Inference on Extreme Quantiles of Unobserved Individual Heterogeneity
Online Appendix

Vladislav Morozov*

December 23, 2025

This online appendix provides additional theoretical and numerical results.

- Section OA.1: extends the paper's setting and results to unbalanced data.
- Section OA.2: presents a deterministic version of conditions (2) and (3) from theorem 3.3.
- Section OA.3: provides three additional simulation studies:
  - Analyzing the choice of extreme order confidence intervals tuning parameters (r, q) and evaluating the tuning parameter rule of thumb proposed in the main text.
  - Evaluating corrected estimators for extreme quantiles.
  - Examining the convergence speed to the normal limit in theorem 3.3.

Contents

OA.1 Unbalanced Data OA-2
OA.2 Deterministic Conditions For Intermediate Order Tail Equivalence Conditions OA-3
OA.3 Additional Simulation Results OA-4
OA.3.1 Choice of Tuning Parameters in Extreme CIs OA-5
OA.3.2 Performance of Corrected Estimators For Quantiles OA-6
OA.3.3 Quality of Intermediate Approximation OA-12

*morozov@uni-bonn.de; Institute of Finance and Statistics, University of Bonn

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, as is the case in the empirical application of section 6. In a meta-analysis setting, individual studies may have varying sample sizes.

It is straightforward to handle unbalanced data settings with the results of the main text, as we show in this section.

Setting with unobservable data 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, θ is distribution according to F. For each T the estimation noise $\varepsilon_{i,T}$ is distributed according to some distribution function G_T; the family $\{G_T\}$ is assumed to satisfy assumption 3. Interest still lies in extreme quantile of F.

Expressing (OA.1.1) in balanced- sample form To conduct inference on extreme quantiles of F under the data generating process (OA.1.1), we represent (OA.1.1) as a special case of the setup studied in the main text.

Define T and $\{\lambda_i\}_{i=1}^N$ so that

$$T_i = \lambda_i T, \quad \lambda_i \geq 1, \quad T \to \infty.$$

By construction $T_i \geq T$; T can be interpreted as a minimal sample size.

We assume that T_i is random. Define $\tilde{\varepsilon}_{i,T} = \lambda_i^{-p}\varepsilon_{i,T_i}$. With this definition (OA.1.1) can be expressed exactly as in eq. (1) in the main text:

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\tilde{\varepsilon}_{i,T}. \tag{OA.1.2}$$

This setup can be intuitively viewed in a hierarchical manner: first θ_i is drawn, then the observed sample size T_i (equivalently λ_i) is drawn for i. Last ε_{i,T_i} is drawn from G_{T_i}. The components are combined into the observable $\vartheta_{i,T}$ according to (OA.1.1).

Transferring main text results To transfer the results of the main text, let $\tilde{G}_T$ be the cdf of $\tilde{\varepsilon}_{i,T}$. One can apply all of the results in the main text by simply replacing G_T with $\tilde{G}_T$ in their statements, as (OA.1.2) suggests.

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. Then the same bound applies to transformed noise, meaning that $\sup_T \mathbb{E}|\tilde{\varepsilon}_{i,T}|^\beta < \infty$. This bound follows from the fact that $\tilde{\varepsilon}_{i,T}$ $\lambda_i^{-p} \leq 1$. 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 (Randomness of T_i)**.** 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 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 $U_1, \ldots, U_N$ be iid Uniform[0, 1]. Let $\delta_{k,N,T}$ be such that it holds that $\delta_{k,N,T} \to 0$ and*

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

as $T, N(T), k(N) \to \infty$, $k(N) = o(N)$. Also let c_N be a monotonic sequence of constants such that $c_N \neq 0$ eventually. Define

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

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.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

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

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

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

The inequality above holds because 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. The last line follows by the assumption of the proposition.

We proceed identically 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 it holds with probability approaching 1 that

$$\sup_{u \in [0, \epsilon]} \{\ldots\} \leq \sup_{u \in \left[0, 1 - U_{k,N}\right]} \{\ldots\} \leq \inf_{u \in \left[0, U_{k,N}\right]} \{\ldots\} \leq \inf_{u \in [0, \frac{k}{N}(1 - \delta_{k,N,T})]} \{\ldots\}.$$

These inequalities imply that on event A_N both the random supremum and the random infimum converge to zero. Since $P(A_N) \to 1$, we conclude convergence i.p. □

OA.3 Additional Simulation Results

In this section, we extend the analysis of the Monte Carlo study of section 5 in the main text. First, in section OA.3.1 we provide simulation support for the rule of remark 8 for choosing the

tuning parameters of extreme-order CIs. Second, in section OA.3.2 we consider the performance of the quantile estimators proposed in section 4. Finally, section OA.3.3 shows that convergence in the intermediate-order theorem 4.5 may be slow, driving relatively poor performance of the corresponding CIs.

