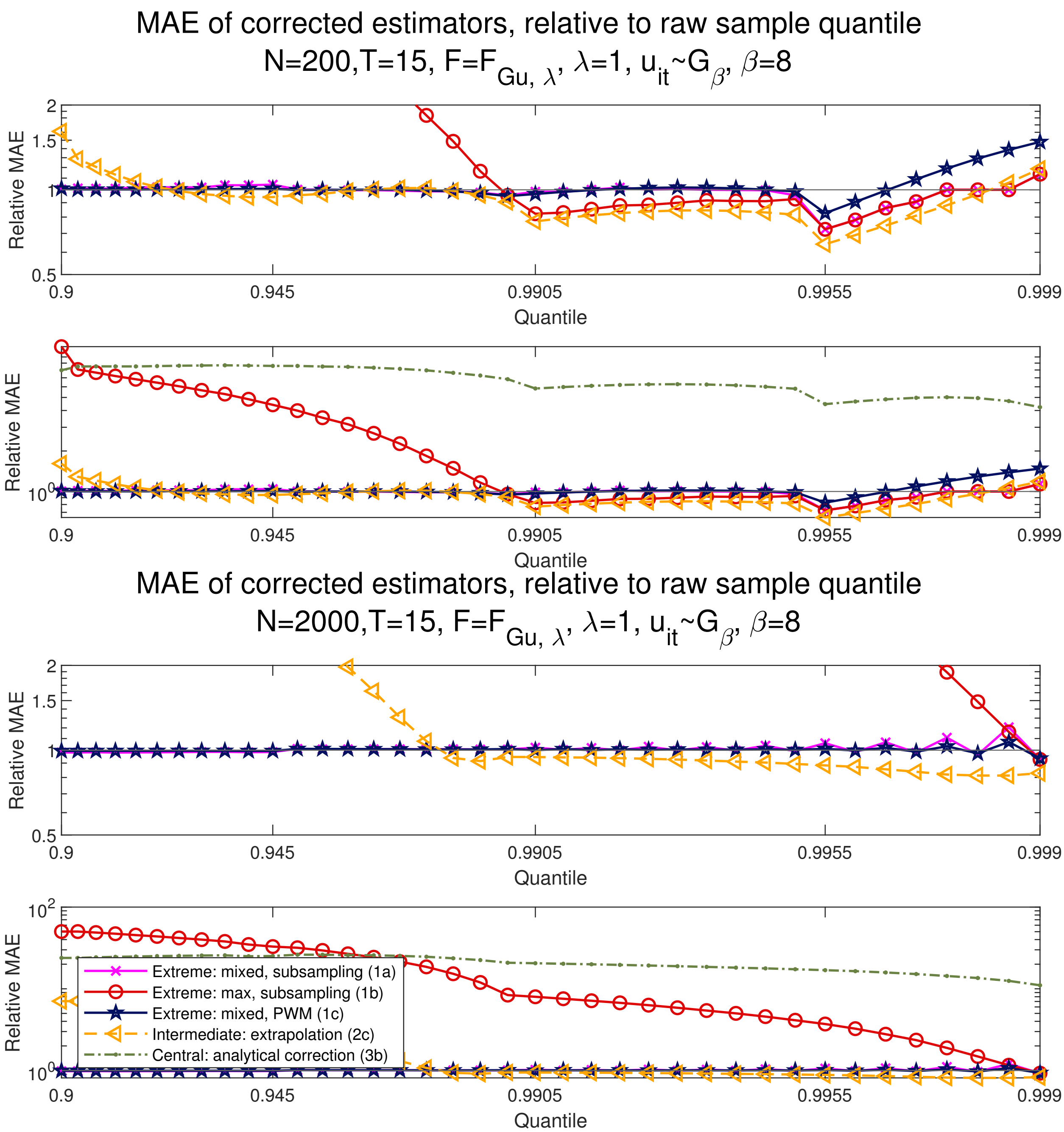


Figure 22: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim G_\beta, \beta = 8$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

