Inference on Extreme Quantiles of Unobserved Individual Heterogeneity
Online Appendix

Vladislav Morozov*

March 13, 2025

This online appendix provides additional theoretical, numerical, and empirical 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: computes the rate conditions for theorems 3.1 and 3.3 for several general distributions of θ_i and $\varepsilon_{i,T}$.
- Section OA.4: provides four additional simulation studies:
  - Expanding section 5 by incorporating more distributions for θ and idiosyncratic disturbances.
  - Analyzing the effect of choices for 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.
- Section OA.5: presents additional empirical results, including detailed results for all confidence interval methods of the paper.

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

†Acknowledgments: I profoundly grateful to Andrea Sy for her generosity and knowledge of the CBI data. 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, 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

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

$$T_i = \lambda_i T, \quad \lambda_i \geq 1, \quad T \to \infty. \tag{OA.1.2}$$

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:

$$\vartheta_{i,T} = \theta_i + \frac{1}{T^p}\tilde{\varepsilon}_{i,T}. \tag{OA.1.3}$$

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.1.3) 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 $U_1, \dots, U_N$ 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

$$s_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1 - \frac{k}{N}(1+\delta) - u\right) - F^{-1}\left(1 - \frac{k}{N}(1-\delta)\right) + \frac{1}{T^p}G_T^{-1}(u)\right), \tag{OA.2.2}$$

$$S_{s,k,N,T}(u,\delta) = \frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1 - \frac{k}{N}(1-\delta) + u\right) - F^{-1}\left(1 - \frac{k}{N}(1+\delta)\right) + \frac{1}{T^p}G_T^{-1}(1-u)\right) \tag{OA.2.3}$$

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

If for some $\epsilon \in (0,1)$

$$\sup_{u\in[0,\epsilon]} s_{s,k,N,T}(u, \delta_{k,N,T}) \to 0, \tag{OA.2.4}$$

$$\inf_{u\in\left[0,\frac{k}{N}(1-\delta_{k,N,T})\right]} S_{s,k,N,T}(u, \delta_{k,N,T}) \to 0, \tag{OA.2.5}$$

then

$$\sup_{u\in[0,\epsilon]} \frac{\sqrt{k}}{c_N}\left(F^{-1}(1 - U_{k,N} - u) - F^{-1}(1 - U_{k,N}) + \frac{1}{T^p}G_T^{-1}(u)\right) \xrightarrow{p} 0, \tag{OA.2.6}$$

$$\inf_{u\in\left[0,U_{k,N}\right]} \frac{\sqrt{k}}{c_N}\left(F^{-1}(1 - U_{k,N} + u) - F^{-1}(1 - U_{k,N}) + \frac{1}{T^p}G_T^{-1}(1-u)\right) \xrightarrow{p} 0. \tag{OA.2.7}$$

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

$$A_N = \left\{\left|\frac{N}{k}U_{k,N} - 1\right| \leq \delta_{k,N,T}\right\}. \tag{OA.2.8}$$

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

$$\inf_{u\in\left[0,U_{k,N}\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1-U_{k,N}+u\right)-F^{-1}\left(1-U_{k,N}\right)+\frac{1}{T^p}G_T^{-1}\left(1-u\right)\right) \tag{OA.2.9}$$

$$=\inf_{u\in\left[0,U_{k,N}\right]}\frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}+u\right)-F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}\right)+\frac{1}{T^p}G_T^{-1}\left(1-u\right)\right) \tag{OA.2.10}$$

$$\leq \inf_{u\in[0,\frac{k}{N}(1-\delta_{k,N,T})]} S_{s,k,N,T}(u,\delta_{k,N,T}) \tag{OA.2.11}$$

$$\to 0. \tag{OA.2.12}$$

The inequality above holds as 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 in the exact same manner for the supremum: on A_N it holds that

$$\sup_{u\in[0,\epsilon]}\frac{\sqrt{k}}{c_N}\left(F^{-1}(1-U_{k,N}-u)-F^{-1}\left(1-U_{k,N}\right)+\frac{1}{T^p}G_T^{-1}\left(u\right)\right) \tag{OA.2.13}$$

$$=\sup_{u\in[0,\epsilon]}\frac{\sqrt{k}}{c_N}\left(F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}-u\right)-F^{-1}\left(1-\frac{k}{N}\frac{N}{k}U_{k,N}\right)+\frac{1}{T^p}G_T^{-1}\left(u\right)\right) \tag{OA.2.14}$$

$$\geq \sup_{u\in[0,\epsilon]} s_{s,k,N,T}(u,\delta_{k,N,T}) \tag{OA.2.15}$$

$$\to 0. \tag{OA.2.16}$$

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

$$\sup_{u\in[0,\epsilon]}\{\dots\}\leq \sup_{u\in\left[0,1-U_{k,N}\right]}\{\dots\}\leq \inf_{u\in\left[0,U_{k,N}\right]}\{\dots\}\leq \inf_{u\in[0,\frac{k}{N}(1-\delta_{k,N,T})]}\{\dots\}. \tag{OA.2.17}$$

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 Example Rate Conditions

To build intuition for the sufficient rate conditions of propositions 3.2 and 3.4, we consider three explicit but general examples.

OA.3.1 Specification of Examples

Specifications of F We consider three specifications for F that differs in the sign of the extreme value index γ:

$\gamma > 0$ $F_{Fr,\kappa}(\theta)$ is defined as

$$F_{Fr,\kappa}(\theta) = 1 - (\theta + 1)^{-\kappa}, \kappa > 0, \theta \in [0, \infty). \tag{OA.3.1}$$

$F_{Fr,\kappa}$ satisfies assumption 2 with $\gamma = 1/\kappa > 0$. Let $a_N = N^{1/\kappa} - 1, b_N = 0$. Then $\theta_{N,N}/(N^{1/\kappa} - 1)$ is asymptotically distributed as a Fréchet random variable.

$\gamma = 0$ Let $F = F_{Gu,\lambda}$ be the exponential distribution with parameter λ.

$$F_{Gu,\lambda}(\theta) = 1 - e^{-\lambda x}, \quad \theta \in [0, \infty). \tag{OA.3.2}$$

$F_{Gu,\lambda}$ 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 according to the Gumbel distribution.

$\gamma < 0$ Let $\theta_F < \infty$ and define $F_{W,\alpha}$ as

$$F_{W,\alpha}(\theta) = 1 - ((\theta_F - \theta)/\theta_F)^{\alpha}, \quad \alpha > 0, \quad \theta \in [0, \theta_F], \quad \theta_F = F^{-1}(1) < \infty. \tag{OA.3.3}$$

$F_{W,\alpha}$ 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.

The above examples for F are general in the following sense. If F satisfies assumption 2 with $\gamma \neq 0$, it differs from one of the examples cdfs only by a slowly varying function (possible in a generalized sense for $\gamma = 0$, see de Haan and Ferreira (2006))

Specifications for G_T For G_T, we consider two specifications denoted $G_{\beta,T}$ and $G_{Normal,T}$:

1. $G_{\beta,T}$ is a symmetric distribution with median μ_T and moments of order $< \beta$. It is defined by its density
$$g_{\beta}(x) = (\beta/2)(1 + |x - \mu_T|)^{-\beta-1}, \beta > 0, x \in \mathbb{R}, \tag{OA.3.4}$$
where μ_T is a bounded sequence.
2. $G_{Normal,T}$ is a $N(\mu_T, \sigma_T^2)$ distribution where (μ_T, σ_T^2) is a bounded sequence.

These examples are the leading cases of distributions satisfying the assumptions of proposition 3.2.

OA.3.2 Examples of Rate Conditions for EVT (Theorem 3.1)

We first compute rate conditions that ensure that conditions (TE-Inf) and (TE-Sup) hold in each example. A detailed verification can be found below in subsection OA.3.4.

Example 1 ($\gamma > 0$)**.** In this example, there is no restriction on magnitudes of N and T if $\varepsilon_{i,T}$ has more than κ moments. If $G_T = G_{\beta,T}$, 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. This result is intuitive, if F has a heavier tail, it will be more pronounced in the data and dominate the tail of G_T.

Example 2 ($\gamma = 0$)**.** 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 ($\gamma < 0$)**.** 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

$$\frac{N^{1/\beta-\gamma}(\log(T))^{1/\beta}}{T^p} \to 0. \tag{OA.3.5}$$

If $G_T = G_{normal,T}$, the conditions can be equivalently written as

$$\frac{N^{-\gamma}\sqrt{\log(N)}}{T^p} \to 0. \tag{OA.3.6}$$

Note the similarity to the rate conditions of proposition 3.2.

OA.3.3 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. Considering the structure of the conditions (2) and (3) in the main text yields the following rate conditions in our examples.

Example 1 ($\gamma > 0$, 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 $\nu > 0$ it holds that $N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta-1/\kappa)+\nu/\beta}/T^p \to 0$. If $G_T = G_{normal,T}$, then it is sufficient that $N^{\delta/2+(1-\delta)(-1/\kappa)}\sqrt{\log(N)}/T^p \to 0$.

Example 2 ($\gamma = 0$, 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 $\nu > 0$ it holds that $N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta)+\nu/\beta}/T^p \to 0$. Similarly, if $G_T = G_{normal,T}$, it is sufficient that $N^{\delta/2}\sqrt{\log(N)}/T^p \to 0$.

Example 3 ($\gamma < 0$, 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:

$$G_{\beta,T} : \frac{N^{\delta/2(1+1/\beta)+(1-\delta)(-\gamma+1/\beta)+\nu/\beta}}{T^p} \to 0, \tag{OA.3.7}$$

$$G_{normal,T} : \frac{N^{\delta/2+(1-\delta)(-\gamma)}\sqrt{\log(N)}}{T^p} \to 0, \tag{OA.3.8}$$

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

OA.3.4 Verification of Examples

In this section we provide a detailed verification of the claims made in sections OA.3.2-OA.3.3.

Useful Expressions

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

Distributions The densities of the example distributions are given by

$$f_{Fr,\kappa}(\theta) = \kappa(\theta+1)^{-\kappa-1}\ , \tag{OA.3.9}$$

$$f_{Gu,\lambda} = \lambda e^{-\lambda x}, \tag{OA.3.10}$$

$$f_{W,\alpha}(\theta) = \frac{\alpha}{\theta_F}\left(\frac{\theta_F-\theta}{\theta_F}\right)^{\alpha-1}, \quad \theta \in [0,\theta_F]. \tag{OA.3.11}$$

Two specifications for cdf of estimation noise:

$$G_{\beta,T}(x) = \begin{cases} 1 - \frac{1}{2}(1+(x-\mu_T))^{-\beta}, & x \geq \mu_T \\ \frac{1}{2}(1-(x-\mu_T))^{-\beta} & x < \mu_T, \end{cases} \tag{OA.3.12}$$

$$G_{Normal,T}(x) = \Phi\left(\frac{x-\mu_T}{\sigma_T}\right)\ . \tag{OA.3.13}$$

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):

$$F^{-1}_{Fr,\kappa}(y) = (1-y)^{-1/\kappa} - 1, \qquad U_{F_{Fr,\kappa}}(y) = y^{1/\kappa} - 1, \tag{OA.3.14}$$

$$F^{-1}_{Gu,\lambda}(y) = -\frac{\log(1-y)}{\lambda}, \qquad U_{Gu,\lambda} = \frac{\log y}{\lambda}, \tag{OA.3.15}$$

$$F^{-1}_{W,\alpha}(y) = \theta_F - \theta_F(1-y)^{1/\alpha}, \qquad U_{F_{W,\alpha}}(y) = F^{-1}\left(1-\frac{1}{y}\right) = \theta_F - \theta_F y^{-1/\alpha}, \tag{OA.3.16}$$

$$G^{-1}_{\beta,T}(\tau) = \begin{cases} (2(1-\tau))^{-1/\beta} - 1 + \mu_T & \tau \geq \frac{1}{2}, \\ 1 - (2\tau)^{-1/\beta} + \mu_T & \tau < \frac{1}{2}. \end{cases} \tag{OA.3.17}$$