OA.3.1 Choice of Tuning Parameters in Extreme CIs

The extreme-order CIs of example 3 have two tuning parameters (r, q):

(1) r controls the order statistic $\vartheta_{N-r,N,T}$ in the numerator of statistic (6). Intuitively, the CI uses that order statistic as a base "anchor" point for the CI (note that the CI does not include $\vartheta_{N-r,N,T}$ if both critical values are positive).

(2) q controls the order statistic $\vartheta_{N-q,N,T}$ in the denominator of statistic (6). This denominator may be interpreted as a random scaling factor for the statistic.

In this section we provide some suggestions for selecting (r, q), along with simulation evidence regarding the impact of these choices.

From a theoretical point of view, the problem of choosing (r, q) is challenging. Any fixed choice of (r, q) leads to a valid limit in theorem 4.3. Moreover, there does not exist a general consistent rule for selecting "appropriate" values for (r, q) (Müller and Wang, 2017, p. 1431).

Given the above challenge, we propose a selection rule based on simulations and some heuristic considerations. Let τ be the quantile of interest. Let $\tau \geq 1/2$ without loss of generality. As specified in remark 8, our recommendation for (r, q) takes the following form:

- Let $(1-\tau)N \leq 100$ (fewer than 100 points available to the right of the sample τth quantile). For q, we suggest values in the range 2-4 in smaller cross-sections ($N \lesssim 2000$) for CIs with simulated critical values; and $q = 2$ with subsampled critical values. A value up to 30 may be used in larger cross-sections ($N \gtrsim 10000$). For r, we suggest $r = \lfloor (1-\tau)N \rfloor$. This choice matches $\vartheta_{N-r,N,T}$ to the corresponding order statistic.
- If $(1-\tau)N > 100$, we suggest $q = 2$ and $r = 0$.

Four remarks are in order. First, in a strict theoretical sense, matching $\vartheta_{N-r,N,T}$ to the sample τth quantile is somewhat improper in the framework of theorem 4.3. While $r \leq 100$ under the rule regardless of N, this choice of r is sample size-dependent. However, this rule shows favorable simulation performance (see below). It is also intuitive — it is motivated by the idea of constructing the CI based on the nearest order statistic. Second, recall that we recommend using extreme-order CIs only if the quantile of interest τ satisfies $(1-\tau)N \leq 100$ (section 5). Thus, we recommend matching $\vartheta_{N-r,N,T}$ to the sample τth quantile whenever we recommend extreme approximations. Third, we find that CIs with simulated critical values are less sensitive to the choice of q in the region with $(1-\tau)N \leq 100$ than CIs with subsampled critical values. One may thus use larger values of q with simulated critical values and obtain shorter CIs. These recommendations are reflected in the tuning parameters chosen in section 5. Fourth, if one wishes to apply the extreme CIs even if $(1-\tau)N > 100$, we suggest basing the intervals on the sample maximum — setting $r = 0$.

To evaluate the above rule, we conduct an additional simulation study. The data-generating process is as in section 5, to which we refer for details. We evaluate the coverage and length properties extreme-order CIs, both with subsampled and simulated critical values. We consider several values of q spaced between $q = 2$ and $q = 28$. We also contrast the performance of setting $r = 0$ and matching r to the sample τth quantile.

The results are presented in figures OA.1-OA.4. Specifically, figures OA.1-OA.3.1 report the coverage and length properties for a range of values of q. Throughout, we match r to the sample quantiles. Figures OA.3-OA.4 instead contrast setting $r = 0$ vs. matching r to the sample quantile. To see how the choices of r and q interact, we report the results for $q = 2$ and $q = 4$.

Overall, we find that the above rule leads to correct coverage. In particular, if $(1 - \tau)N \leq 100$, the extreme CIs with the above (r, q) have favorable length properties and are the only viable option for inference. In contrast, if $(1 - \tau)N > 100$, the length of extreme CIs grows as τ falls. This effect is due to the approximation of theorem 4.3 becoming progressively less informative towards the center of the distribution. In this region, the extreme CIs are an order of magnitude longer than the central-order CI with analytical correction (e.g. compare figs. OA.3 to fig. 1 in the main text). Since central-order approximations have correct coverage (under their rate conditions on (N, T)), we recommend using them when $(1 - \tau)N > 100$, see section 5.