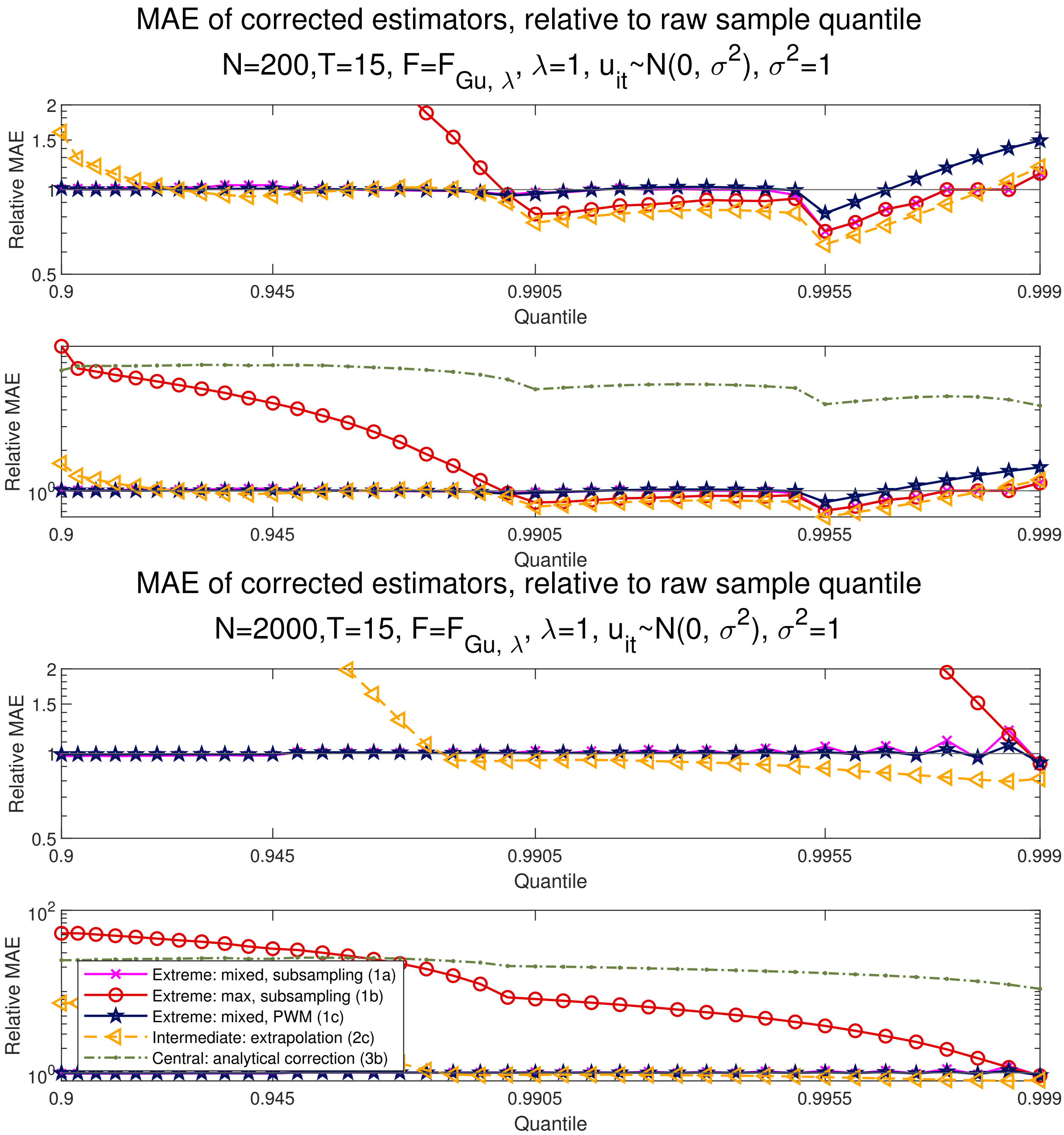


Figure 23: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

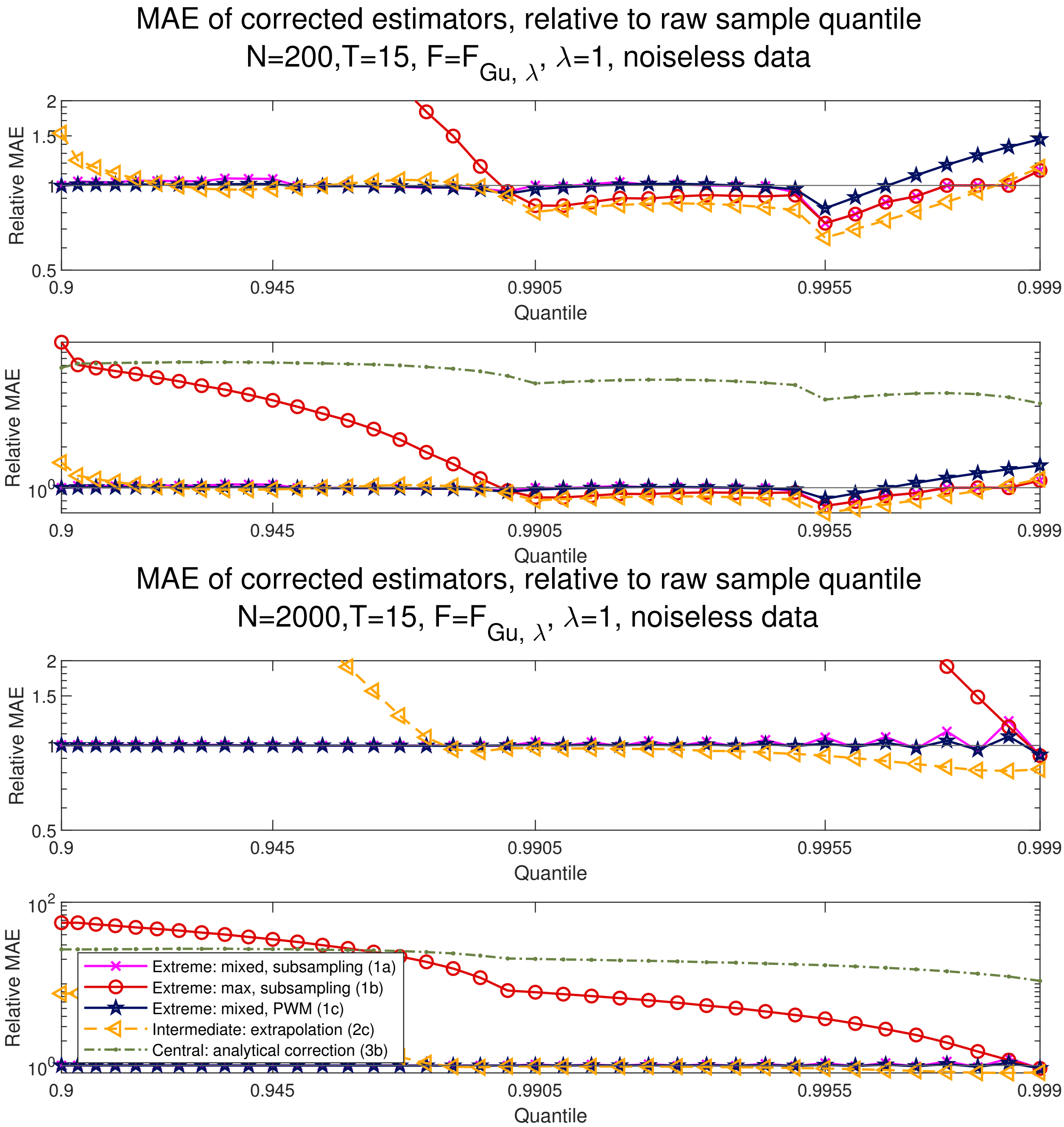


Figure 24: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{Gu,\lambda}$, $\lambda = 1$, noiseless data. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

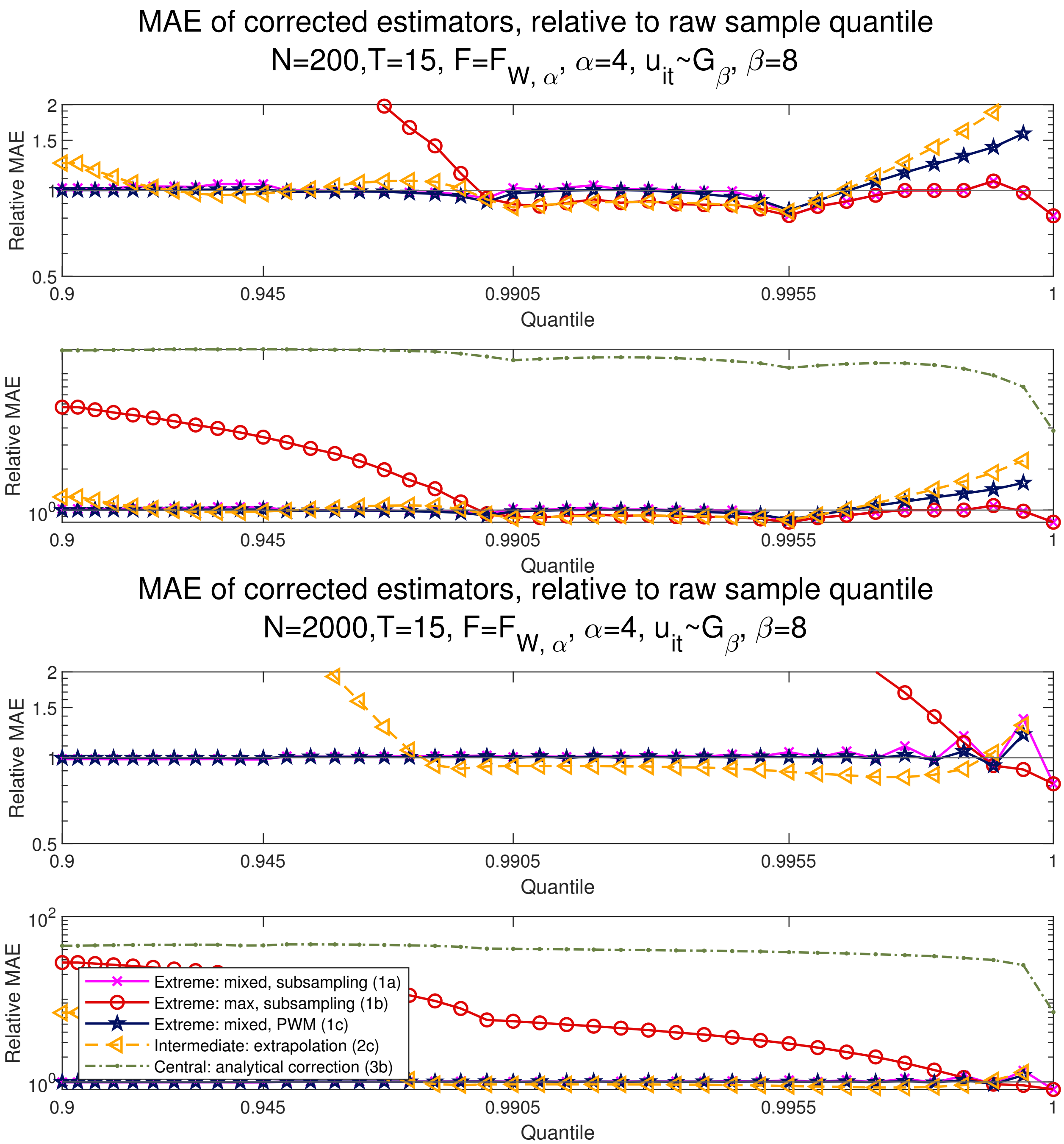


Figure 25: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim G_\beta, \beta = 8$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

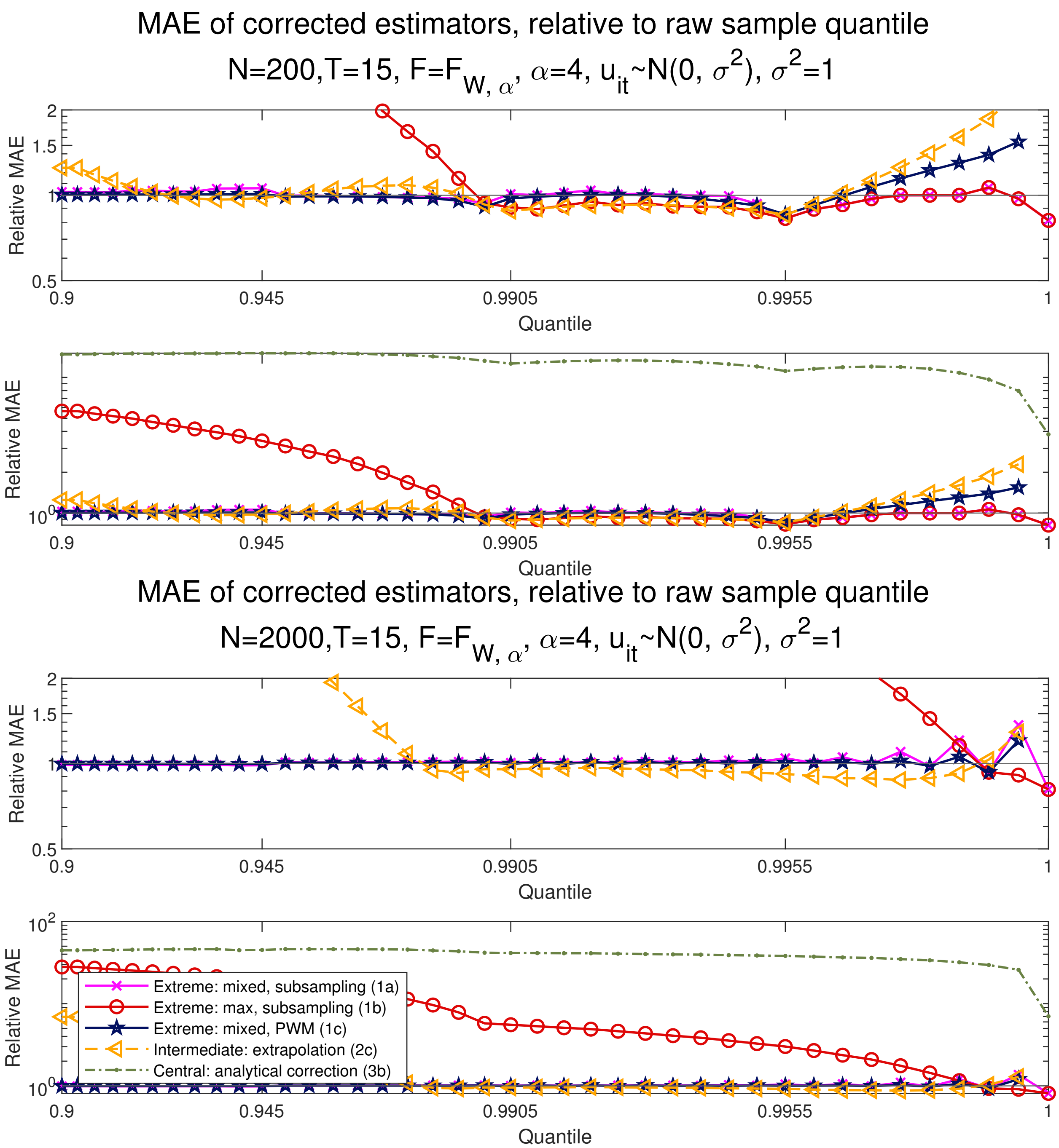


Figure 26: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, $u_{it} \sim N(0, 1)$. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