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

$$S_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}\left(1-\frac{1}{N\tau}+u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right) + \frac{1}{T^p}G_T^{-1}(1-u)\right), \tag{OA.3.18}$$

$$s_{\tau,N,T}(u) = \frac{1}{a_N}\left(F^{-1}\left(1-\frac{1}{N\tau}-u\right) - F^{-1}\left(1-\frac{1}{N\tau}\right) + \frac{1}{T^p}G_T^{-1}(u)\right). \tag{OA.3.19}$$

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

$$\limsup_{N,T\to\infty} \inf_{u\in\left[0,\frac{1}{N\tau}\right]} S_{\tau,N,T}(u) \leq 0. \tag{OA.3.20}$$

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

$$\liminf_{N,T\to\infty} \sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \geq 0. \tag{OA.3.21}$$

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

$$\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \leq \sup_{u\in[0,1-1/N\tau]} s_{\tau,N,T}(u) \leq \inf_{u\in[0,1/N\tau]} S_{\tau,N,T}(u). \tag{OA.3.22}$$

This implies that

$$\limsup_{N,T\to\infty} \sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \leq \liminf_{N,T\to\infty} \inf_{u\in\left[0,\frac{1}{N\tau}\right]} S_{\tau,N,T}(u). \tag{OA.3.23}$$

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

$$\lim_{N,T\to\infty} \sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) = \lim_{N,T\to\infty} \inf_{u\in\left[0,\frac{1}{N\tau}\right]} S_{\tau,N,T}(u) = 0. \tag{OA.3.24}$$

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 OA-8 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

$$u_{S,\tau,N,T} = \frac{1}{N\tau}\frac{1}{\log(T)+1} \in \left[0, 1 - \frac{1}{N\tau}\right] . \tag{OA.3.25}$$

We show that $S_{\tau,N,T}(u_{S,\tau,N,T}) \to 0$, where $S_{\tau,N,T}$ is defined in eq. (OA.3.18).

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.3.4, we obtain that

$$F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\right) = (N\tau)^{1/\kappa} - 1, \tag{OA.3.26}$$

$$F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) = \left(N\tau\frac{\log(T)+1}{\log(T)}\right)^{1/\kappa} - 1. \tag{OA.3.27}$$

Then as $N, T \to \infty$

$$\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\right) - F^{-1}_{Fr,\kappa}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right] \tag{OA.3.28}$$

$$= \tau^{1/\kappa} - \tau^{1/\kappa}\left(\frac{\log(T)+1}{\log(T)}\right)^{1/\kappa} \tag{OA.3.29}$$

$$\to 0 . \tag{OA.3.30}$$

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.3.4, we obtain that

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1 - u_{S,\tau,N,T}\right) \tag{OA.3.31}$$

$$= \frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1 - \frac{1}{N\tau}\frac{1}{\log(T)+1}\right) \tag{OA.3.32}$$

$$\sim \frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T}{N^{1/\kappa}T^p} . \tag{OA.3.33}$$

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

$$\frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p} \to 0. \tag{OA.3.34}$$

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

$$\frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{Normal,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p}+\frac{\mu_T}{N^{1/\kappa}T^p} \ . \tag{OA.3.35}$$

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

$$\frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p} \to 0. \tag{OA.3.36}$$

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

$$u_{s,\tau,N,T} = \frac{1}{N\tau}\frac{1}{\log(T)}. \tag{OA.3.37}$$

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

$$\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\right)-F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\frac{\log(T)+1}{\log(T)}\right)\right] \tag{OA.3.38}$$

$$= \tau^{1/\kappa}-\tau^{1/\kappa}\left(\frac{\log(T)}{\log(T)+1}\right)^{1/\kappa} \tag{OA.3.39}$$

$$\to 0 \ . \tag{OA.3.40}$$

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

$$\frac{1}{a_N}\frac{1}{T^p}G^{-1}_{\beta,T}\left(u_{s,\tau,N,T}\right) \tag{OA.3.41}$$

$$= \frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{\beta,T}\left(\frac{1}{N\tau}\frac{1}{\log(T)}\right) \tag{OA.3.42}$$

$$\sim \frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p}+\frac{\mu_T(\log(T))^{1/\beta}}{N^{1/\kappa}T^p} \ . \tag{OA.3.43}$$

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

$$\frac{N^{1/\beta-1/\kappa}(\log(T))^{1/\beta}}{T^p} \to 0. \tag{OA.3.44}$$

The condition is the same as for the infimum.

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

$$\frac{1}{N^{1/\kappa}-1}\frac{1}{T^p}G^{-1}_{Normal,T}\left(\frac{1}{N\tau}\frac{1}{\log(T)}\right) \sim \frac{\sqrt{\log(N)}}{N^{1/\kappa}T^p} \ , \tag{OA.3.45}$$

which again matches the condition derived for the infimum.

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

$$\limsup_{N,T\to\infty}\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) \le \liminf_{N,T\to\infty}\inf_{u\in\left[0,\frac{1}{N\tau}\right]} S_{\tau,N,T}(u) \ . \tag{OA.3.46}$$

If the above rate conditions on N and T hold, it holds that $\liminf_{N,T\to\infty}\sup_{u\in[0,\epsilon]} s_{\tau,N,T}(u) = 0, \limsup_{N,T\to\infty}\inf_{u\in[0,\frac{1}{N\tau}]} S_{\tau,N,T}(u) = 0$. 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.3.25) and (OA.3.37)) 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

$$\frac{N^{1/\beta}(\log(\cdots\log(T))^{1/\beta}}{N^{1/\kappa}T^{1/2}} \to 0\ . \tag{OA.3.47}$$

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

$$\frac{1}{N^{1/\kappa}-1}\left[F^{-1}_{Fr,\kappa}\left(1-\frac{1}{N\tau}\right) - F^{-1}_{Fr,\kappa}\left(1-\frac{1-c}{N\tau}\right)\right] = \tau^{1/\kappa} - \tau^{1/\kappa}(1-c)^{-1/\kappa} \neq 0, \tag{OA.3.48}$$

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

Example 2, page OA-8 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.3.25). Then since

$$F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\right) = \frac{\log(N\tau)}{\lambda}\ , \tag{OA.3.49}$$

$$F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) = \frac{\log(N\tau)+\log\left(\frac{\log(T)+1}{\log(T)}\right)}{\lambda}\ , \tag{OA.3.50}$$

we obtain that

$$\frac{1}{a_N}\left[F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\right) - F^{-1}_{Gu,\lambda}\left(1-\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right] = \frac{\log\left(\frac{\log(T)+1}{\log(T)}\right)}{\lambda} \to 0. \tag{OA.3.51}$$

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

$$\frac{1}{a_N T^p}G^{-1}_{\beta,T}\left(1-\frac{1}{N\tau}\left(1-\frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T(\log(T))^{1/\beta}}{T^p}\ . \tag{OA.3.52}$$

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

$$\frac{N^{1/\beta}(\log(T))^{1/\beta}}{T^{1/2}} \to 0. \tag{OA.3.53}$$

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

$$\frac{1}{a_N T^p} G^{-1}_{normal,T}\left(1 - \frac{1}{N\tau}\left(1 - \frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{\sqrt{\log(N)}}{T^p} + \frac{\mu_T}{T^p}, \tag{OA.3.54}$$

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

$$\frac{\sqrt{\log(N)}}{T^p} \to 0 \tag{OA.3.55}$$

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 OA-8 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.3.25). First examine the $F^{-1}_{W,\alpha}$ terms. Using the expressions for inverses given in section OA.3.4, we obtain

$$F^{-1}_{W,\alpha}\left(1 - \frac{1}{N\tau}\right) = \theta_F - \theta_F\left(\frac{1}{N\tau}\right)^{1/\alpha}, \tag{OA.3.56}$$

$$F^{-1}_{W\alpha}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right) = \theta_F - \theta_F\left(\frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)^{1/\alpha}, \tag{OA.3.57}$$

hence

$$\frac{N^{1/\alpha}}{\theta_F}\left(F^{-1}_{W,\alpha}\left(1 - \frac{1}{N\tau}\right) - F^{-1}_{W,\alpha}\left(1 - \frac{1}{N\tau}\frac{\log(T)}{\log(T)+1}\right)\right) \tag{OA.3.58}$$

$$\propto \frac{1}{\tau^{1/\alpha}} - \frac{1}{\tau^{1/\alpha}}\left(\frac{\log(T)}{\log(T)+1}\right)^{1/\alpha} \tag{OA.3.59}$$

$$\to 0. \tag{OA.3.60}$$

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

$$\frac{1}{a_N}\frac{1}{T^p} G^{-1}_{\beta,T}\left(1 - \frac{1}{N\tau}\left(1 - \frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\alpha+1/\beta}(\log(T))^{1/\beta}}{T^p} + \frac{\mu_T N^{1/\alpha}}{T^p}. \tag{OA.3.61}$$

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

$$\frac{N^{1/\alpha+1/\beta}(\log(T))^{1/\beta}}{T^{1/2}} \to 0. \tag{OA.3.62}$$

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

$$\frac{1}{a_N}\frac{1}{T^p} G^{-1}_{normal,T}\left(1 - \frac{1}{N\tau}\left(1 - \frac{\log(T)}{\log(T)+1}\right)\right) \sim \frac{N^{1/\alpha}\sqrt{\log(N)}}{T^p} + \frac{\mu_T N^{1/\alpha}}{T^p}. \tag{OA.3.63}$$

For this term to decay it is sufficient that

$$\frac{N^{1/\alpha}\sqrt{\log(N)}}{T^p} \to 0. \tag{OA.3.64}$$

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

Intermediate Order Statistics, Examples 1, 2, 3, Page OA-8

Example cdfs satisfy assumption 4 First we establish that cdfs F of examples 1-3 satisfy assumption 4, hence theorem 3.3 can be applied to the examples.

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

$$\frac{1-F_{Fr,\kappa}}{f_{Fr,\kappa}}(\theta) = \frac{(\theta+1)^{-\kappa}}{\kappa(\theta+1)^{-\kappa-1}} = \frac{1}{\kappa}(\theta+1), \tag{OA.3.65}$$

so

$$\left(\frac{1-F_{Fr,\kappa}}{f_{Fr,\kappa}}\right)' = \frac{1}{\kappa} = \gamma. \tag{OA.3.66}$$

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

$$\frac{1-F_{Gu,\lambda}}{f_{Gu,\lambda}} = \frac{e^{-\lambda x}}{\lambda e^{-\lambda x}} = \frac{1}{\lambda}, \tag{OA.3.67}$$

so

$$\left(\frac{1-F_{Gu,\lambda}}{f_{Gu,\lambda}}\right)' = 0 = \gamma. \tag{OA.3.68}$$

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

$$\frac{1-F_{W,\alpha}}{f_{W,\alpha}}(\theta) = \frac{\left(\frac{\theta_F-\theta}{\theta_F}\right)^{\alpha}}{\frac{\alpha}{\theta_F}\left(\frac{\theta_F-\theta}{\theta_F}\right)^{\alpha-1}} = \frac{1}{\alpha}(\theta_F-\theta), \tag{OA.3.69}$$

from which it follow that

$$\left(\frac{1-F_{W,\alpha}}{f_{W,\alpha}}\right)' = -\frac{1}{\alpha} = \gamma \tag{OA.3.70}$$

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.3.18):

$$\tilde{S}_{N,T}(u) = \frac{\sqrt{k}}{c_N}\left(F^{-1}(1-U_{k,N}+\tilde{u}_{S,N,T}) - F^{-1}\left(1-U_{k,N}\right) + \frac{1}{T^p}G_T^{-1}\left(1-\tilde{u}_{S,N,T}\right)\right), \tag{OA.3.71}$$

$$\tilde{s}_{N,T}(u) = \frac{\sqrt{k}}{c_N}\left(F^{-1}(1-U_{k,N}-\tilde{u}_{s,N,T}) - F^{-1}\left(1-U_{k,N}\right) + \frac{1}{T^p}G_T^{-1}\left(\tilde{u}_{s,N,T}\right)\right). \tag{OA.3.72}$$

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 to conclude that with probability approaching 1 $\liminf_{N,T\to\infty}\sup_{u\in[0,\epsilon]}\tilde{s}_{N,T}(u)\geq 0$. Proceeding as above, we conclude that conditions (2) and (3) of theorem 3.3 hold.