An incorrect choice of (r, q) may lead to undercoverage in two scenarios:

(1) Drop in coverage for $\tau \geq 1 - 1/N$ (see $\tau \geq 0.995$ for $N = 200$ on fig. OA.1). This issue is driven by setting q too large.

(2) Drop in coverage for lower τ, particularly if $(1 - \tau)N > 100$ (see $\tau \leq 0.99$ for $(N, T) = (10000, 10)$ on fig OA.3). This issue occurs if r is set too high and the corresponding tail equivalence conditions do not hold.

The first issue arises as the choice of q involves a length-robustness trade-off (e.g., figure OA.1). Choosing a smaller value of q leads to a smaller denominator and hence more variability in statistic (6). This increased variability leads to larger critical values. The associated CIs have higher coverage at the price of being longer. This increased variability is important to cover quantiles that are "outside the range of the data", that is, when $\tau \leq 1 - 1/N$.

The second issue affects CIs with $r = \lfloor (1 - \tau)N \rfloor$ for more central τ. In this case, the behavior of $\vartheta_{N-r,N,T}$ is better described by theorem 3.3 than theorem 3.1. The corresponding rate conditions on (N, T) are progressively stricter as τ falls (remarks 3-4). A larger value of q may partially offset the issue by increasing interval length (figure OA.3).

OA.3.2 Performance of Corrected Estimators For Quantiles

In this section we assess the performance of the adjusted quantile estimators proposed in example 4 and in section 4.3 of the main text. We again work in the setup of section 5 in the main text.

Figure OA.5 summarizes our results. The performance of the estimators is measured in terms of mean squared error (MSE). As a benchmark for performance, we use the simple quantile estimator