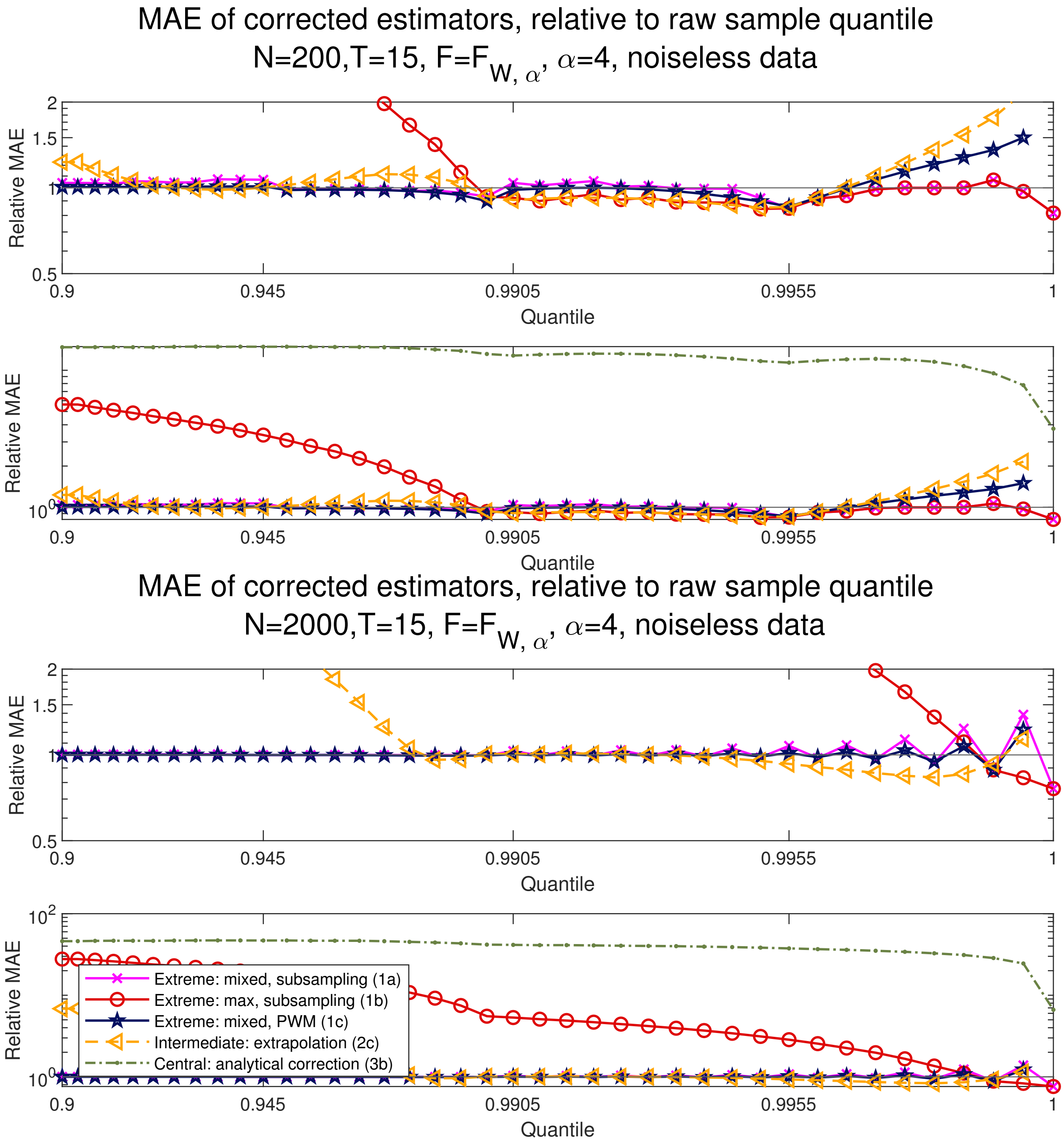


Figure 27: Mean absolute error of corrected estimators for quantiles, different approximations, MAE relative to unadjusted sample quantile, $N = 200, 2000, T = 15$. $\theta \sim F_{W,\alpha}$, $\alpha = 4$, noiseless data. Top panel: scale from 0.5 to 2. Bottom panel: scale from 0.5 to 15.

OA.5 Supplemental Materials for the Empirical Application

This section provides additional results related to our empirical application to differences in firm productivity between denser and less dense areas in the setting of Combes et al. (2012) (CDGPR12). In section OA.5.1 we discuss in detail the issue of estimating the EV index γ, including choosing the corresponding tuning parameter k and the difference between several estimators of γ. In sections OA.5.2 and OA.5.3 we provide additional results related to the analysis in section 6. We construct and compare confidence intervals using all the methods discussed in sections 4 and 5. We find that the results presented in the main text are robust to the choice of the CI used. Further, we consider the results based on a full dataset of CDGPR12 rather than a subsample of 2000 firms.

OA.5.1 Estimation of the EV Index

We begin by providing further details about estimation of the EV index γ. We consider several different estimators for γ and discuss the impact of the choice of k used to form the estimators.

We consider the following estimators for γ:

(1) The Hill (1975) estimator, as defined in remark 9 in the main text. It is consistent if $\gamma \geq 0$; if $\gamma < 0$, the Hill estimator converges in probability to 0.

(2) The probability weighted moment estimator (PWM) of Hosking and Wallis (1987), as defined in remark 9 of the main text. It is consistent for $\gamma < 1$.

(3) The average of the previous two estimators provide a compromise option. It is consistent if $\gamma \in [0, 1)$, otherwise it is asymptotically biased upwards.

Figs. 28 and 29 graphically present our estimates using the above estimators for a range of values of k. In fig. 28 we plot the estimates for the left and right tails of the distribution of productivity in areas with below-median employment density (ABMED). Fig. 29 shows the estimates for areas with above-median employment density (AAMED). In both cases k is allowed to range from 0 to $0.075N$ (we assume that up to 7.5% of the relevant sample tail displays approximate Pareto behavior).

The EV index estimators are positive and fairly large for both the left and the right tails of both the AAMED and the ABMED distributions for a wide range of values of k. In all cases, the estimated values of γ indicate fairly heavy left and right tails. However, there some differences between the values of the estimators. Generally, the Hill estimator produces the highest estimates (generally ≈ 0.38 for all tails except the right tail of AAMED, where the value is ≈ 0.32). The PWM estimator generally yields lower estimates (generally ≈ 0.25 except for the right tail of AAMED which has $\hat{\gamma}_{PWM} \approx 0.2$).

All estimators are fairly insensitive to choice of k, as figs. 28 and 29 show. The only exception occurs for k small (≤ 50), where the estimators change rapidly with k, before broadly stabilizing for larger values of k. Our choice of $k = \lfloor N^{3/5} \rfloor$ lies in the stable region for both AAMED and ABMED.

The point estimates obtained in the main text are obtained by considering the average of the Hill and the PWM estimators for $k = \lfloor N^{3/5} \rfloor$, which corresponds to $k = 95$ for both AAMED and

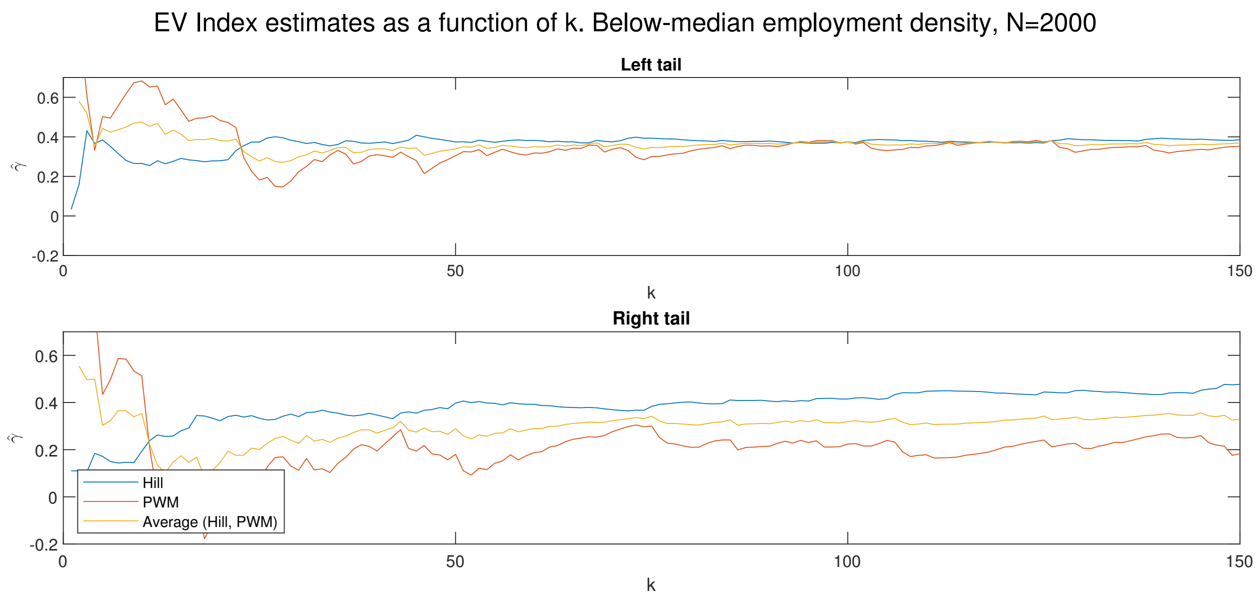


Figure 28: Estimates of EV index γ for a range of values of k. Areas with below-median employment density. Top panel: left tail. Bottom panel: right tail