First, we establish the following elementary lemma.

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

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

$$\left(1+\frac{1}{N^\rho}\right)^\gamma-1=f\left(\frac{1}{N}\right)-1=\frac{1}{N}f'\left(\frac{\varkappa}{N}\right),\varkappa\in[0,1]. \tag{OA.3.73}$$

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

$$\left|\frac{1}{N}f'\left(\frac{\varkappa}{N}\right)\right|=O\left(\frac{1}{N^\rho}\left|\left(1+\frac{\varkappa^\rho}{N^\rho}\right)^{\gamma-1}\right|\right)=O\left(N^{-\rho}\right). \tag{OA.3.74}$$

□

Example 1, page OA-8 Let $F=F_{Fr,\kappa}$. Compute the normalizing functions of theorem 3.3.

$$F^{-1}_{Fr,\kappa}\left(1-\frac{k}{N}\right)=U_{F_{Fr,\kappa}}\left(\frac{N}{k}\right)=\left(\frac{N}{k}\right)^{1/\kappa}-1, \tag{OA.3.75}$$

$$c_N=\frac{N}{k}\left(\left(\frac{1}{1-F_{Fr,\kappa}}\right)^{-1}\right)'\left(\frac{N}{k}\right)=\frac{N}{k}U'_{F_{Fr,\kappa}}\left(\frac{N}{k}\right)=\frac{1}{\kappa}\left(\frac{N}{k}\right)^{1/\kappa}. \tag{OA.3.76}$$

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

$$\tilde{u}_{S,N,T}=U_{k,N}\frac{1}{N^\rho+1}\in\left[0,U_{k,N}\right]. \tag{OA.3.77}$$

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.3.4.

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

$$\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\left[F^{-1}_{Fr,\kappa}\left(1-U_{k,N}\frac{N^\rho}{N^\rho+1}\right)-F^{-1}_{Fr,\kappa}\left(1-U_{k,N}\right)\right] \tag{OA.3.78}$$

$$=\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\left[\left(\frac{1}{U_{k,N}}\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa}-\left(\frac{1}{U_{k,N}}\right)^{1/\kappa}\right] \tag{OA.3.79}$$

$$=\left(\frac{N}{k}U_{k,N}\right)^{-1/\kappa}\sqrt{k}\left[\left(\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa}-1\right]. \tag{OA.3.80}$$

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

$$\sqrt{k}\left[\left(\frac{N^\rho+1}{N^\rho}\right)^{1/\kappa}-1\right] \to 0. \tag{OA.3.81}$$

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

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

$$\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-U_{k,N}\frac{1}{N^\rho+1}\right) \tag{OA.3.82}$$

$$\sim \frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{N^{\rho/\beta}}{U^{1/\beta}_{k,N}T^p} \tag{OA.3.83}$$

$$= \left(\frac{N}{k}U_{k,N}\right)^{-1/\beta}\frac{k^{1/2+1/\kappa-1/\beta}N^{\rho/\beta}}{N^{1/\kappa-1/\beta}T^p}. \tag{OA.3.84}$$

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

$$\frac{k^{1/2+1/\kappa-1/\beta}}{N^{1/\kappa-1/\beta-\rho/\beta}T^p} \to 0. \tag{OA.3.85}$$

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

$$\frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{normal,T}\left(1-U_{k,N}\frac{1}{N^\rho+1}\right) \tag{OA.3.86}$$

$$= \frac{k^{1/2+1/\kappa}}{N^{1/\kappa}}\frac{1}{T^p}G^{-1}_{normal,T}\left(1-\left(\frac{N}{k}U_{k,N}\right)\frac{N^\delta}{N}\frac{1}{N^\rho+1}\right) \tag{OA.3.87}$$

$$\sim \frac{k^{1/2+1/\kappa}\sqrt{\log(N)}}{N^{1/\kappa}}\frac{1}{T^p}. \tag{OA.3.88}$$

Thus, the scaled G term decays if

$$\frac{k^{1/2+1/\kappa}\sqrt{\log(N)}}{N^{1/\kappa}}\frac{1}{T^p} \to 0. \tag{OA.3.89}$$

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.

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

Rate conditions if $k = N^\delta$ Conditions (OA.3.81), (OA.3.85), and (OA.3.89) 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.3.81) transforms to the requirement that $N^{\delta/2-\rho} \to 0$, which holds if $\rho > \delta/2$. Condition (OA.3.85) becomes

$$\frac{N^{\delta/2+\delta/\kappa-\delta/\beta}}{N^{1/\kappa-1/\beta-\rho/\beta}T^p} = \frac{N^{\delta/2+(1-\delta)(1/\beta-1/\kappa)+\rho/\beta}}{T^p} \to 0. \tag{OA.3.90}$$

Similarly, condition (OA.3.89) becomes

$$\frac{N^{\delta/2+(1-\delta)(-1/\kappa)}\sqrt{\log(N)}}{T^p} \to 0 \tag{OA.3.91}$$

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

$$\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta-1/\kappa)+\nu/\beta}}{T^p} \to 0. \tag{OA.3.92}$$

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

Example 2, page OA-9 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

$$F^{-1}_{Gu,\lambda}\left(1-\frac{k}{N}\right) = U_{Gu,\lambda}\left(\frac{N}{k}\right) = \frac{\log(N/k)}{\lambda}, \tag{OA.3.93}$$

$$c_N = \frac{N}{k} \times \left(\left(\frac{1}{1-F_{Gu,\lambda}}\right)^{-1}\right)'\left(\frac{N}{k}\right) = \frac{N}{k}U'_{Gu,\lambda}\left(\frac{N}{k}\right) = \frac{1}{\lambda}\frac{N}{k}\frac{k}{N} = \frac{1}{\lambda}. \tag{OA.3.94}$$

Pick $\tilde{u}_{S,N,T}$ as in eq. (OA.3.77):

$$u = U_{k,N}\frac{1}{N^\rho+1}. \tag{OA.3.95}$$

Then

$$\sqrt{k}\left(F^{-1}_{Gu,\lambda}\left(1-U_{k,N}\frac{N^\rho}{N^\rho+1}\right) - F^{-1}_{Gu,\lambda}\left(1-U_{k,N}\right)\right) \tag{OA.3.96}$$

$$=\sqrt{k}\left(\frac{\log U_{k,N} + \log\frac{N^\rho+1}{N^\rho}}{\lambda} - \frac{\log U_{k,N}}{\lambda}\right) \tag{OA.3.97}$$

$$\sim\sqrt{k}\log\left(1+\frac{1}{N^\rho}\right) \sim \sqrt{k}N^{-\rho}. \tag{OA.3.98}$$

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

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

$$\sqrt{k}\frac{1}{T^p}G^{-1}_{\beta,T}\left(1-U_{k,N}\frac{1}{N^\rho+1}\right) \tag{OA.3.99}$$

$$\sim k^{1/2} \frac{(N)^{\rho/\beta}}{U_{k,N}^{1/\beta} T^p} \tag{OA.3.100}$$

$$= \left(\frac{N}{k} U_{k,N}\right)^{-1/\beta} \frac{k^{1/2-1/\beta}}{N^{-1/\beta-\rho/\beta} T^p}. \tag{OA.3.101}$$

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

$$\frac{k^{1/2-1/\beta}}{N^{-1/\beta-\rho/\beta} T^p} \to 0. \tag{OA.3.102}$$

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

$$\frac{N^{\delta/2-\delta/\beta}}{N^{-1/\beta-\rho/\beta} T^p} = \frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\beta)+\nu/\beta}}{T^p} \to 0, \tag{OA.3.103}$$

where ν is any fixed number > 0.

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

$$\sqrt{k} \frac{1}{T^p} G^{-1}_{normal,T} \left(1 - U_{k,N} \frac{1}{N^{\rho}+1}\right) \sim k^{1/2} \frac{\sqrt{\log(N)}}{T^p}. \tag{OA.3.104}$$

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

$$\frac{N^{\delta/2}\sqrt{\log(N)}}{T^p} \to 0. \tag{OA.3.105}$$

Example 3, page OA-9 Finally, let $F = F_{W,\alpha}$. We proceed as in the previous two examples.

First compute the normalizing functions of theorem 3.3.

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

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

Pick $\tilde{u}_{S,N,T}$ as in eq. (OA.3.77):

$$\tilde{u}_{S,N,T} = U_{k,N} \frac{1}{N^{\rho}+1} \tag{OA.3.106}$$

Then

$$\frac{N^{1/\alpha}}{\frac{\theta_F}{\alpha} k^{1/\alpha-1/2}} \left(F^{-1}_{W,\alpha}\left(1 - U_{k,N}\frac{N^{\rho}}{N^{\rho}+1}\right) - F^{-1}_{W,\alpha}\left(1 - U_{k,N}\right)\right) \tag{OA.3.107}$$

$$\sim \frac{N^{1/\alpha}}{k^{1/\alpha-1/2}} \left[(U_{k,N})^{1/\alpha} - \left(\frac{N^{\rho}}{N^{\rho}+1}\right)^{1/\alpha} (U_{k,N})^{1/\alpha}\right] \tag{OA.3.108}$$

$$=\sqrt{k}\left(\frac{N}{k}U_{k,N}\right)^{-1/\alpha}\left[1-\left(\frac{N^{\rho}}{N^{\rho}+1}\right)^{1/\alpha}\right]. \tag{OA.3.109}$$

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

$$\sqrt{k}\left[1-\left(\frac{N^{\rho}}{N^{\rho}+1}\right)^{1/\alpha}\right]\sim\sqrt{k}N^{-\rho}\to 0. \tag{OA.3.110}$$

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

$$\frac{1}{T^p}G_{\beta,T}^{-1}\left(1U_{k,N}\frac{1}{N^{\rho}+1}\right)\sim\frac{(N)^{\rho/\beta}}{U_{k,N}^{1/\beta}T^p}. \tag{OA.3.111}$$

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

$$\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}}\frac{N^{\rho/\beta}}{U_{k,N}^{1/\beta}T^p}=\frac{N^{1/\alpha+1/\beta+\rho/\beta}}{k^{1/\alpha+1/\beta-1/2}T^p}\left(\frac{N}{k}U_{k,N}\right)^{-1/\beta}\sim\frac{N^{1/\alpha+1/\beta+\rho/\beta}}{k^{1/\alpha+1/\beta-1/2}T^p}\to 0\ . \tag{OA.3.112}$$

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

$$\frac{N^{\delta/2(1+1/\beta)+(1-\delta)(1/\alpha+1/\beta)+\nu/\beta}}{T^p}\to 0. \tag{OA.3.113}$$

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

$$\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}}G_{normal,T}^{-1}\left(1-U_{k,N}\frac{1}{N^{\rho}+1}\right)\sim\frac{N^{1/\alpha}}{k^{1/\alpha-1/2}}\frac{\sqrt{\log(N)}}{T^p}. \tag{OA.3.114}$$

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

$$\frac{N^{\delta/2+(1-\delta)(1/\alpha)}\sqrt{\log(N)}}{T^p}\to 0. \tag{OA.3.115}$$

OA.4 Additional Simulation Results

In this section, we extend the analysis of the Monte Carlo study of section 5 in the the main text. First, section OA.4.1 reports the results of an expanded study in that covers additional distributions for (θ_i, u_{it}) and additional sample sizes. The results match those of section 5 and our recommendations for inference are unchanged. Second, in section OA.4.2 we provide simulation support for the rule of remark 8 for choosing the tuning parameters of extreme-order CIs. Third, in section OA.4.3 we consider the performance of the quantile estimators proposed in section 4. Finally, section OA.4.4 shows that convergence in the intermediate-order theorem 4.5 may be slow, driving relatively poor performance of the corresponding CIs.