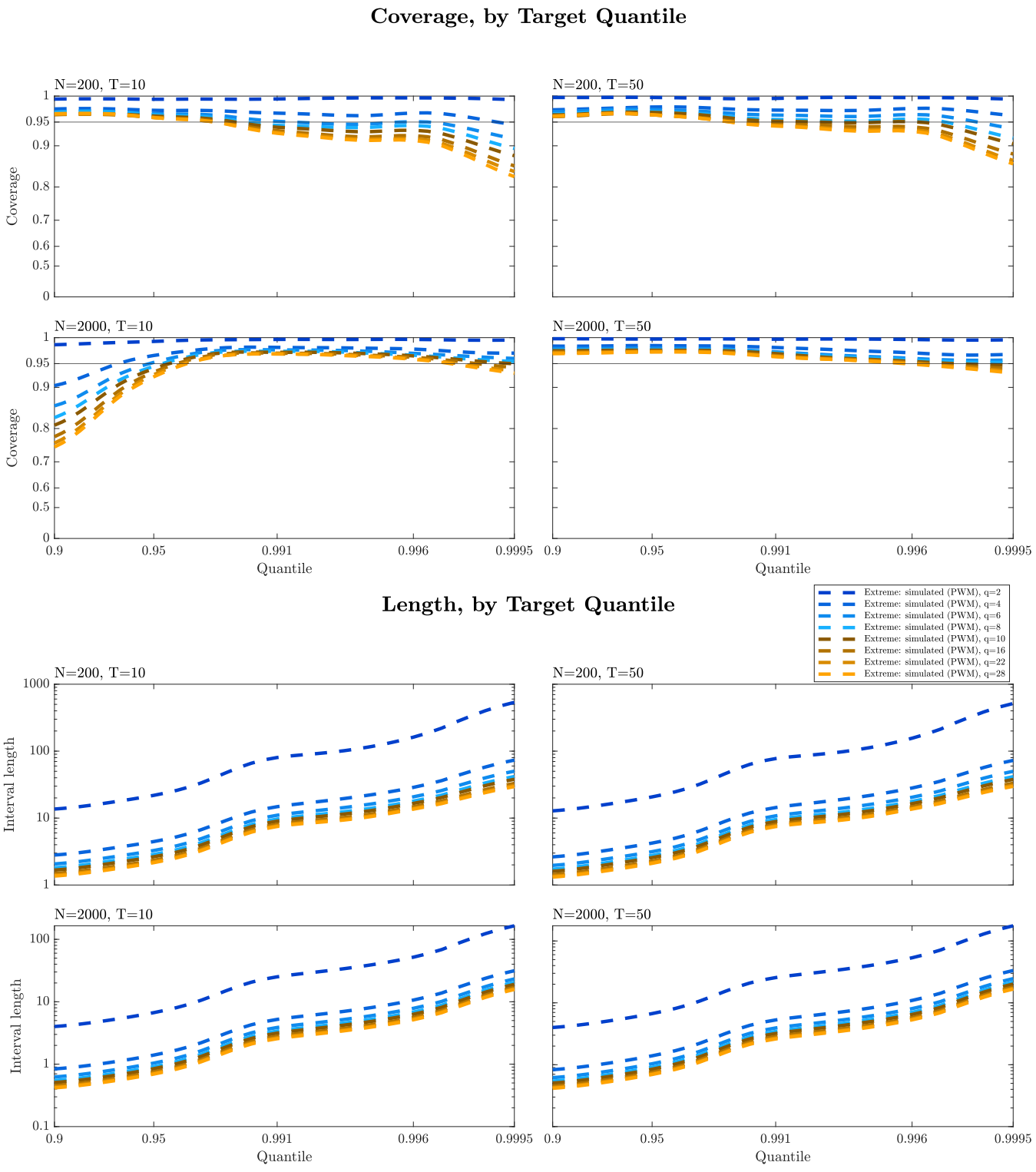

Figure OA.1: Impact of choice of denominator parameter q on coverage and length of a 95% confidence interval based on an extreme value approximation with simulated critical values. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

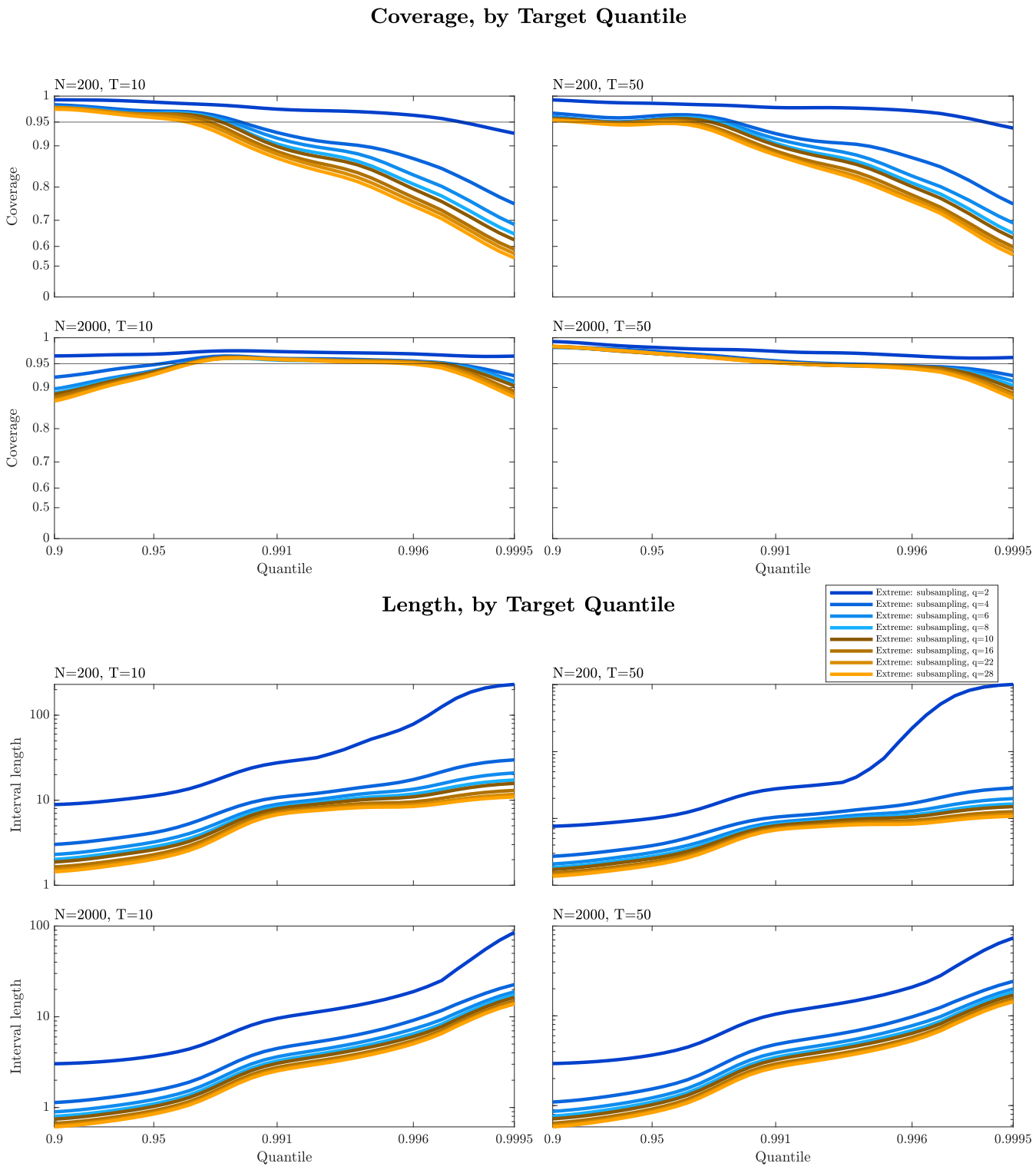

Figure OA.2: Impact of choice of denominator parameter q on coverage and length of a 95% confidence interval based on extreme value approximations with subsampled critical values. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