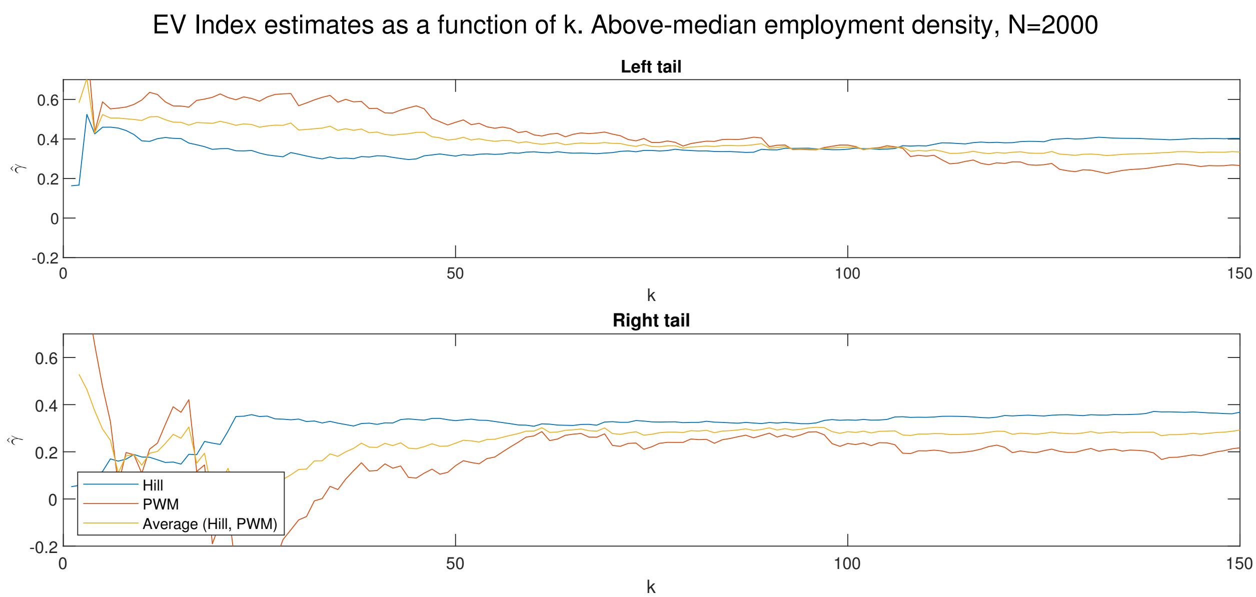


Figure 29: Estimates of EV index γ for a range of values of k. Areas with below-median employment density

ABMED. The estimated EV indices of the ABMED distribution is 0.36 for the left tail and 0.32 for the right tail. The estimates for the AAMED distribution are 0.35 for the left tail and 0.30 for the right tail.

OA.5.2 Equivalence of Tails up to Mean and Variance

We now present additional results for the question of whether mean and variance are sufficient to capture the difference between the AAMED and the ABMED distributions. As in the main text, we first standardize the data so that both datasets have the same mean and variance. We construct 95% confidence intervals for a range of extreme quantiles of standardized data using the confidence intervals considered in sections 4-5 and OA.4.1. The tuning parameters are chosen as in the main text and section 5. We remark that the analytical correction of Jochmans and Weidner (2022) cannot be constructed, as we do not have access to the original data, nor to the estimates of the variance of each $\vartheta_{i,T}$. The quantiles considered are 0.001-0.1 for the left tail and 0.9-0.999 for the right tail.

Almost all methods yield virtually identical confidence intervals on the full dataset, as figs. 30-31 show. The two exceptions are the extrapolation interval and the interval based on theorem 4.5 with subsampled critical values, which are considerably wider for quantiles beyond 0.02 and 0.995. However, these intervals appear to have the same centering as the other CIs throughout. These results are in line with the simulation evidence presented in section OA.4.1. Based on this similarity of intervals, in the main text we report the CI based on theorem 4.3 with subsampled critical values (line "Extreme: mixed, subsampling" on figs. 30-31).

We also repeat our analysis with the full dataset of CDGPR12. Recall that we assume that the estimation noise in $\vartheta_{i,T}$ is assumed to have at least 8 moments. Then the finding of heavy tails of F_{AM} and F_{BM} implies that inference should be reliable, as in this case proposition 3.2 shows that there are no restrictions on the relative sizes of N and T.

The confidence intervals based on the full dataset are reported on figs. 32-33. As above, the different CIs yield broadly the same results for both tails of the distributions for AAMED and ABMED. The results follow the pattern presented in section 5. However, for quantiles below 0.025 and above 0.99 the intervals also display near-zero length, owing to the large cross-sectional size. As in the simulations, the extrapolation-based interval and the interval based on theorem 4.5 with subsampled critical values are considerably wider. Fig. 34 is the counterpart of fig. 3 in the main text using the full sample. For the left tail, the CIs overlap almost completely. The situation is somewhat more delicate for the right tail. The 0.9-0.97th quantiles are statistically different. However, the difference between the two distributions appears minor, and the tails appear to have the same shape. Overall, this finding supports the key assumption CDGPR12 make in their second estimation stage.

95% CIs for extreme quantiles, standardized mean and variance. Below-median density employment, N=2000

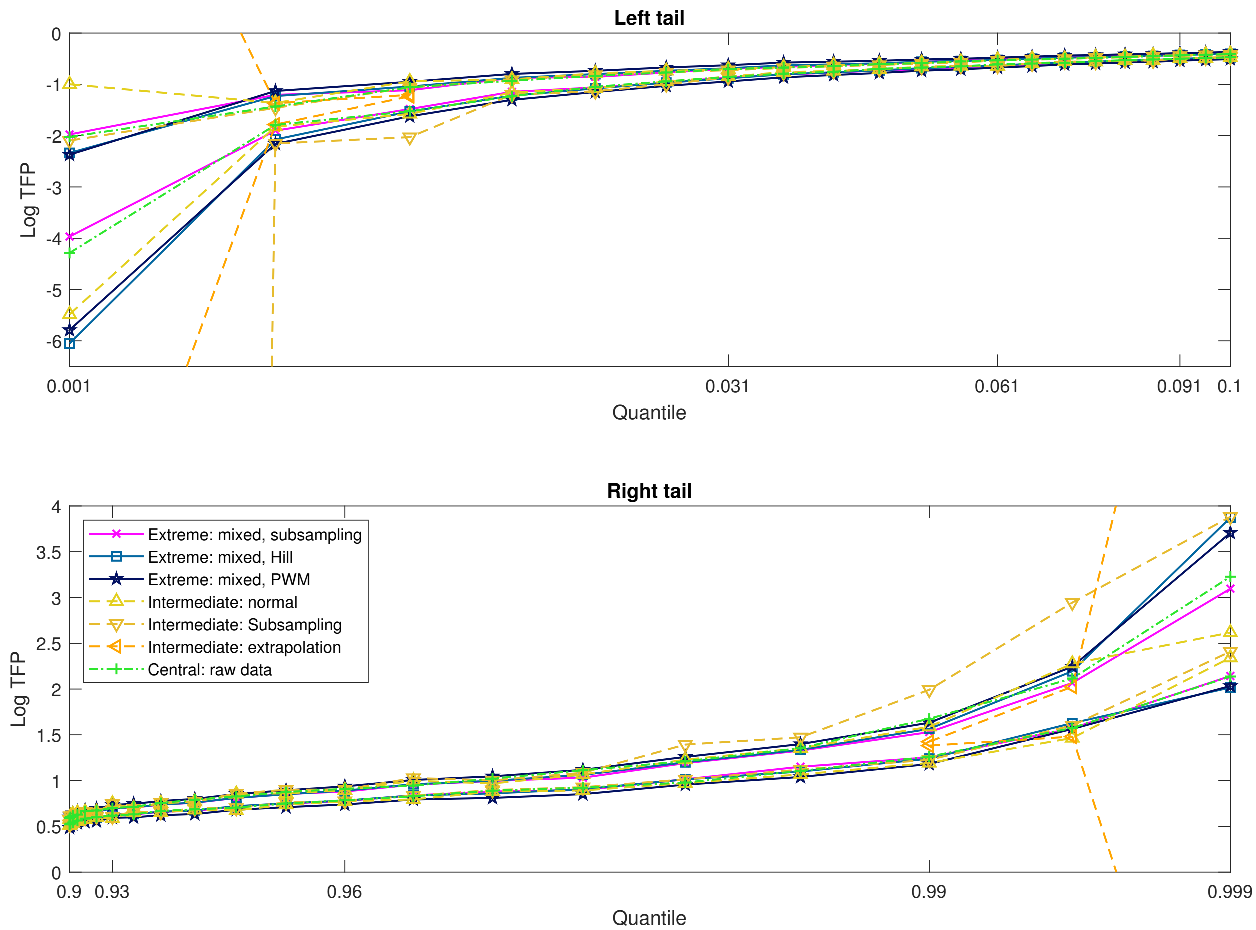


Figure 30: 95% confidence intervals for extreme quantiles, ABMED data, random subsample with $N = 2000$, standardized to have the same mean and variance as AAMED. CIs as in section 5 and OA.4