The design of study is as in section 5, with three additions. First, we add two additional samples sizes throughout: a larger cross-sectional size of $N = 10000$ and an intermediate individual sample size $T = 20$. Second, in section OA.4.1 we explore three additional distributions F for coefficients θ_i and an additional distribution G for u_{it}. Third, in section OA.4.2 we consider additional specifications of the extreme CIs that differ in their tuning parameter values.

OA.4.1 Confidence Intervals: Additional Sample Sizes and Distributions

We begin with an expanded study of properties of the CIs of section 5. We consider four specifications for F (the CDF of θ_i), and three for G, the CDF of u_{it}. For F, we consider t_3 (as in the main text), $F_{Fr,4}$, $F_{Gu,1}$, and $F_{W,4}$ (see section OA.3.2 for definitions). The three new F are prototypical distributions with $\gamma > 0$, $\gamma = 0$, $\gamma < 0$, respectively. For G, we consider $G = G_\beta, \beta = 8$; a noiseless scenario (as in the main text); and an additional specification of $G = N(0, \sigma^2)$ (see section OA.3.2). G_β and $G = N(0, \sigma^2)$ are prototypical distributions that reflect the assumptions of proposition 3.2.

The results for the new distributions and values of (N, T) broadly match the results for $F = t_3$ and sample sizes presented in the main text. We refer to section 5 for a full interpretation. The full results are presented on figs. OA.1-OA.24. Note that figs. OA.1-OA.2 and OA.5-OA.6 are expanded versions of figs. 1 and 2 in the main text, respectively.

Our overall recommendation for inference is identical regardless of the sign of (F, G):

(1) If $(1 - \tau)N \leq 100$, one should use the extreme-order CIs.

(2) If $(1 - \tau)N > 100$, a central-order CI should be used with the correction of Jochmans and Weidner (2024).

A particular point of interest concerns the intermediate-order CI of theorem 4.5. It is a viable option for inference with good length properties if N is sufficiently large and the rate conditions for it hold, as figs. OA.1-OA.24 and section OA.4.4 below suggest. Its properties generally improve as N increases (e.g. figs. OA.1 and OA.5, shifting the target quantile so that a fixed number of order statistics fall to the right of it). These improvement reflect better approximation quality of the limit $N(0, 1)$ law for larger N (see also section OA.4.4 below). The impact of the rate restrictions can be appreciated by contrasting figs. OA.1 and OA.5 for $N = 10000$.