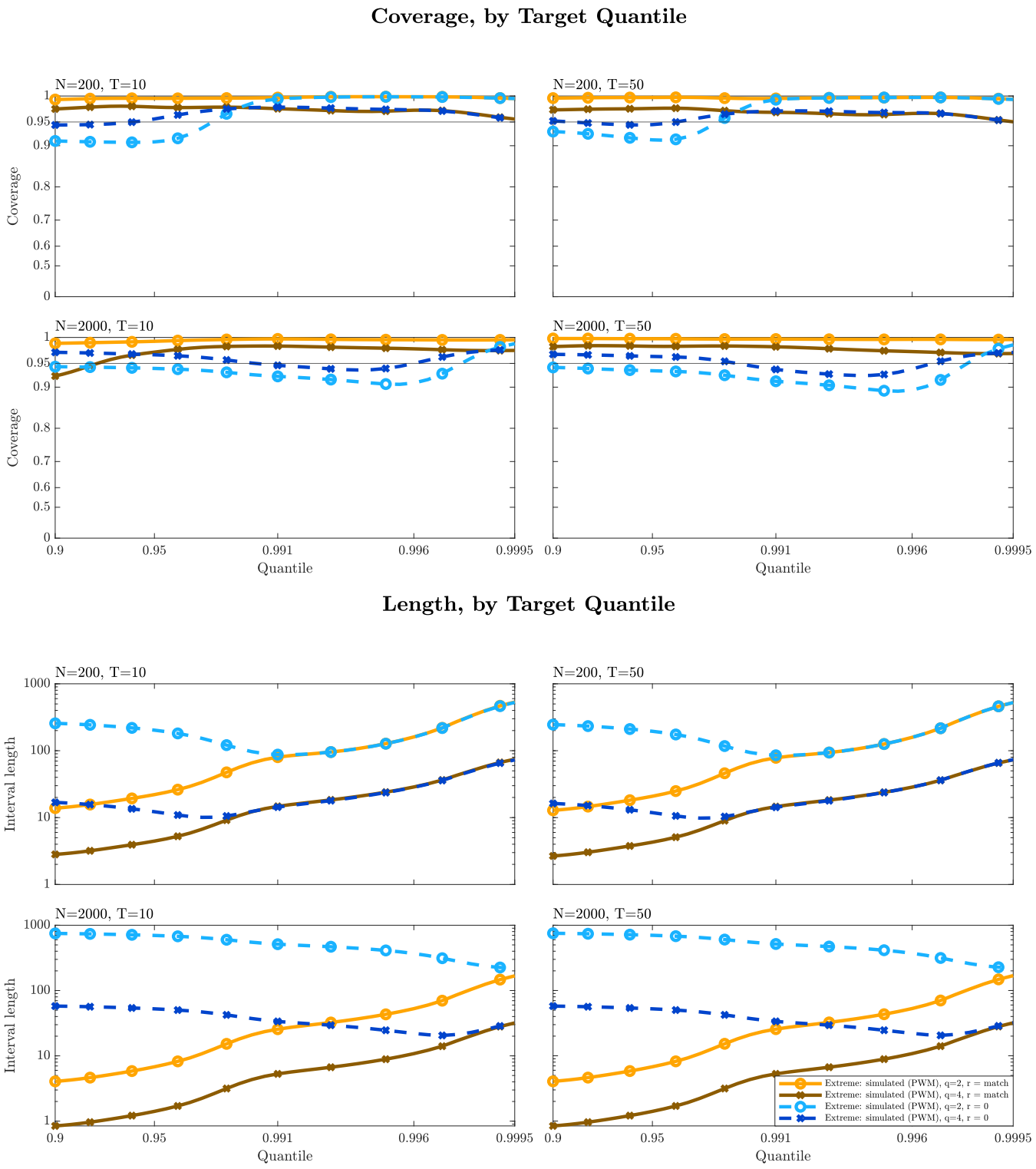

Figure OA.3: Impact of tuning parameter choice on coverage and length of a 95% confidence interval based on extreme value approximations with simulated critical values, Note: nonlinear x- and y-axes.