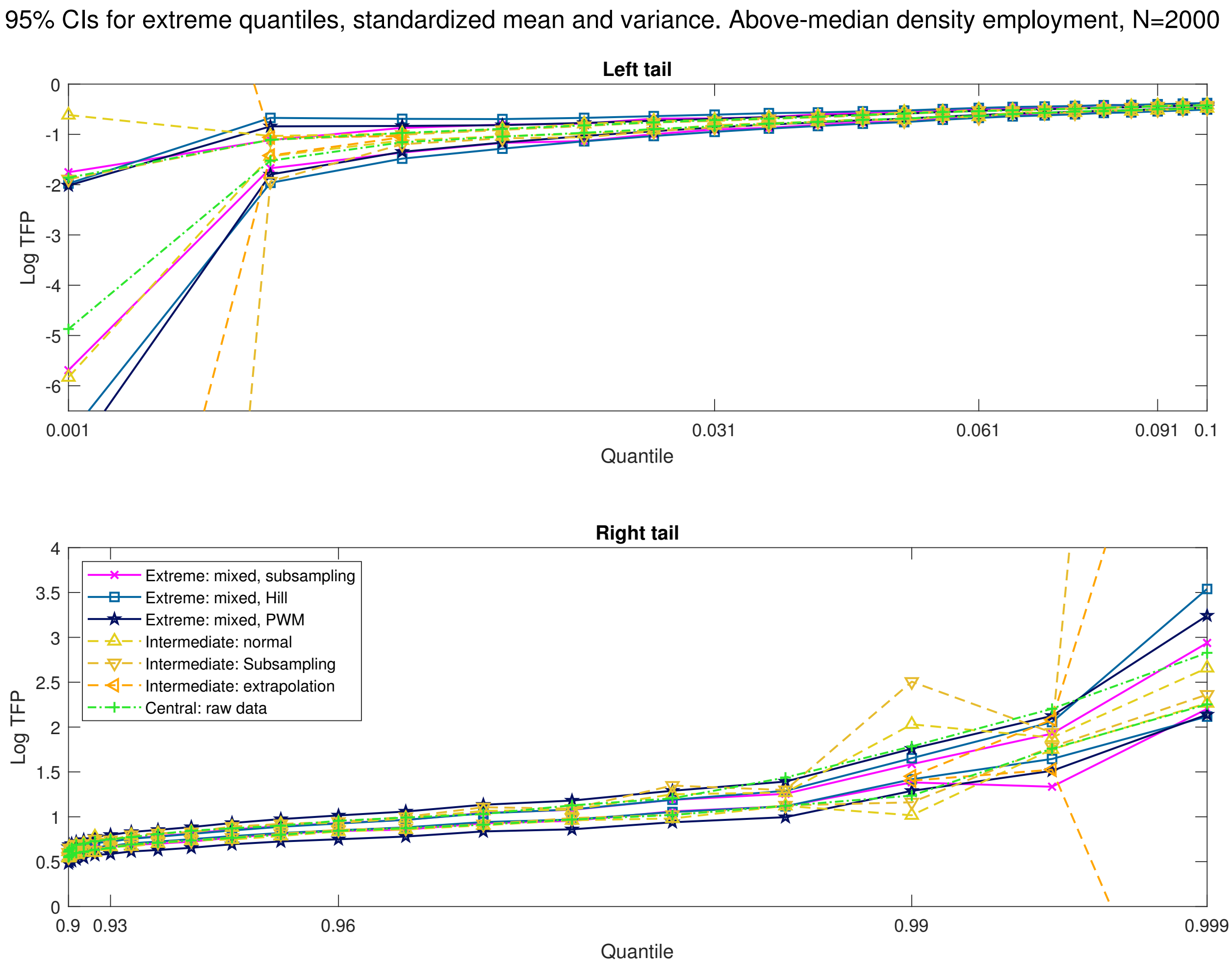


Figure 31: 95% confidence intervals for extreme quantiles, AAMED data, random subsample with $N = 2000$, standardized to have the same mean and variance as ABMED. CIs as in section 5 and OA.4

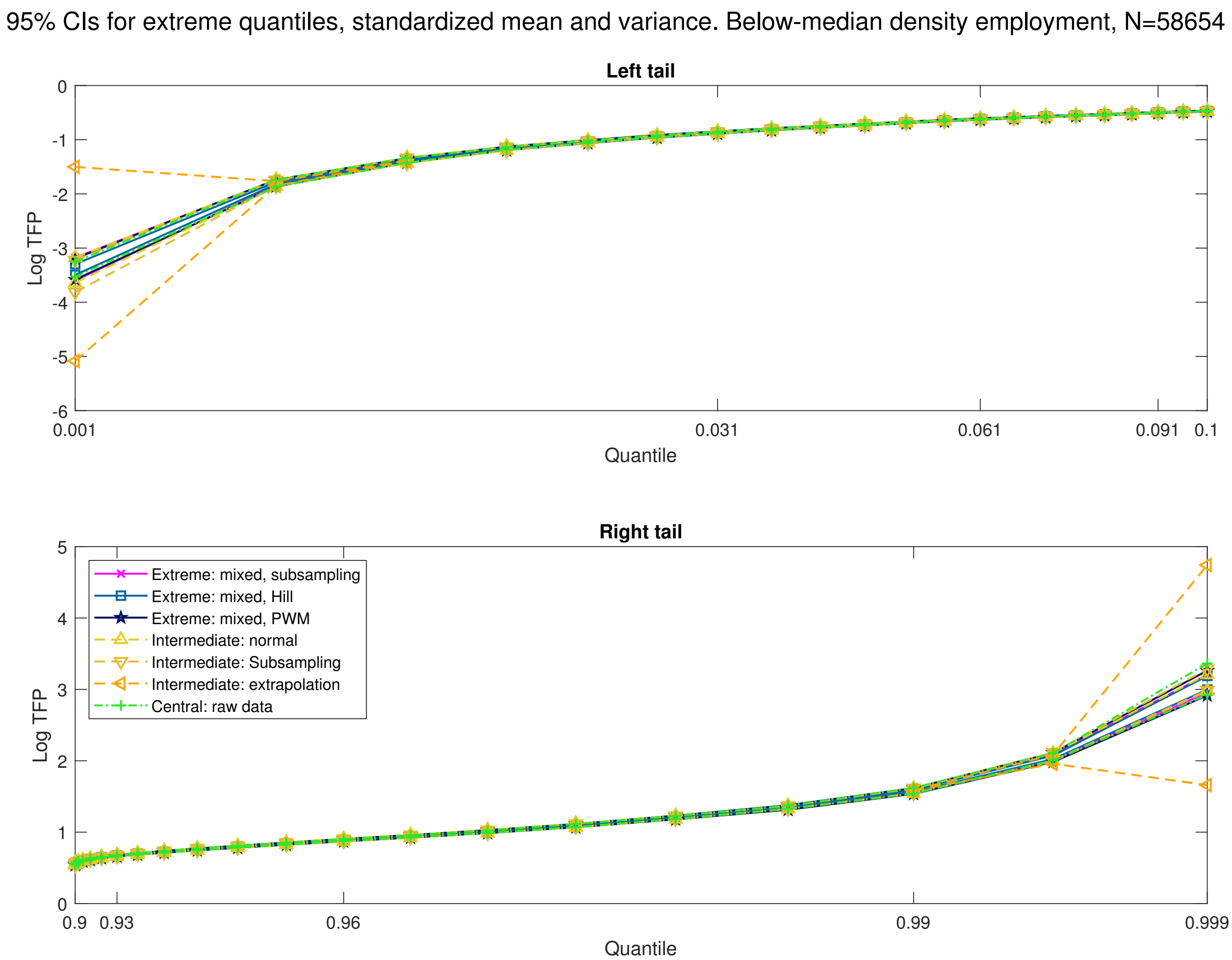


Figure 32: 95% confidence intervals for extreme quantiles, ABMED data, full sample standardized to have the same mean and variance as AAMED. CIs as in section 5 and OA.4

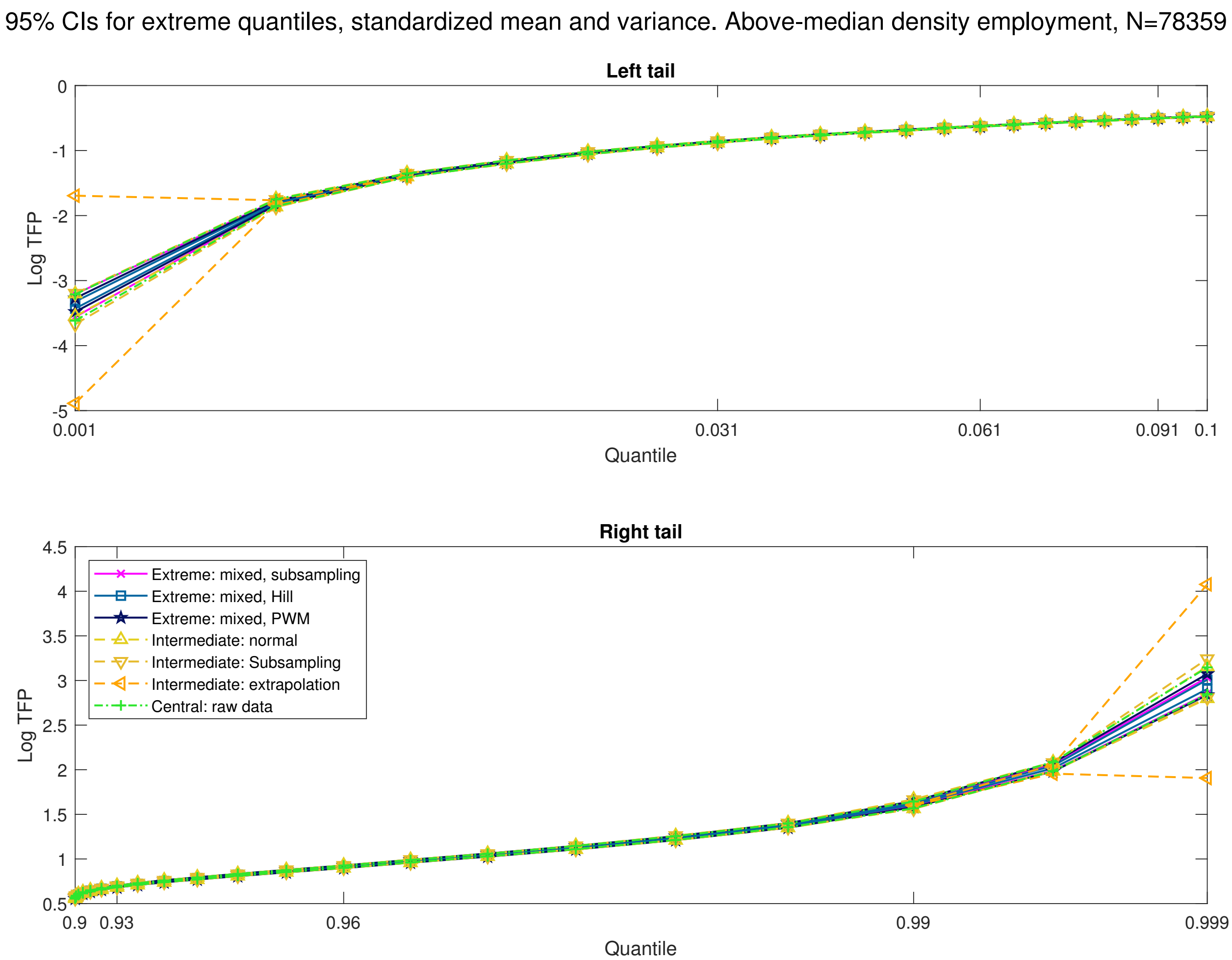


Figure 33: 95% confidence intervals for extreme quantiles, AAMED data, full sample standardized to have the same mean and variance as ABMED. CIs as in section 5 and OA.4