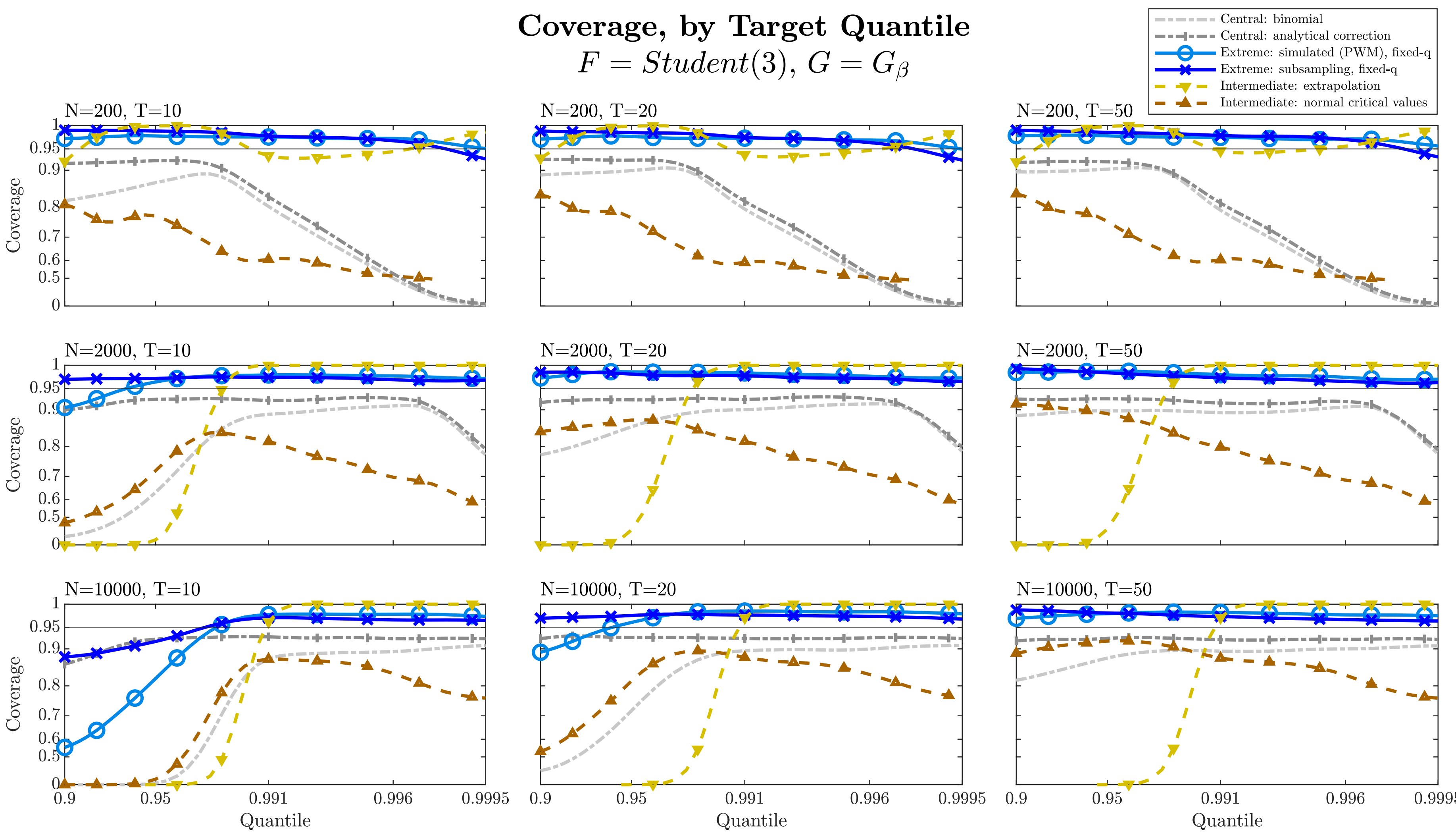

Figure OA.1: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

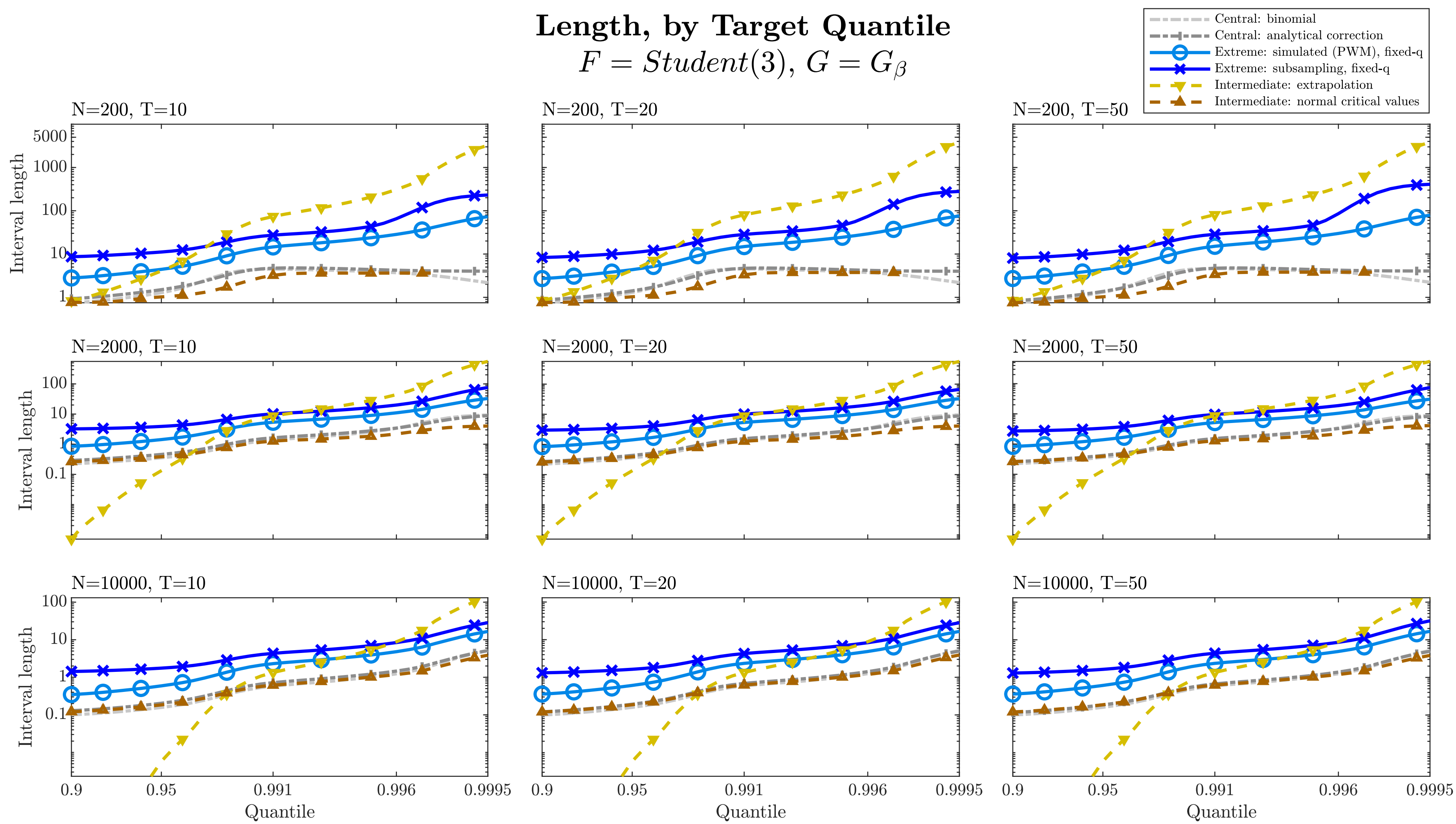

Figure OA.2: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

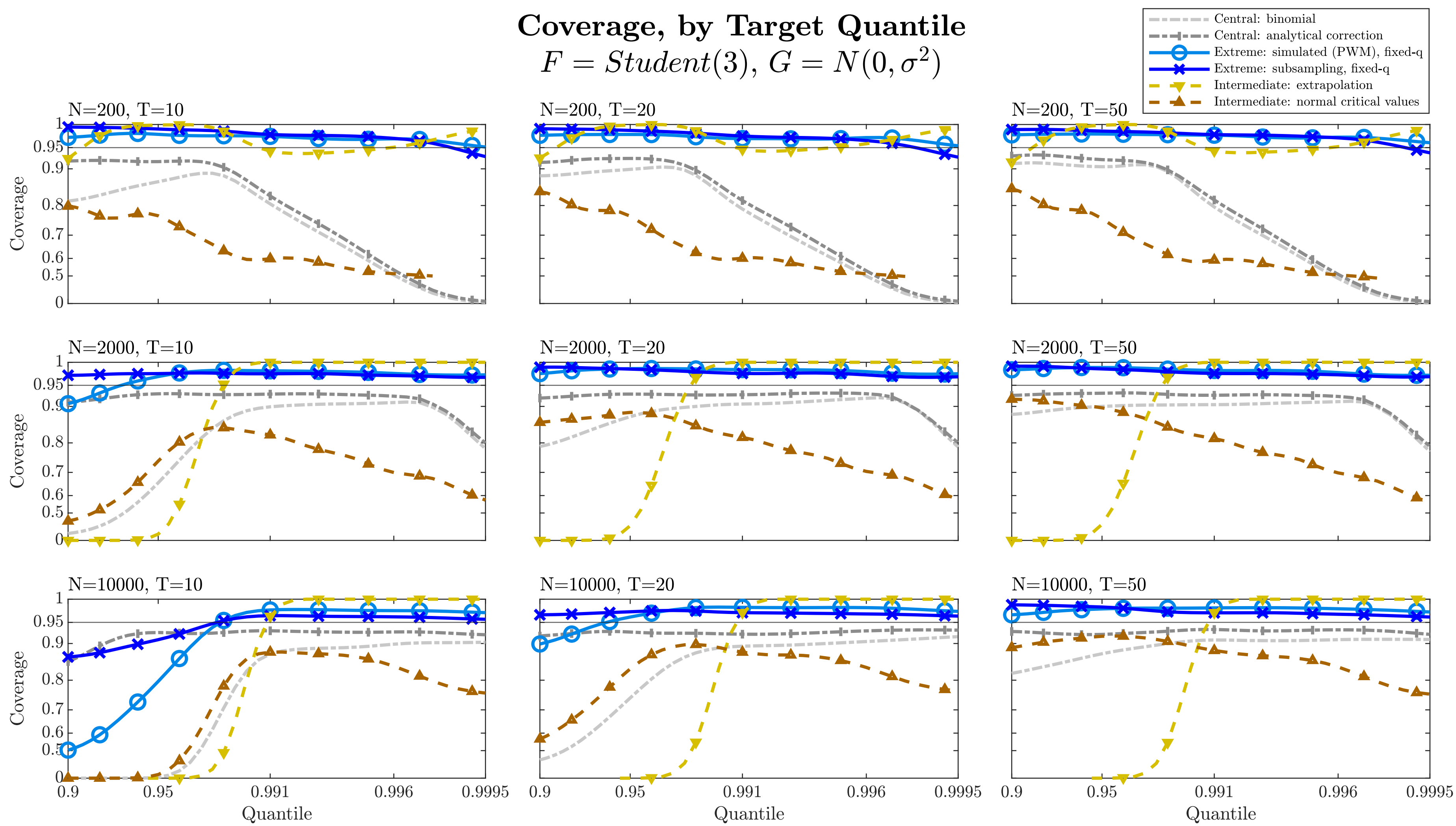

Figure OA.3: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

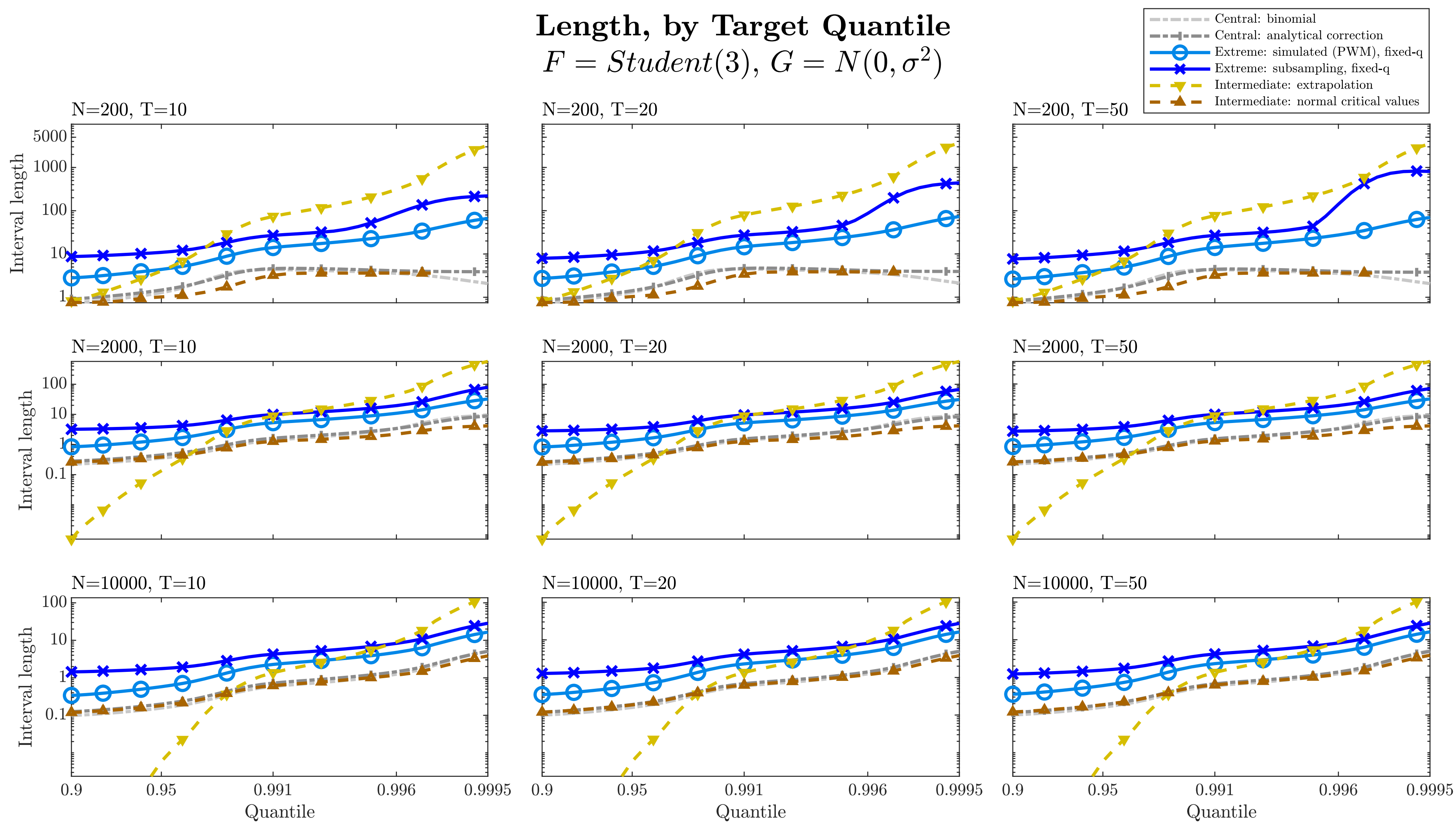

Figure OA.4: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

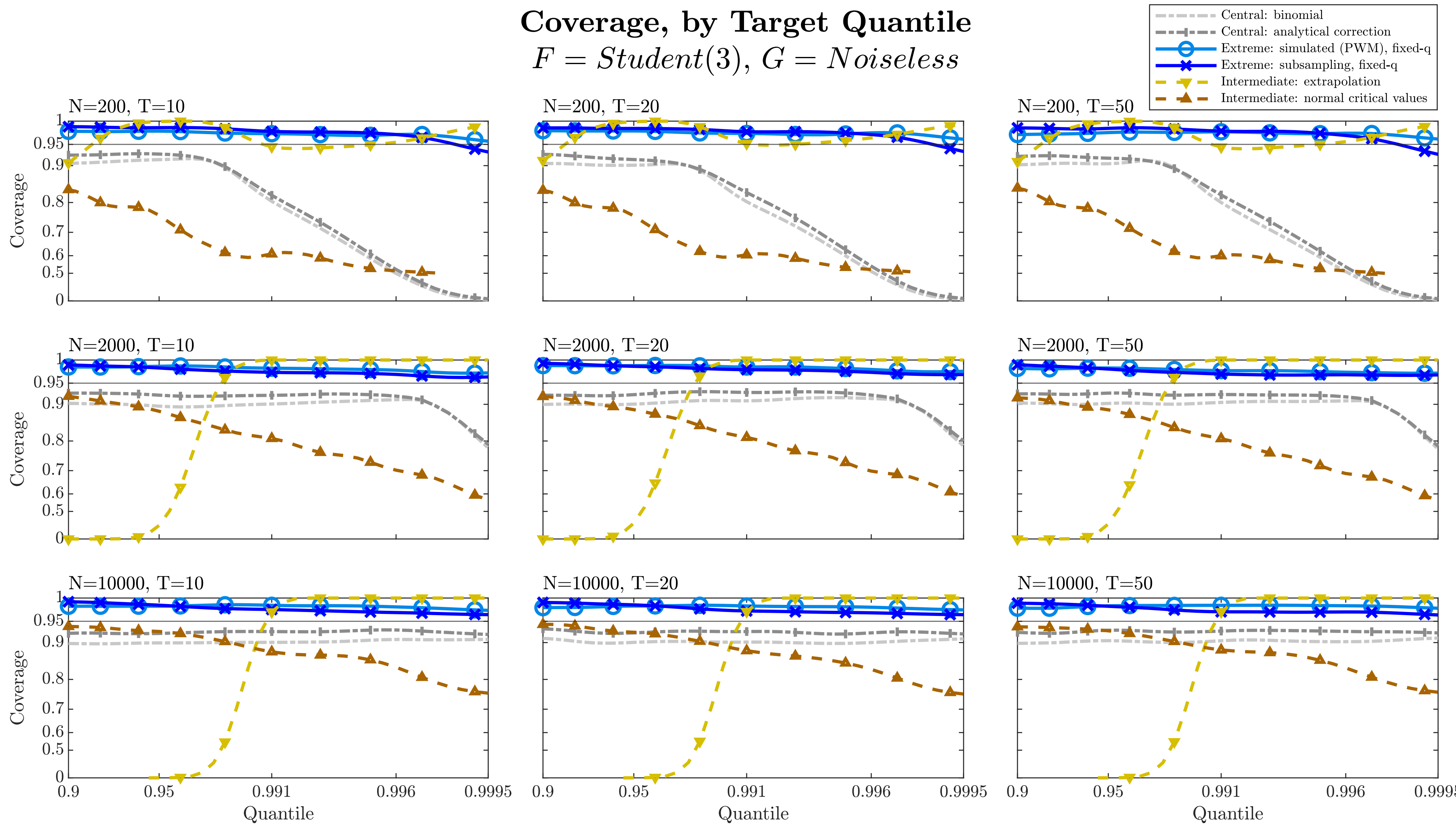

Figure OA.5: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