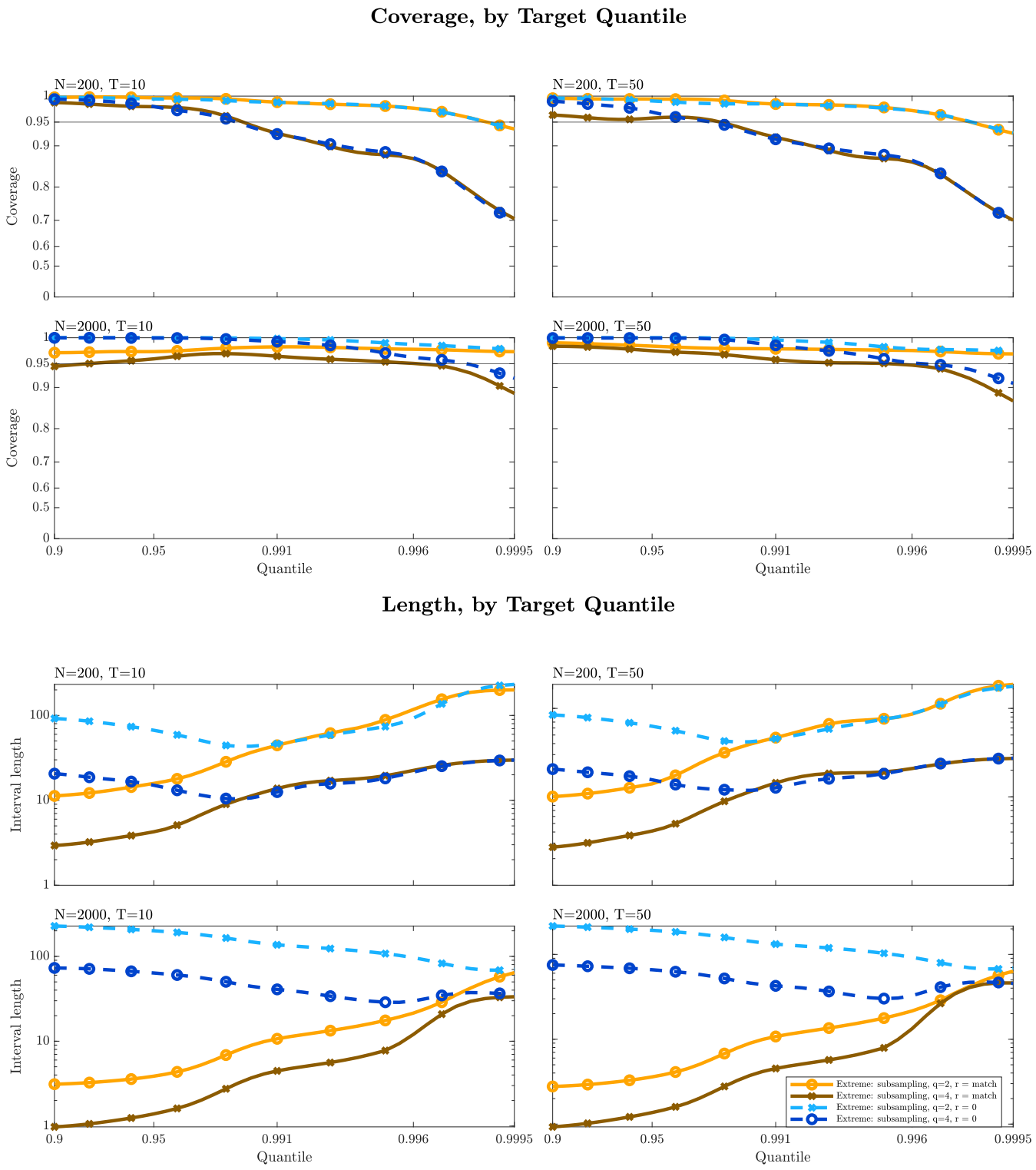

Figure OA.4: Impact of tuning parameter choice on coverage and length of a 95% confidence interval based on extreme value approximations with subsampled critical values. Note: nonlinear x- and y-axes.

obtained by linearly interpolating the two nearest order statistics. For each estimator considered, we report the ratio of the estimator's MSE to the MSE of the benchmark estimator. Values greater than 1 mean that the estimator performs worse than the benchmark, while values below 1 signify better performance.

Our recommendation for the estimation strategy depends on the target quantile τ:

(1) For quantiles below the 0.99th, we suggest using the central-order estimator of theorem 4.6. For such quantiles, the corrected estimator offers an improvement on the benchmark estimator. The improvement is particularly large if N is larger and stronger estimation noise is present.