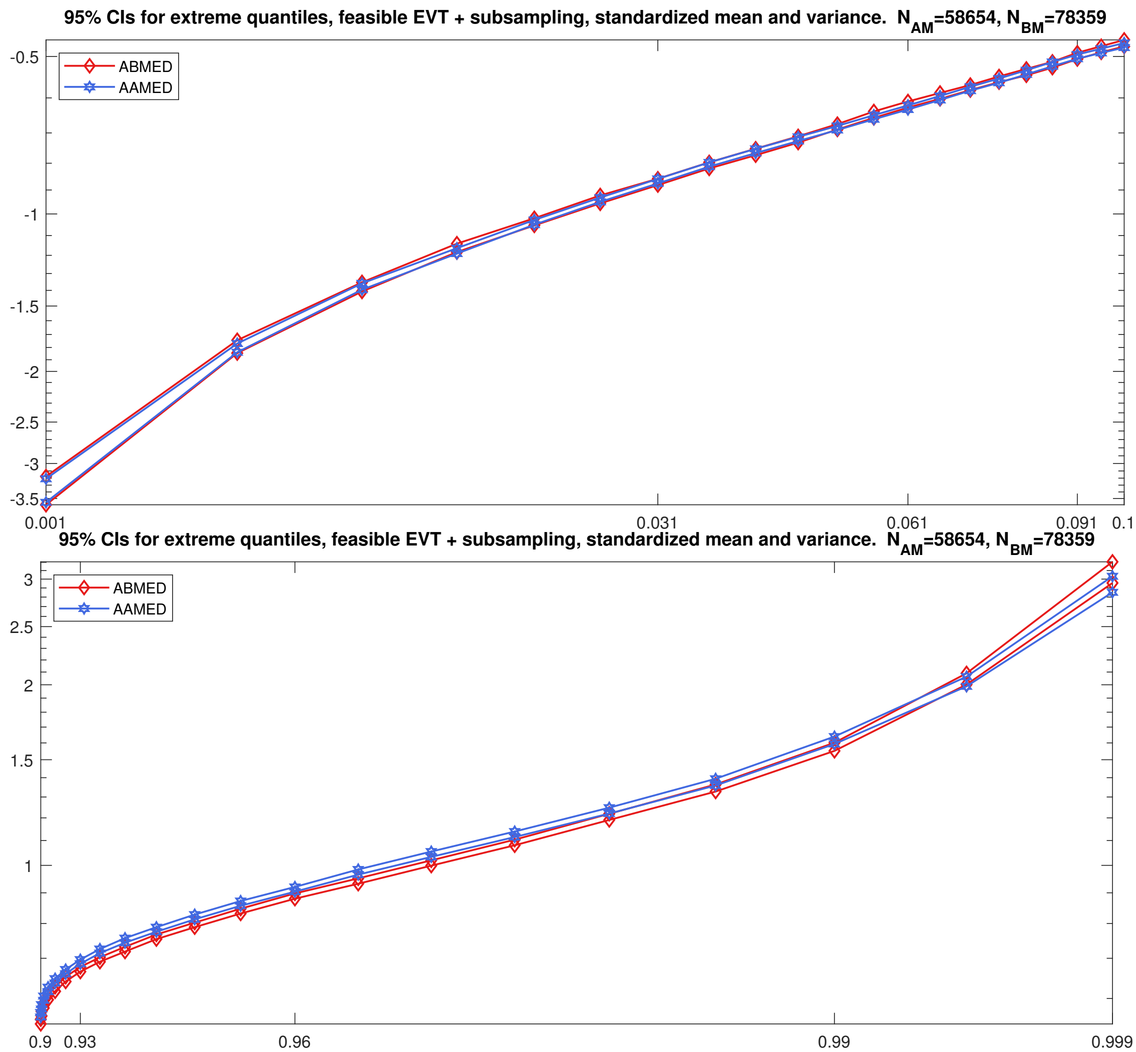


Figure 34: 95% confidence intervals for extreme quantiles (between corresponding lines), AAMED and ABMED data, full sample standardized to have the same mean and variance. CIs based on theorem 4.3 with critical values estimated by subsampling

OA.5.3 Confidence Intervals for Extreme Quantiles of Productivity

We provide analysis analogous to the previous section for both tails of the productivity distribution in AAMED and ABMED. As in the main text, the data is not adjusted. For the left tail we consider the 0.001-0.1th quantiles; for the right tail the 0.9-0.999th quantiles. The complete results are presented on figs. 35-37 for the subsample used in the main text and on figs. 38-40 for the full sample. We omit the comparison of different CIs, as the results are identical to those of section OA.5.2.

There is a significant difference between the low quantiles of F_{AM} and F_{BM} when using the full dataset. As fig. 40 shows, the 0.001-0.07th quantiles in ABMED are significantly higher than those in AAMED. Informally, this can be phrased as the least productive firms in less dense areas being more productive than the least productive firms in denser areas. Mathematically, such a

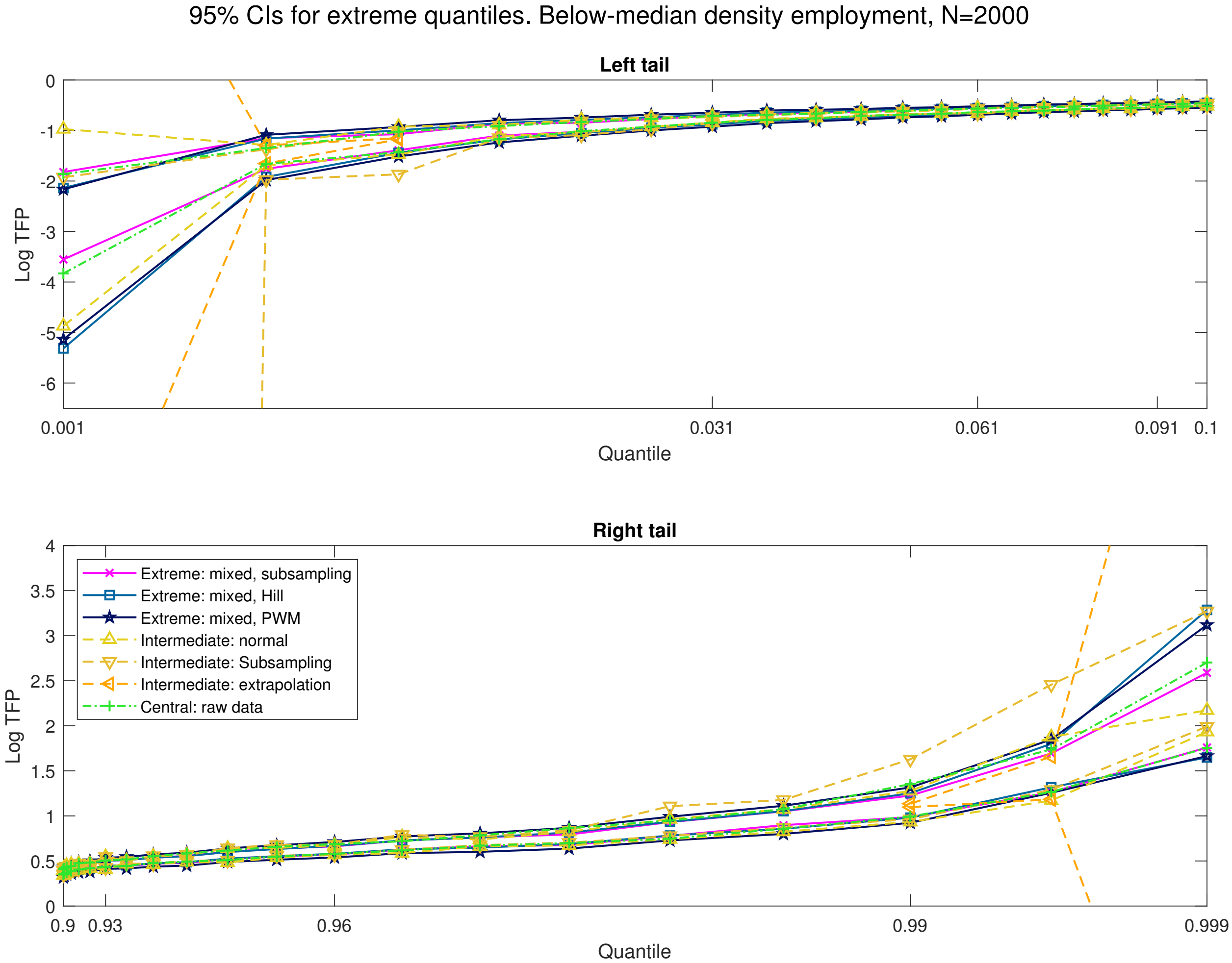


Figure 35: 95% confidence intervals for extreme quantiles, ABMED data, random subsample with $N = 2000$. CIs as in section 5 and OA.4

finding is natural. F_{AM} and F_{BM} have the same tails up to mean and variance. However, F_{AM} has higher variance and thus a heavier tail. As F_{BM} has a lower mean, eventually the tails of F_{BM} and F_{AM} cross. This crossover happens between the 0.07th and the 0.1th quantiles (see fig. 40). Economically, this result seems to go against the hypothesis of stronger competition in dense areas. However, we conjecture that the effect is due to difference in sectoral composition between AAMED and ABMED, rather than a failure of the hypothesis. As the publicly available data from Combes et al. (2012) does not include sectoral labeling, we leave a detailed investigation of the point to future research.