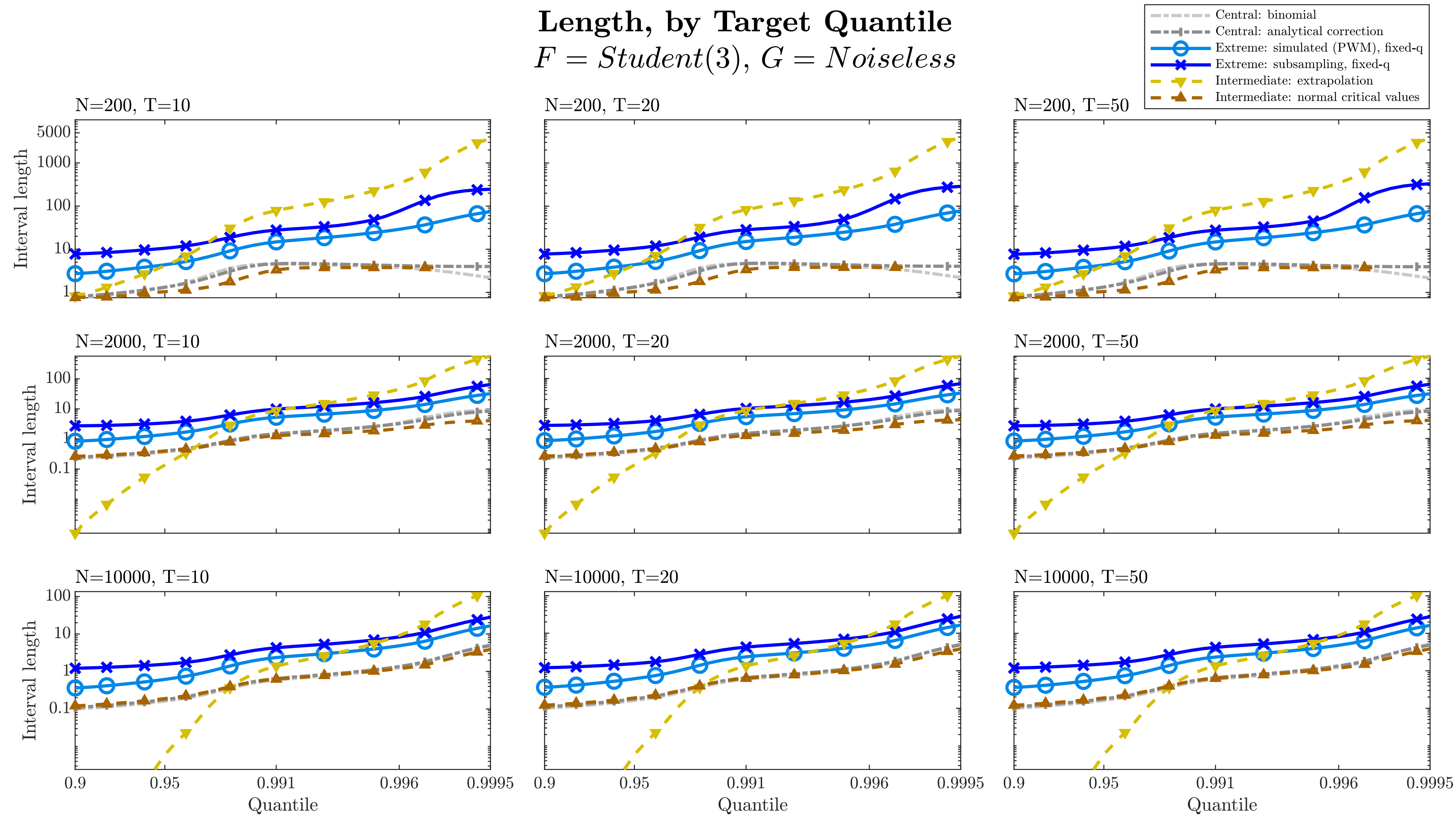

Figure OA.6: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

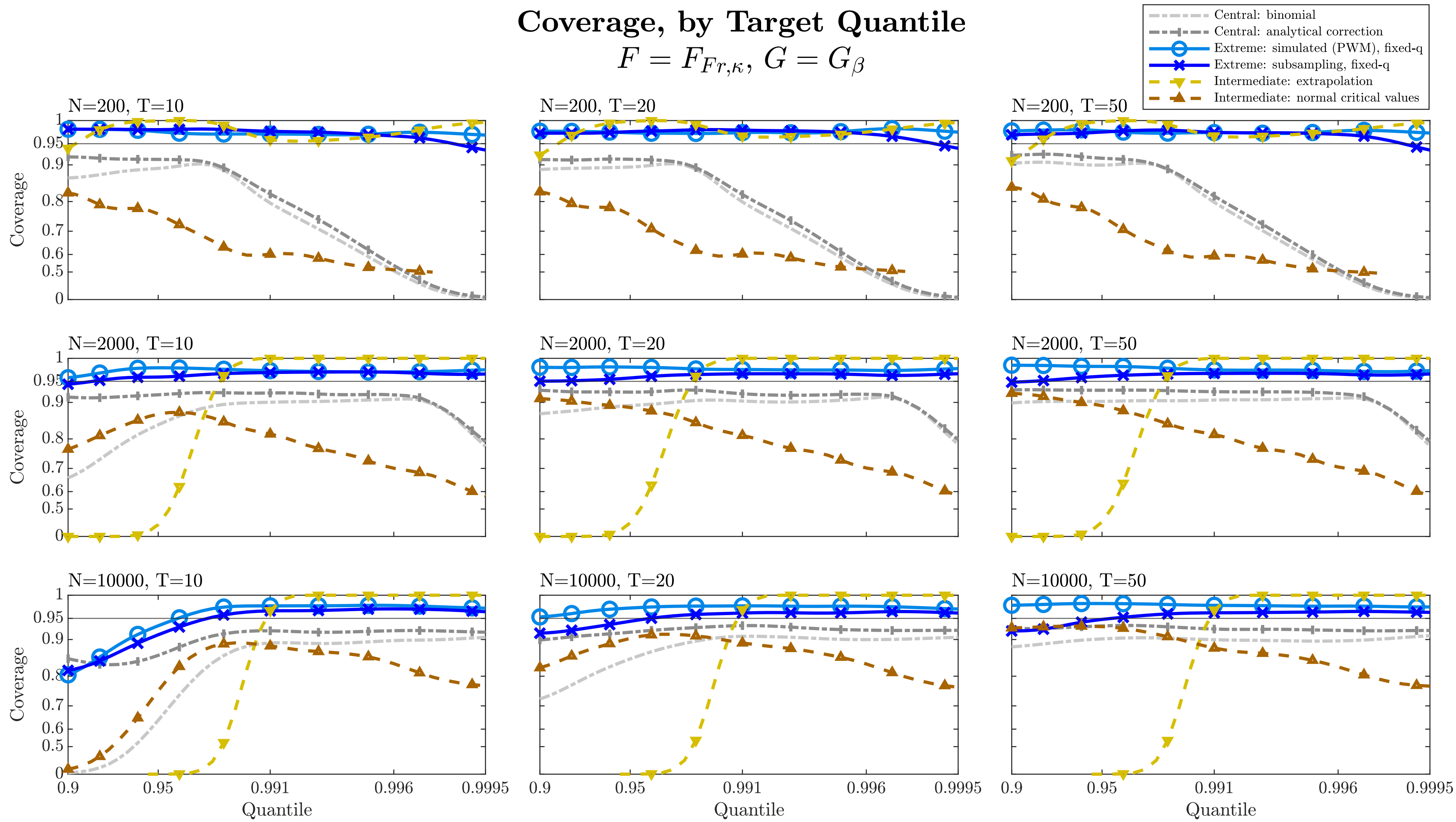

Figure OA.7: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,\kappa}$, $u_{it} \sim G_\beta, \beta = 8$, Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

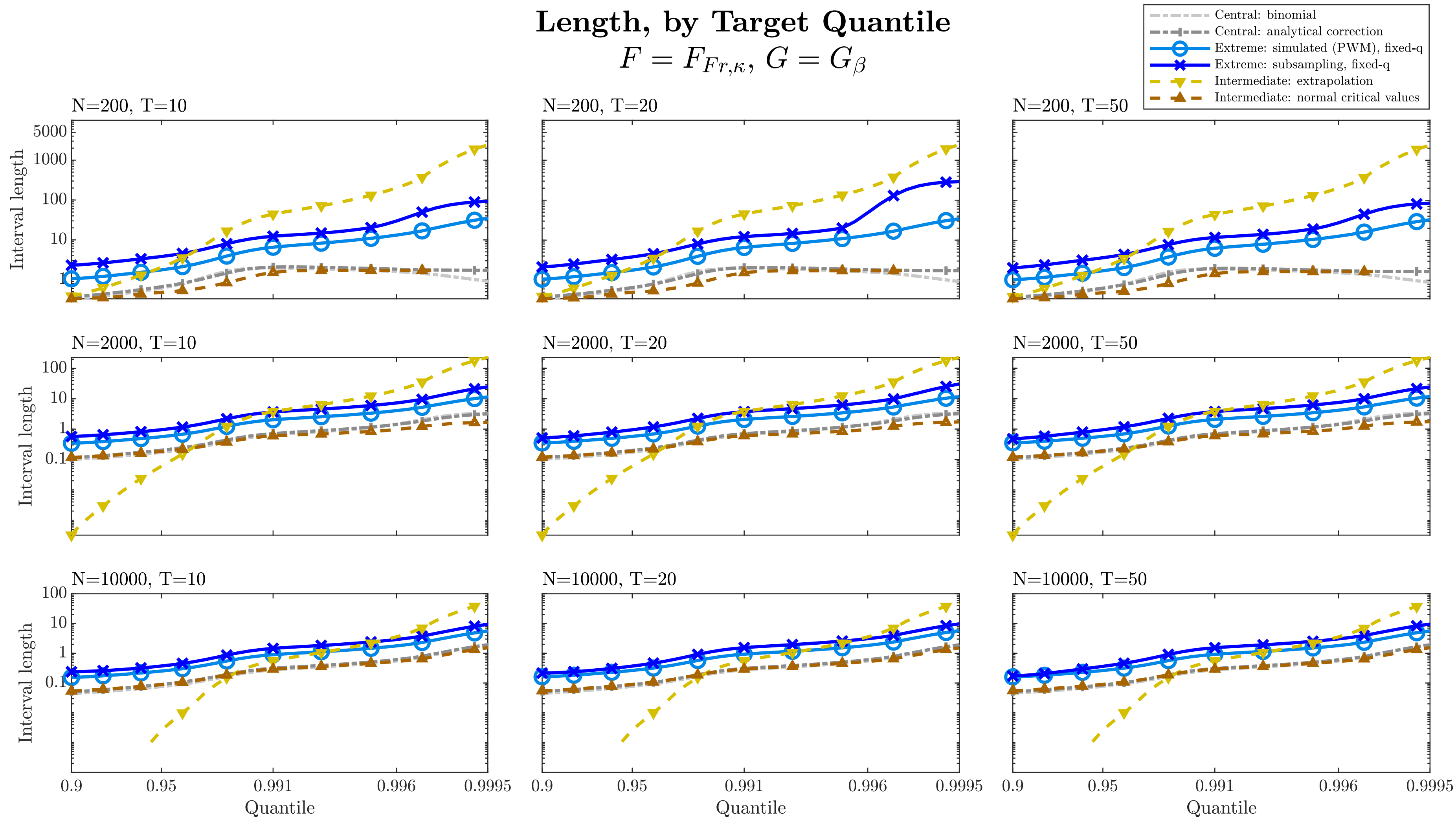

Figure OA.8: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,\kappa}$, $u_{it} \sim G_\beta, \beta = 8$. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