(2) Beyond the 0.99th quantile, we recommend the "textbook" intermediate-order extrapolation estimator. It offers fairly sizable improvements in the MSE for such quantiles for all the simulation settings considered.

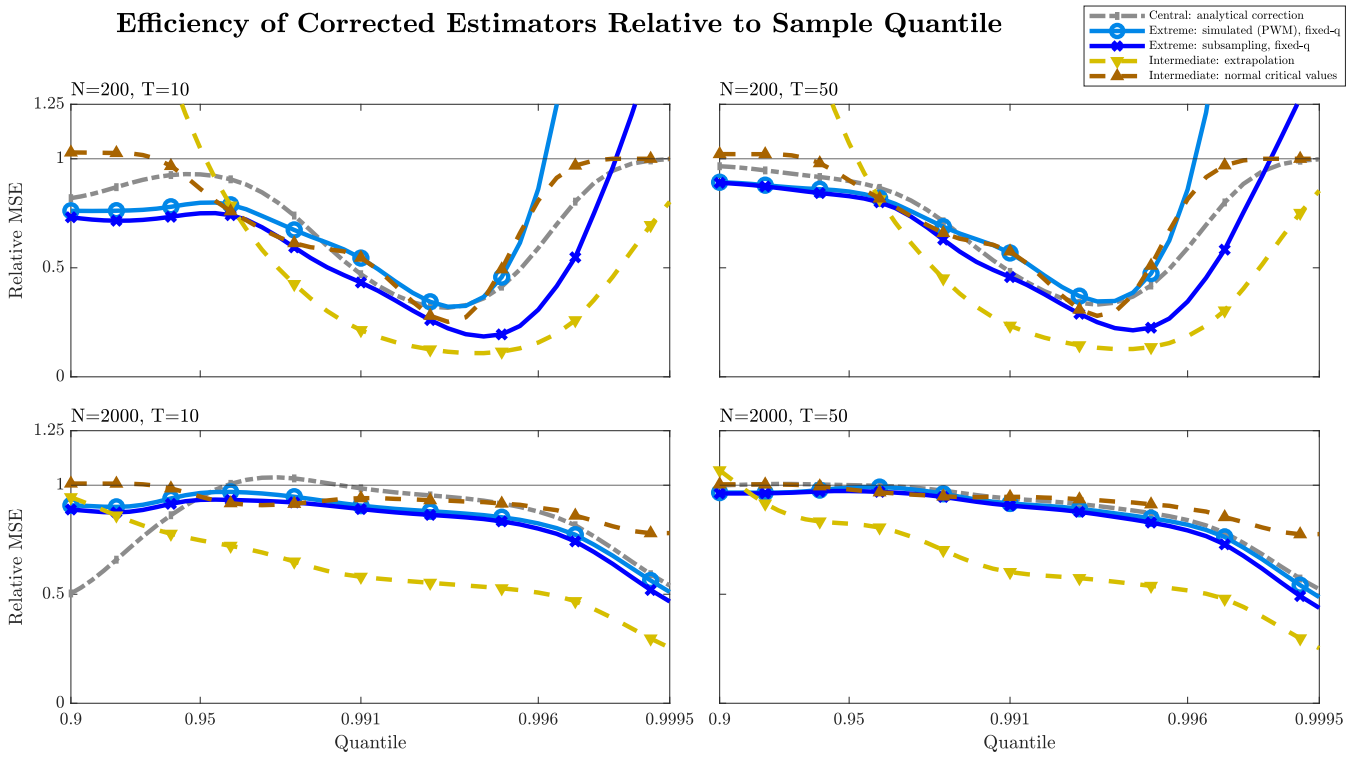

Figure OA.5: MSE of various corrected estimators relative to the MSE of sample quantiles

One may also use the extreme-order corrections of example 4, particularly for larger values of N. These offer uniform improvements on the benchmark estimator for $N = 2000$. For $N = 200$, the two corrections perform worse than the benchmark above the .995th quantile for heavier-tailed distributions (those with the EV index $\gamma > 0$ — $F_{Fr,\kappa}, t_3$).

OA.3.3 Quality of Intermediate Approximation

The confidence intervals based on theorem 4.5 perform poorly in the simulation study of section 5, with somewhat better coverage rates for $N = 2000$. This issue in performance is due to two related factors (Müller and Wang, 2017):

(1) Convergence of intermediate-order order statistics to normal limits may be slow (see section 2.4 in de Haan and Ferreira (2006) for some quantitative results).

(2) The approximation may be poor if $k = k(N)$ is not large.