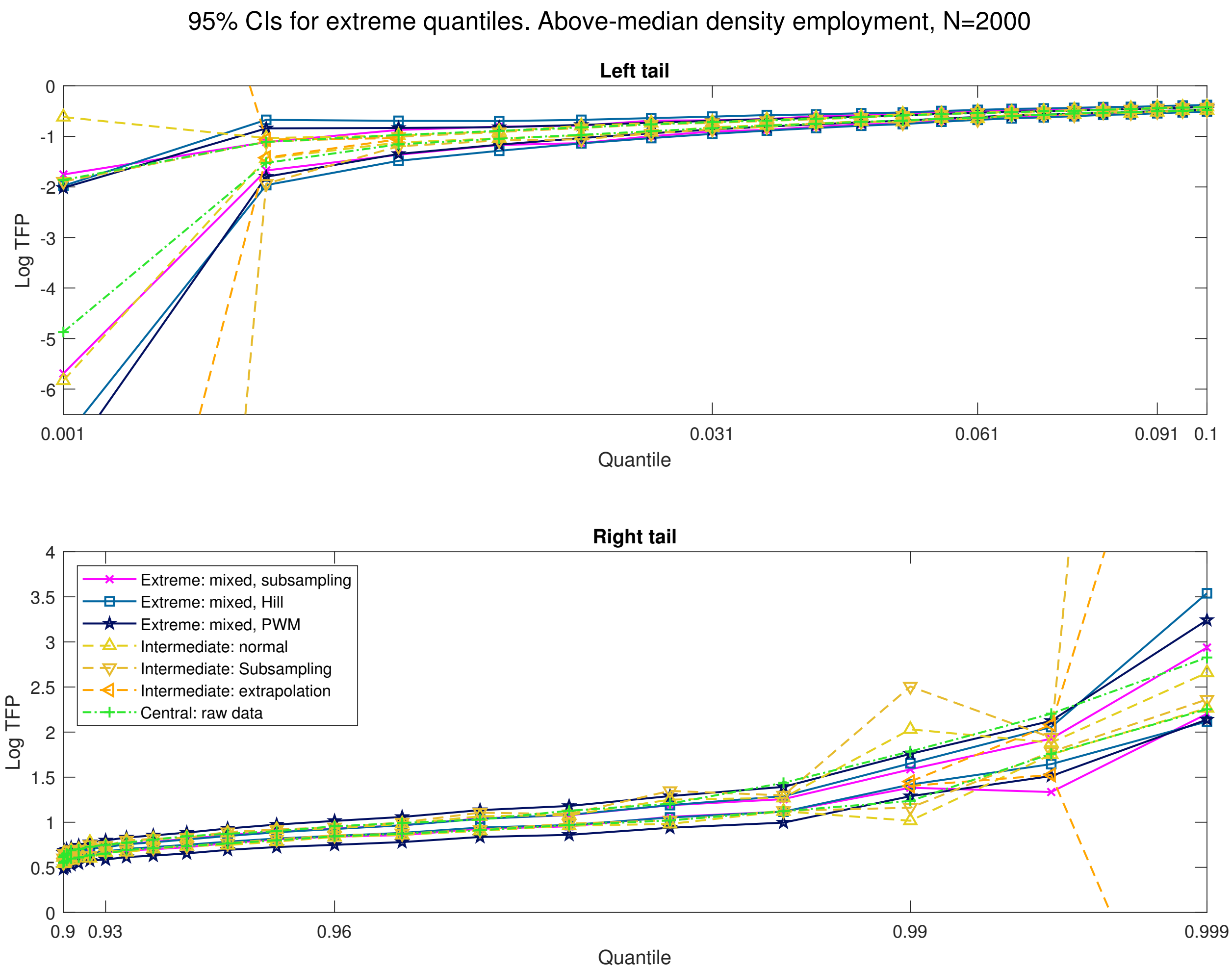
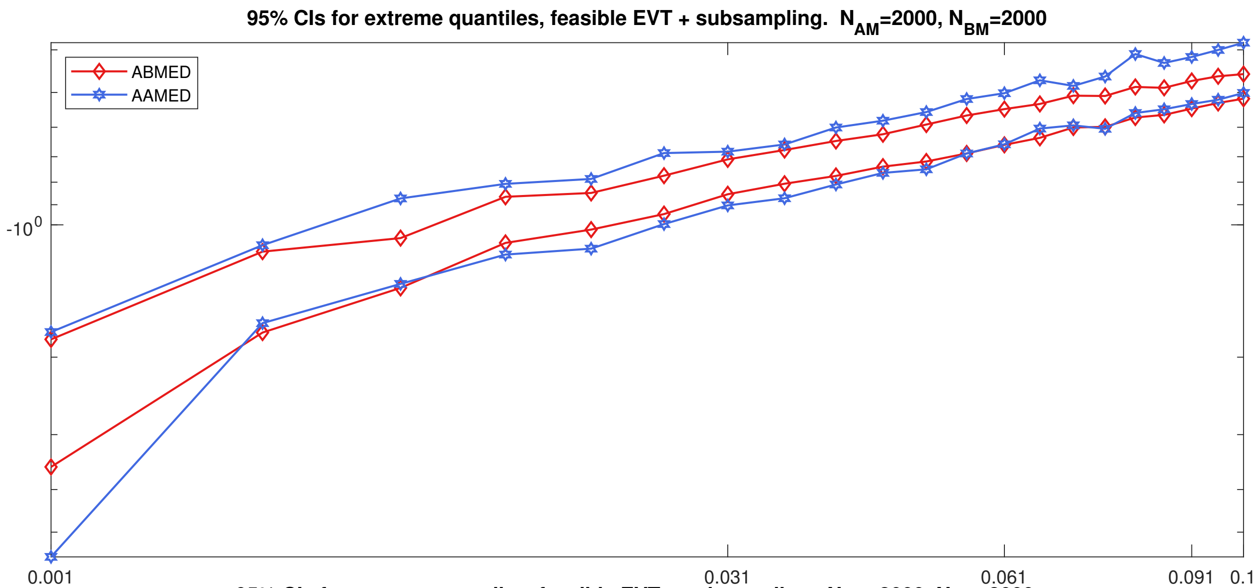


Figure 36: 95% confidence intervals for extreme quantiles, AAMED data, random subsample with $N = 2000$. CIs as in section 5 and OA.4



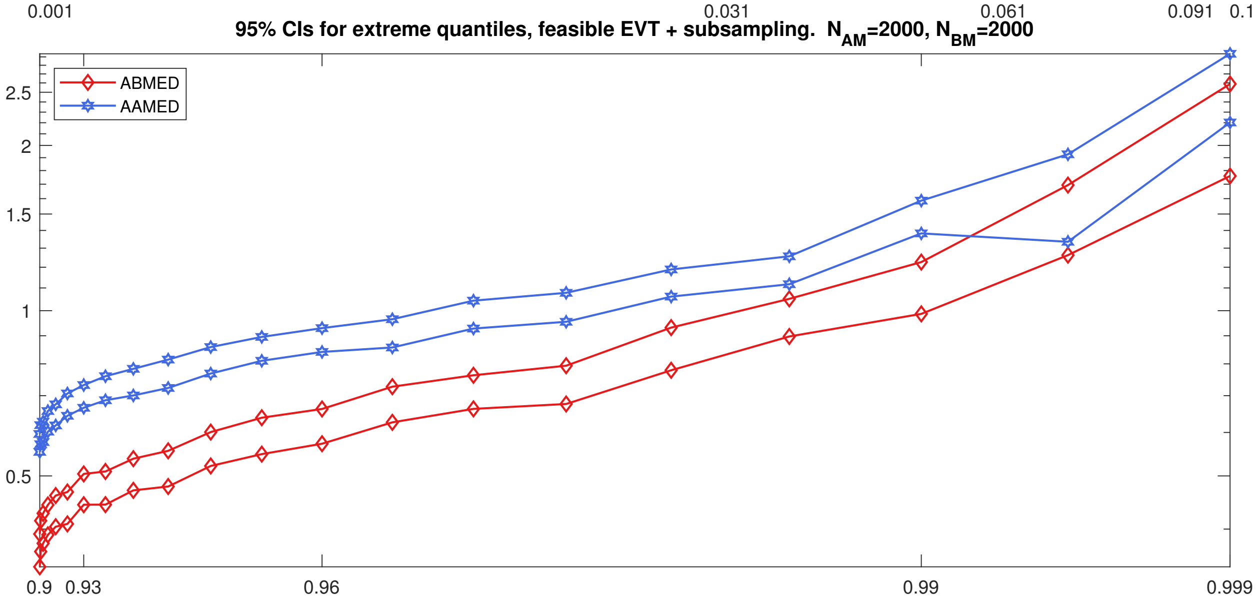


Figure 37: 95% confidence intervals for extreme quantiles, AAMED and ABMED data, random subsample with $N = 2000$. CIs based on theorem 4.3 with critical values estimated by subsampling