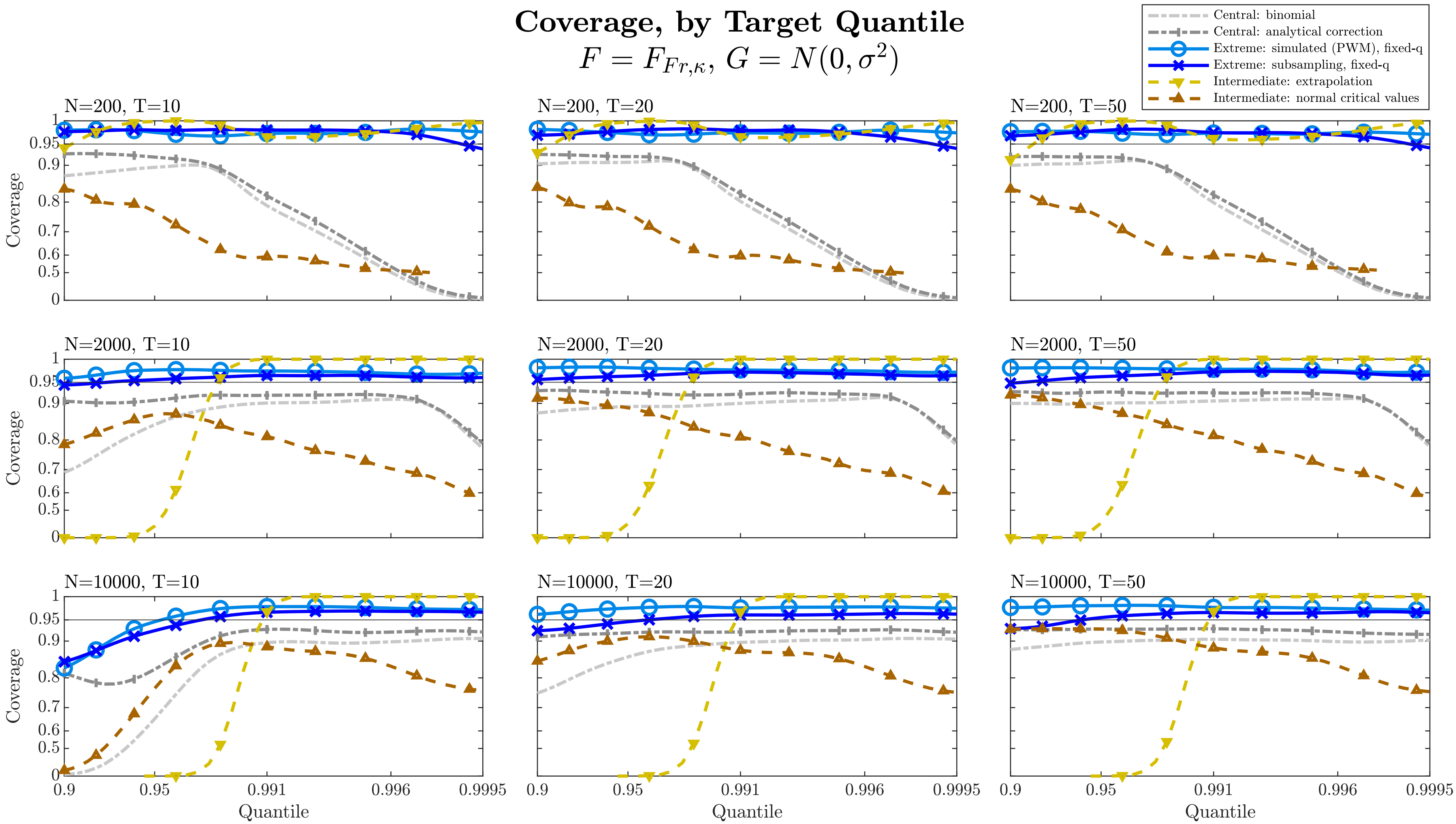

Figure OA.9: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

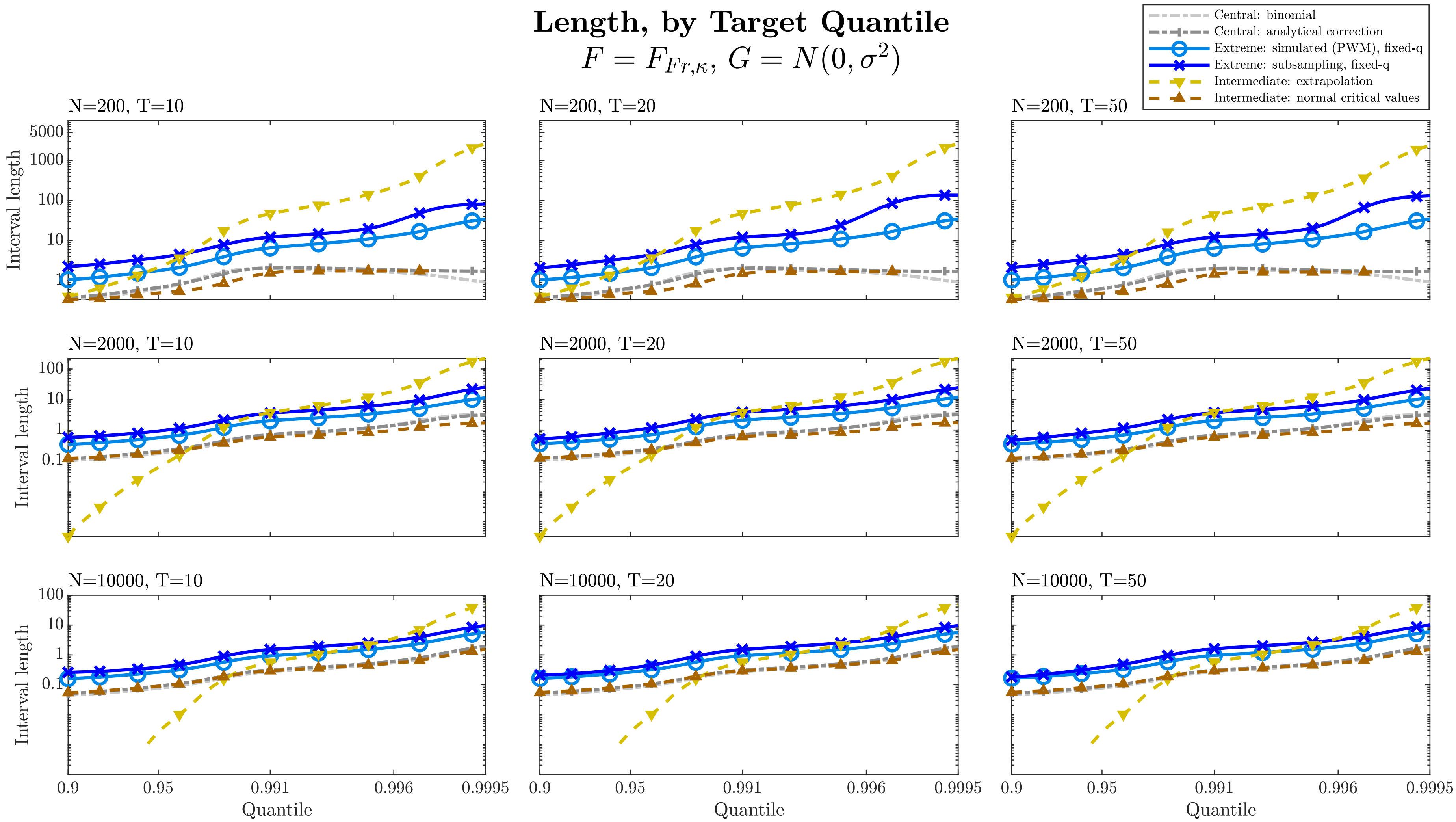

Figure OA.10: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

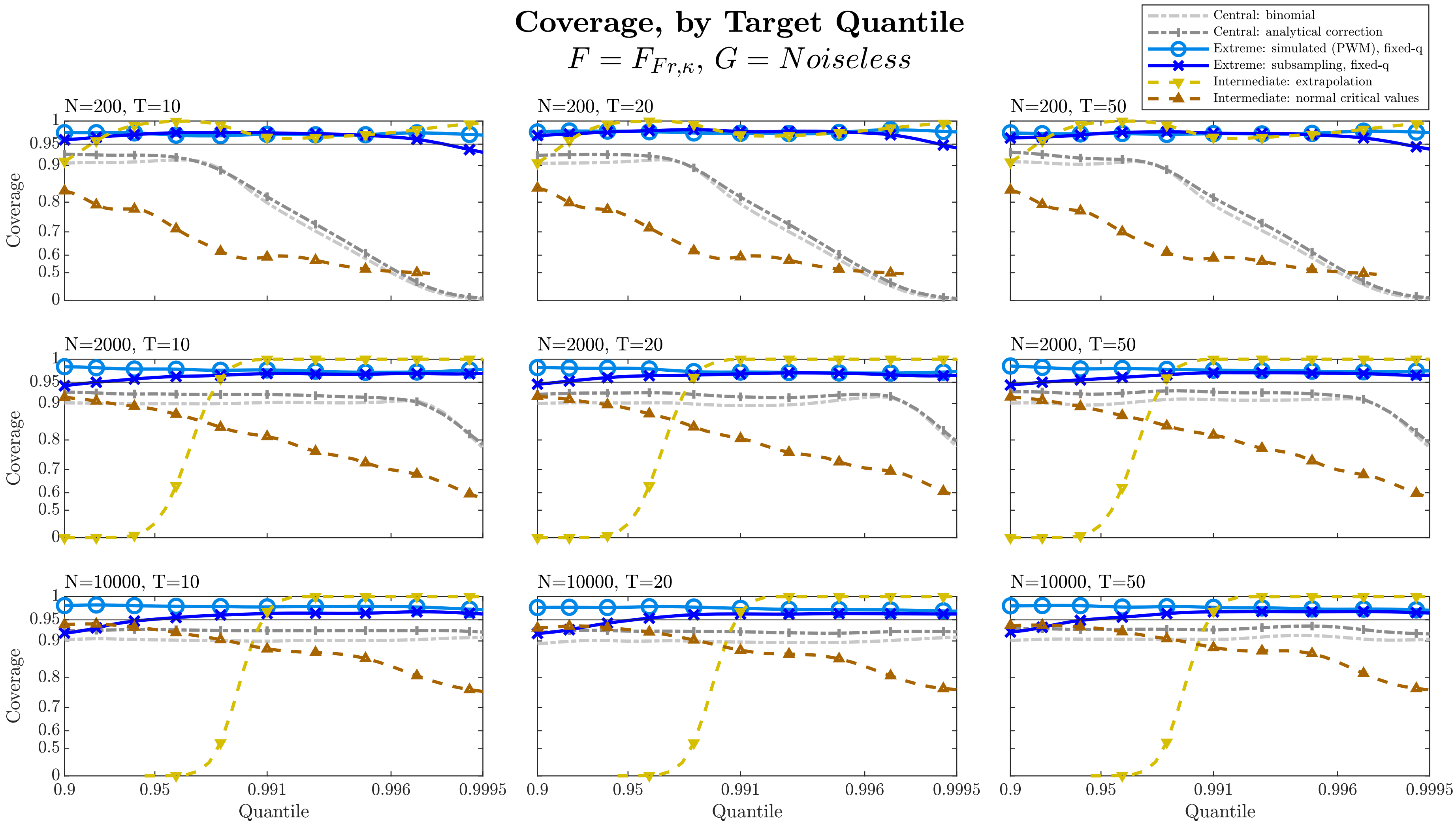

Figure OA.11: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

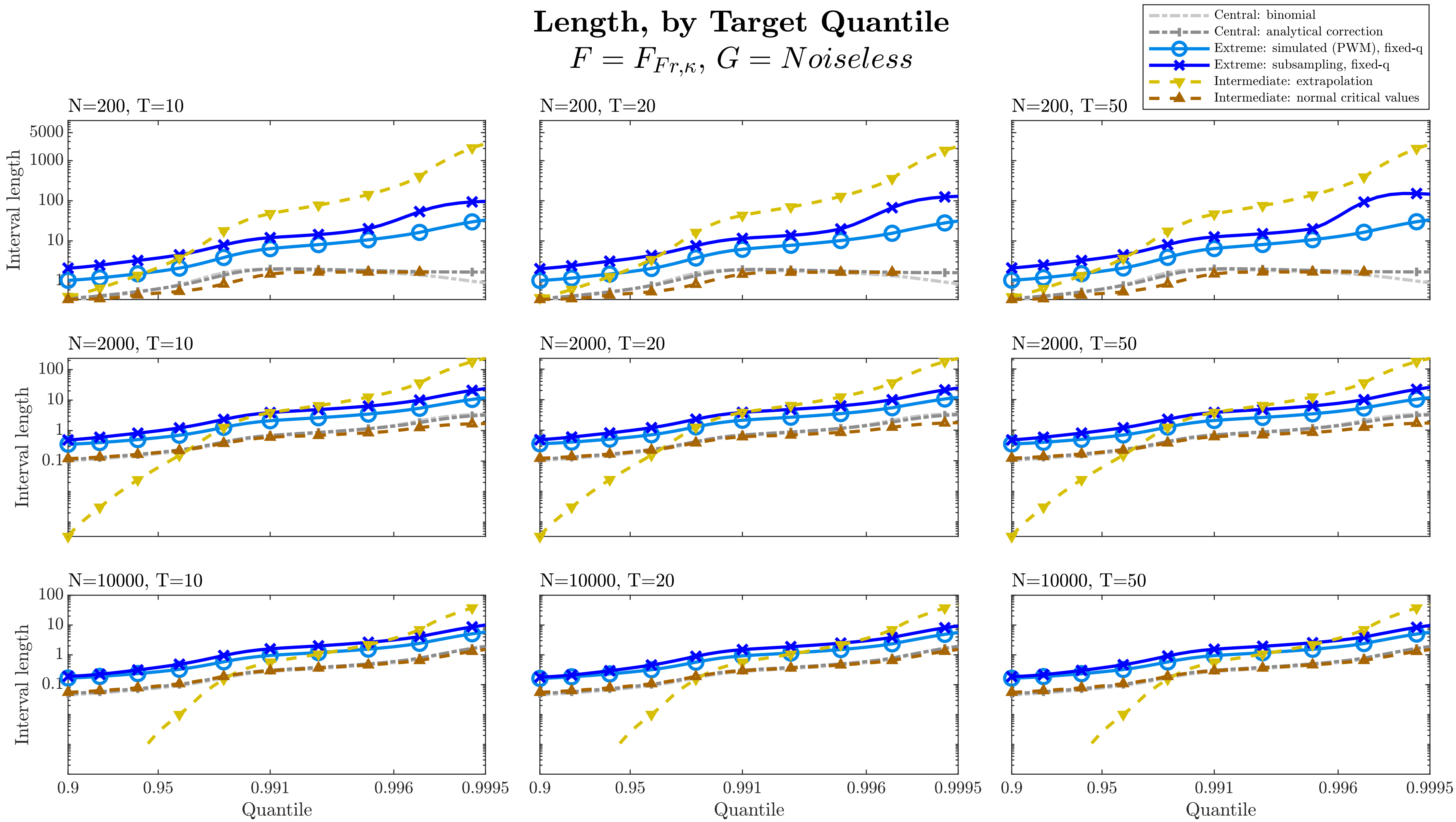

Figure OA.12: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

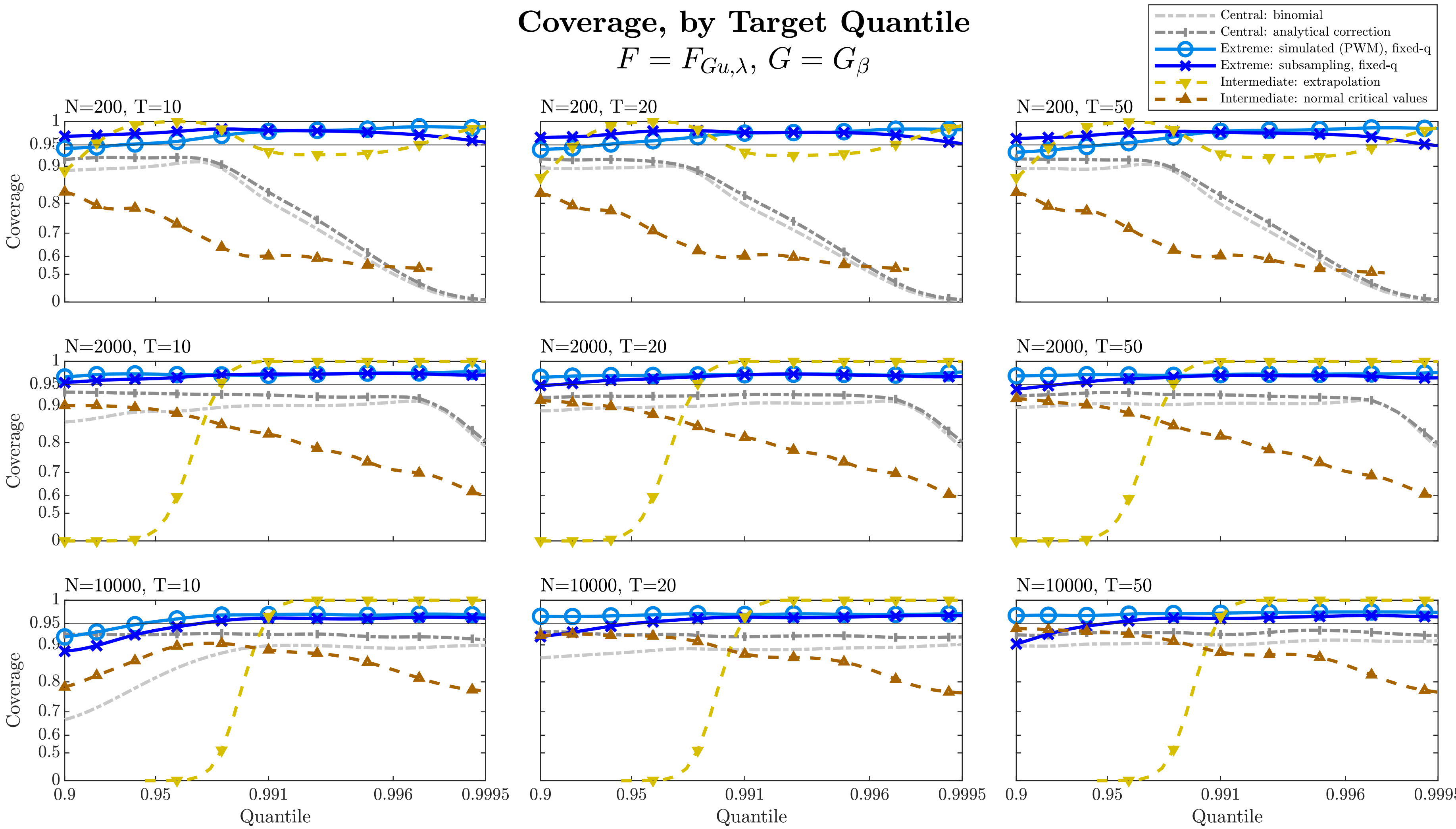