To quantify the extent of these issues, we perform an additional simulation. The data-generating process is as in section 5. The coefficients θ_i are distributed according to t_3, where t_3 is Student's t-distribution with 3 degrees of freedom. We only consider the noiseless case with $u_{it} = 0$. The values of N considered are $N = 200, 2000, 10000, 50000$.

On figure OA.6, we report the distribution of the statistic of theorem 4.5 for four target quantiles: $\tau = 0.9, 0.95, 0.99, 0.999$. The corresponding value of k is determined as $k = \lfloor (1-\tau)N \rfloor$. We also plot the PDF of the limiting $N(0,1)$ distribution.

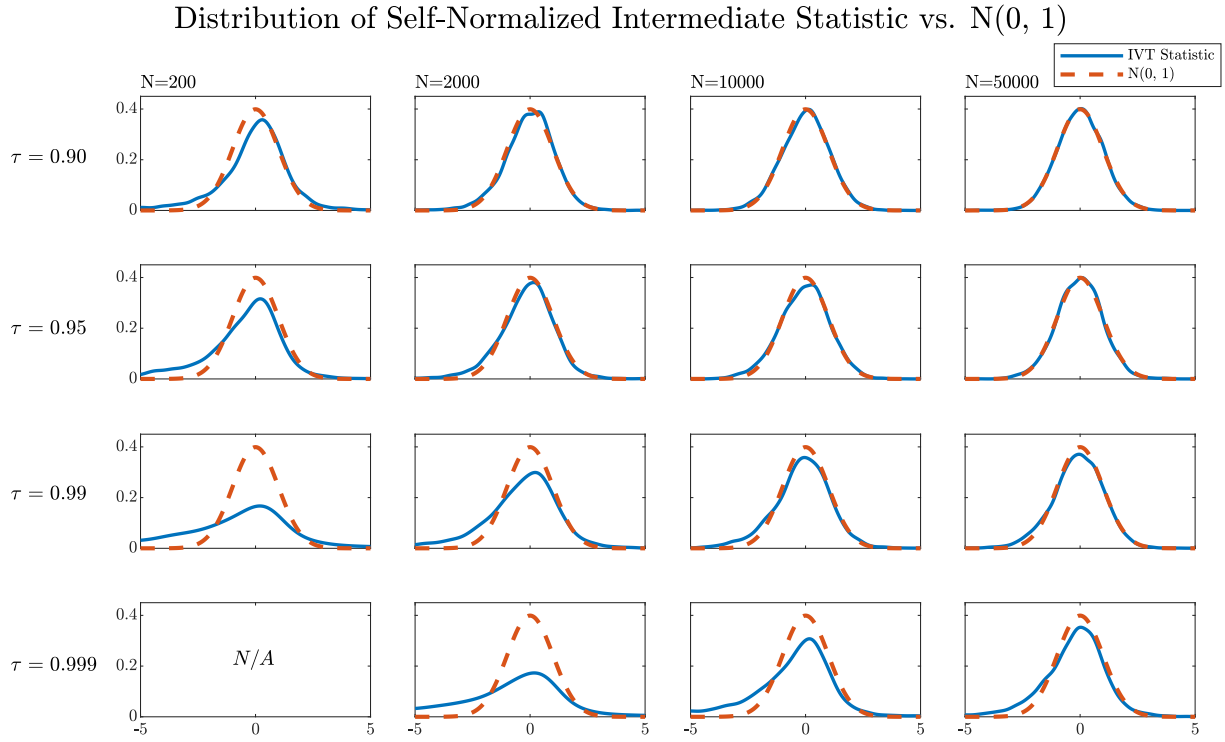

Figure OA.6: Distribution of the self-normalized intermediate order statistic of theorem 4.5 vs. the limit $N(0,1)$ distribution. $\theta \sim t_3$, noiseless data. Rows index target quantiles τ, column index cross-sectional sample sizes N.

The simulation results demonstrate the above two convergence issues. First, the first two rows of fig. OA.6 show that convergence to normality is quite slow. Even for $N = 10000$, the distribution of the statistics for $\tau = 0.9$ and $\tau = 0.99$, shows noticeable left skew and a correspondingly heavier left tail. Second, the statistic generally shows significant non-normality if $k \leq 100$ (equivalently, if $(1 - \tau)N \leq 100$). This non-normality is most evident on the bottom two rows of fig. OA.6.

As a consequence, we can only recommend using intermediate-order approximations in large cross-sections — a conclusion similar to that of Müller and Wang (2017).

References

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

Müller U. K. and Wang Y. (2017). Fixed-k Asymptotic Inference About Tail Properties. *Journal of the American Statistical Association*, 112(519):1334–1343. ISSN 1537274X. doi: 10.1080/01621459.2016.1215990.