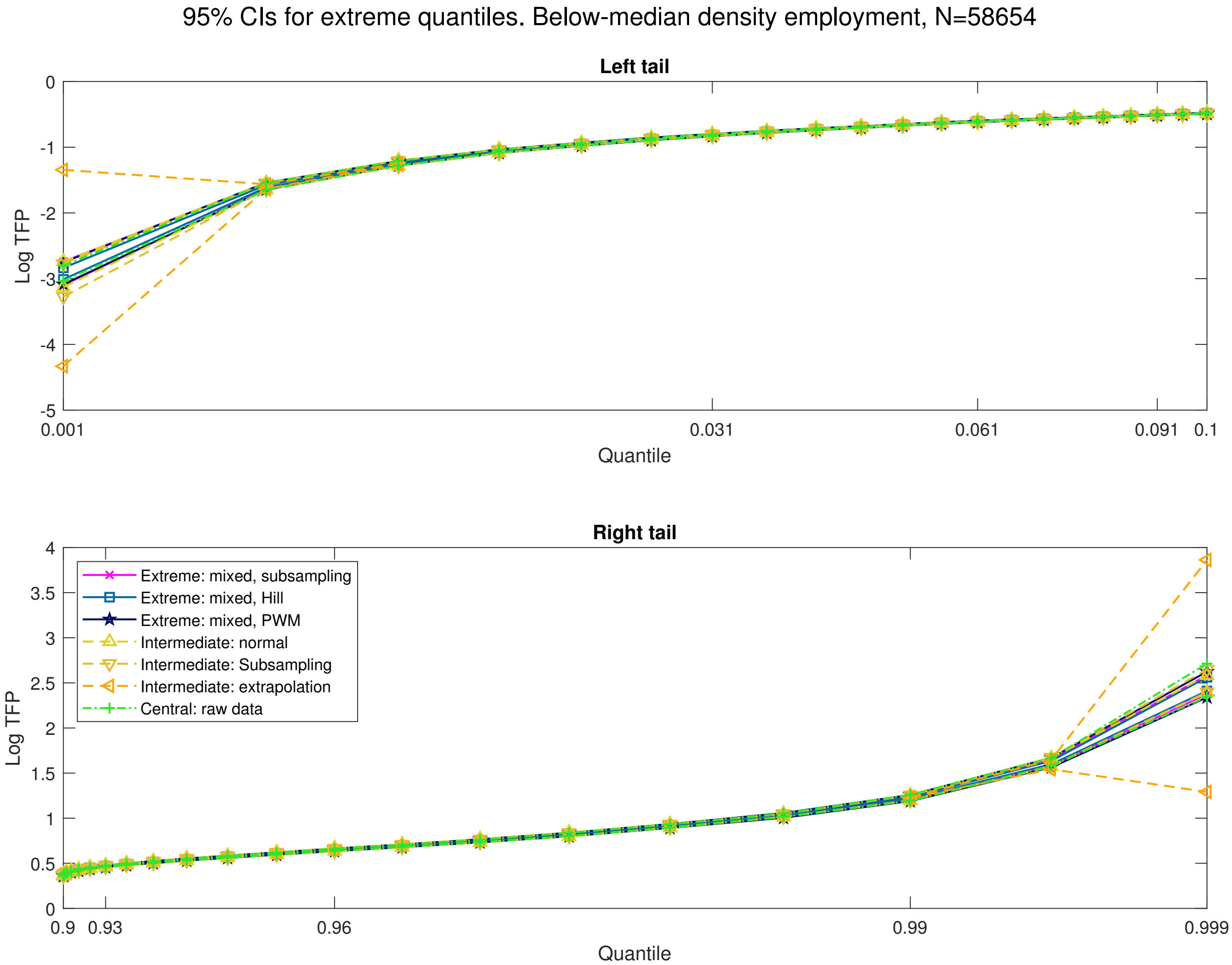


Figure 38: 95% confidence intervals for extreme quantiles, ABMED data, full sample. CIs as in section 5 and OA.4

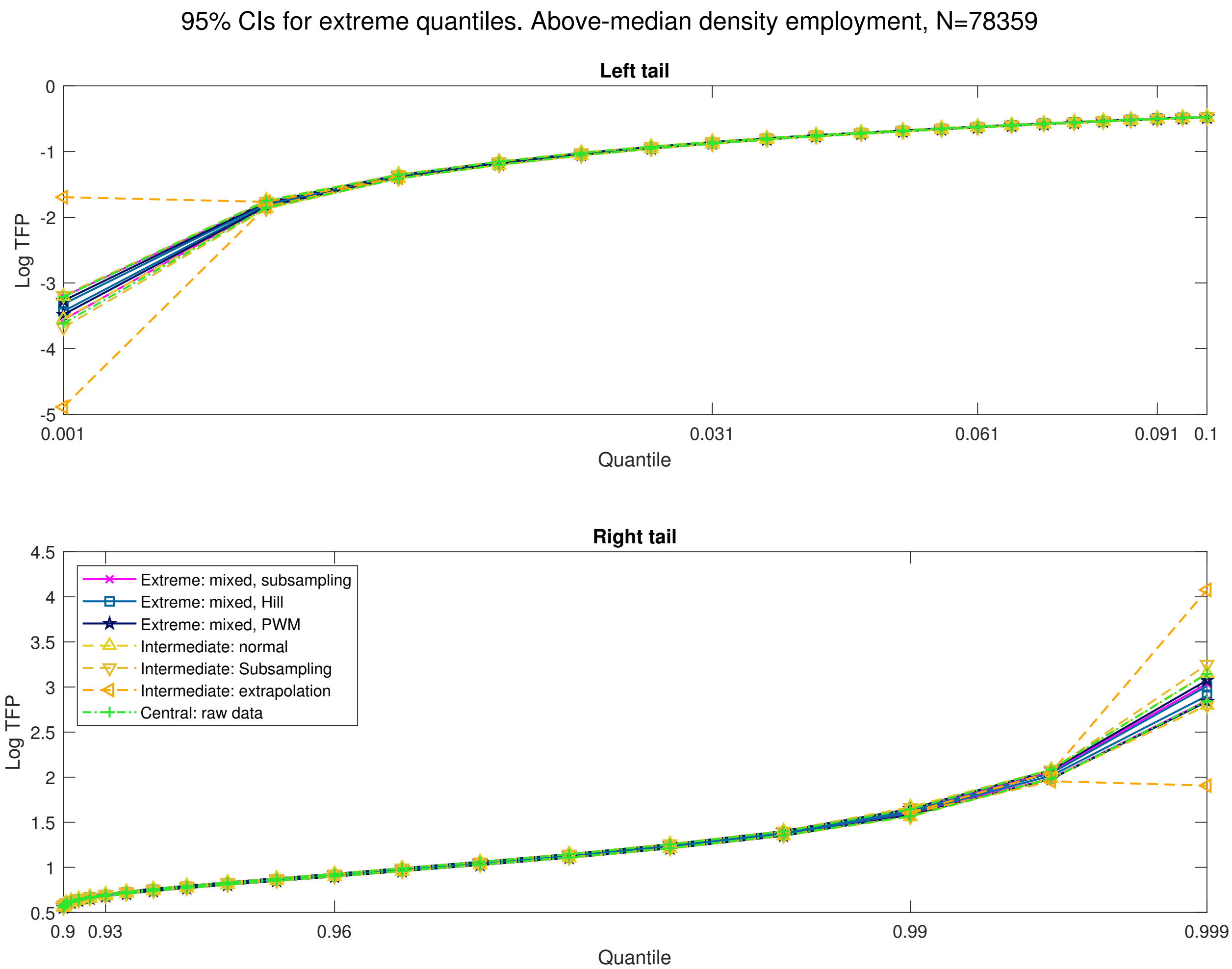


Figure 39: 95% confidence intervals for extreme quantiles, AAMED data, full sample. CIs as in section 5 and OA.4

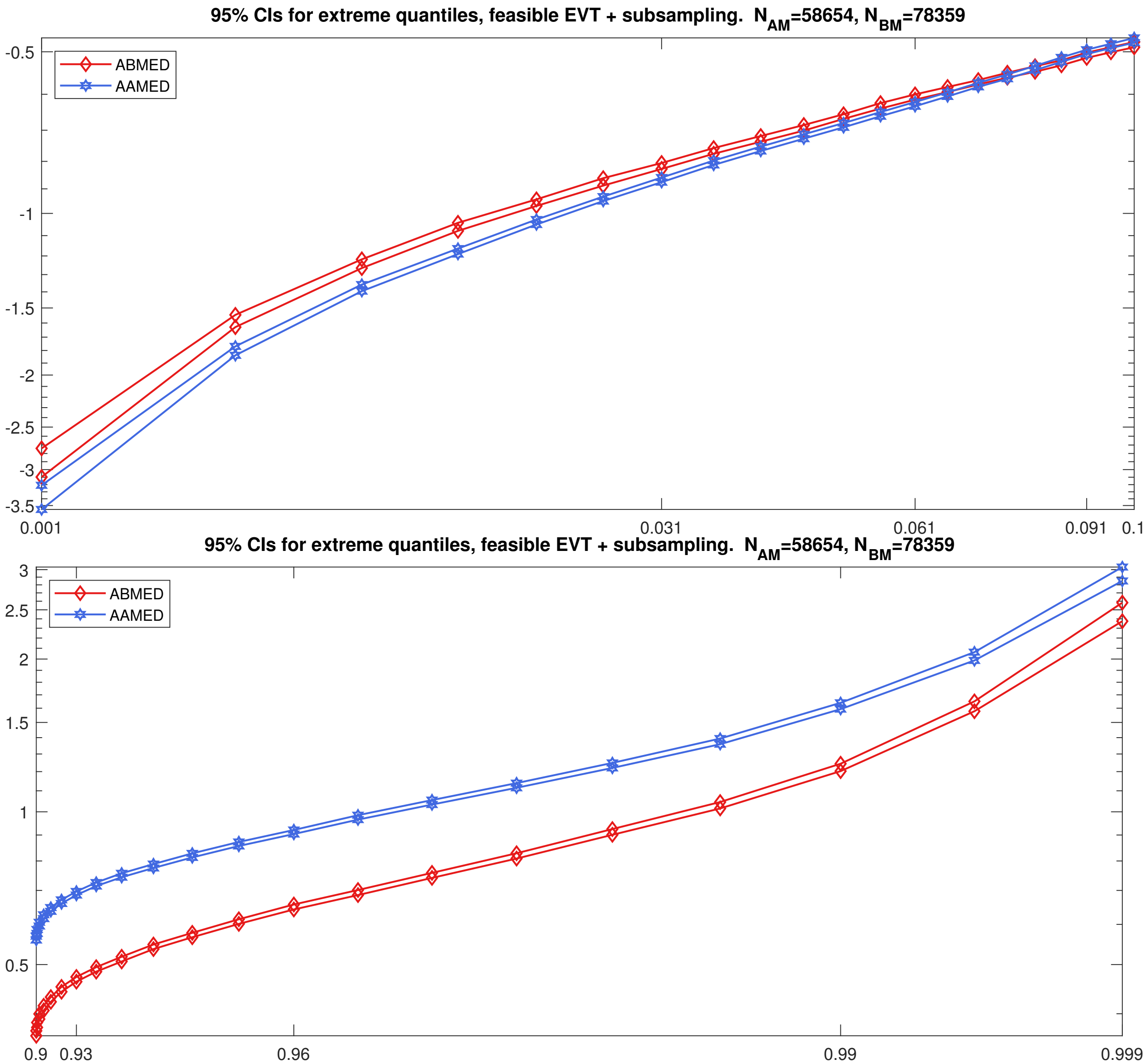


Figure 40: 95% confidence intervals for extreme quantiles of AAMED and ABMED distributions. CIs based on theorem 4.3 with critical values estimated by subsampling

References

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. ISSN 0012-9682. doi: 10.3982/ecta8442.

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

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

J. R. M. Hosking and J. R. Wallis. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):339–349, 1987. doi: 10.2307/1269343.

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