Figure OA.13: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim G_\beta, \beta = 8$, Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

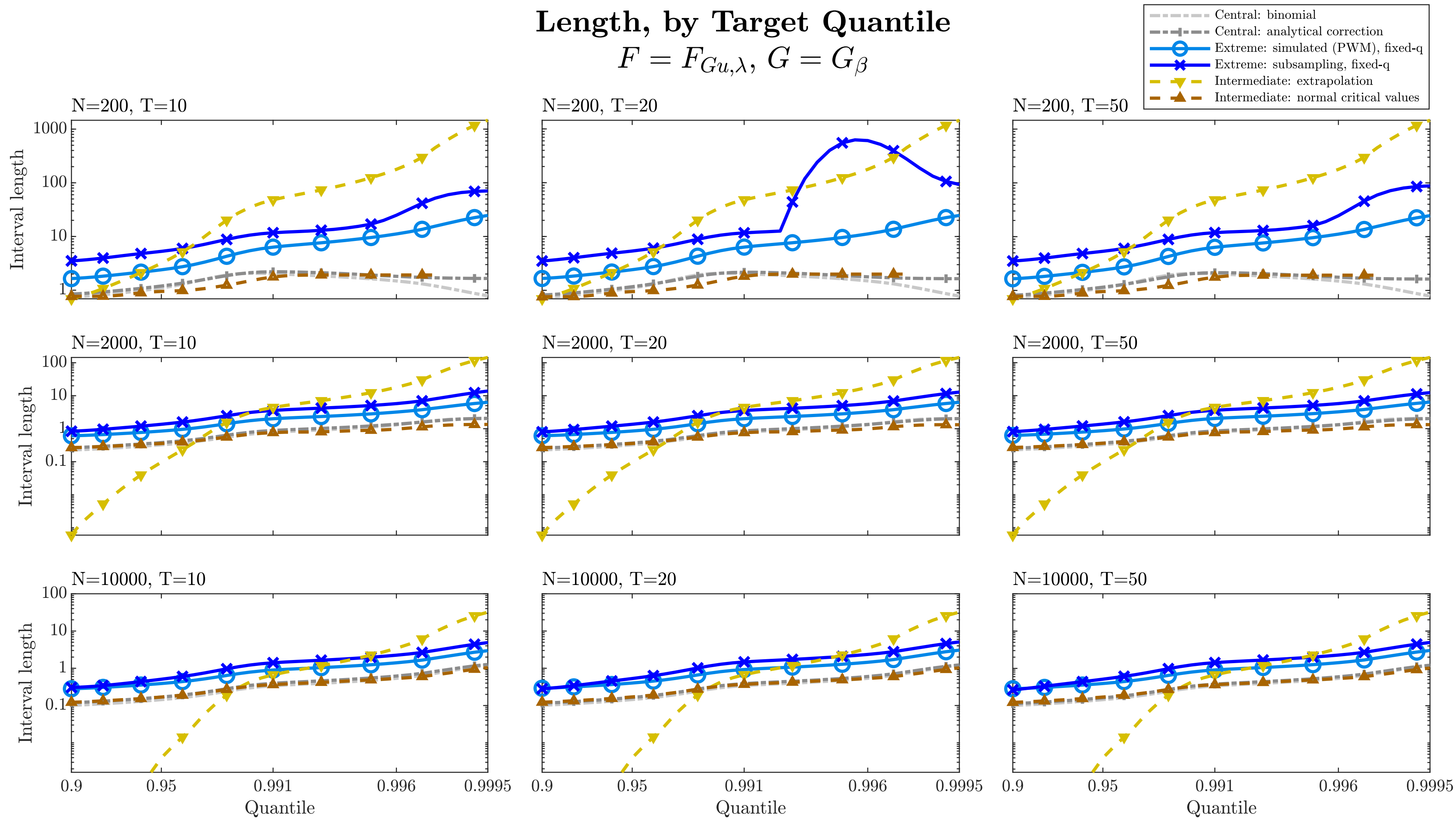

Figure OA.14: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim G_\beta, \beta = 8$. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

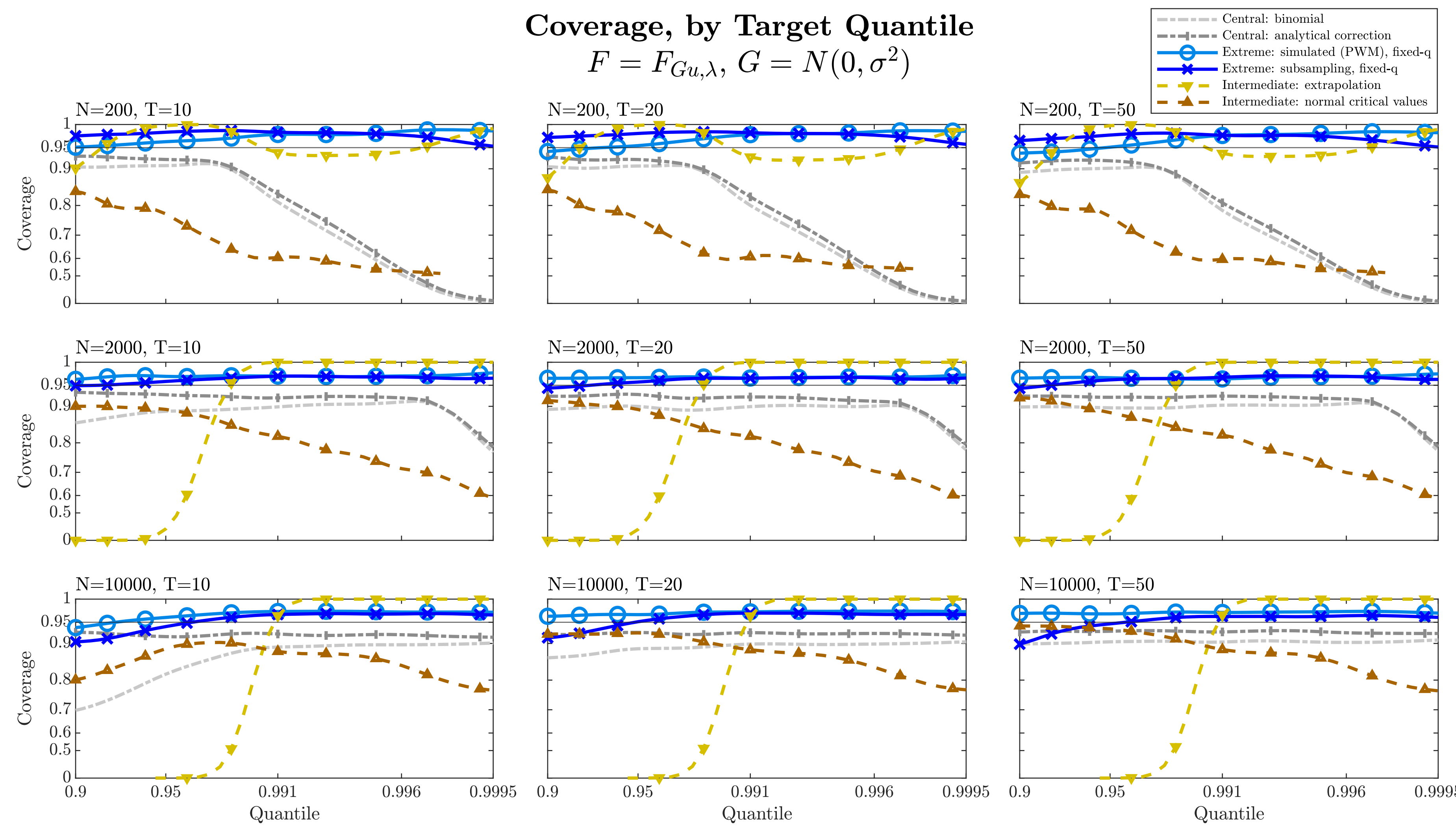

Figure OA.15: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

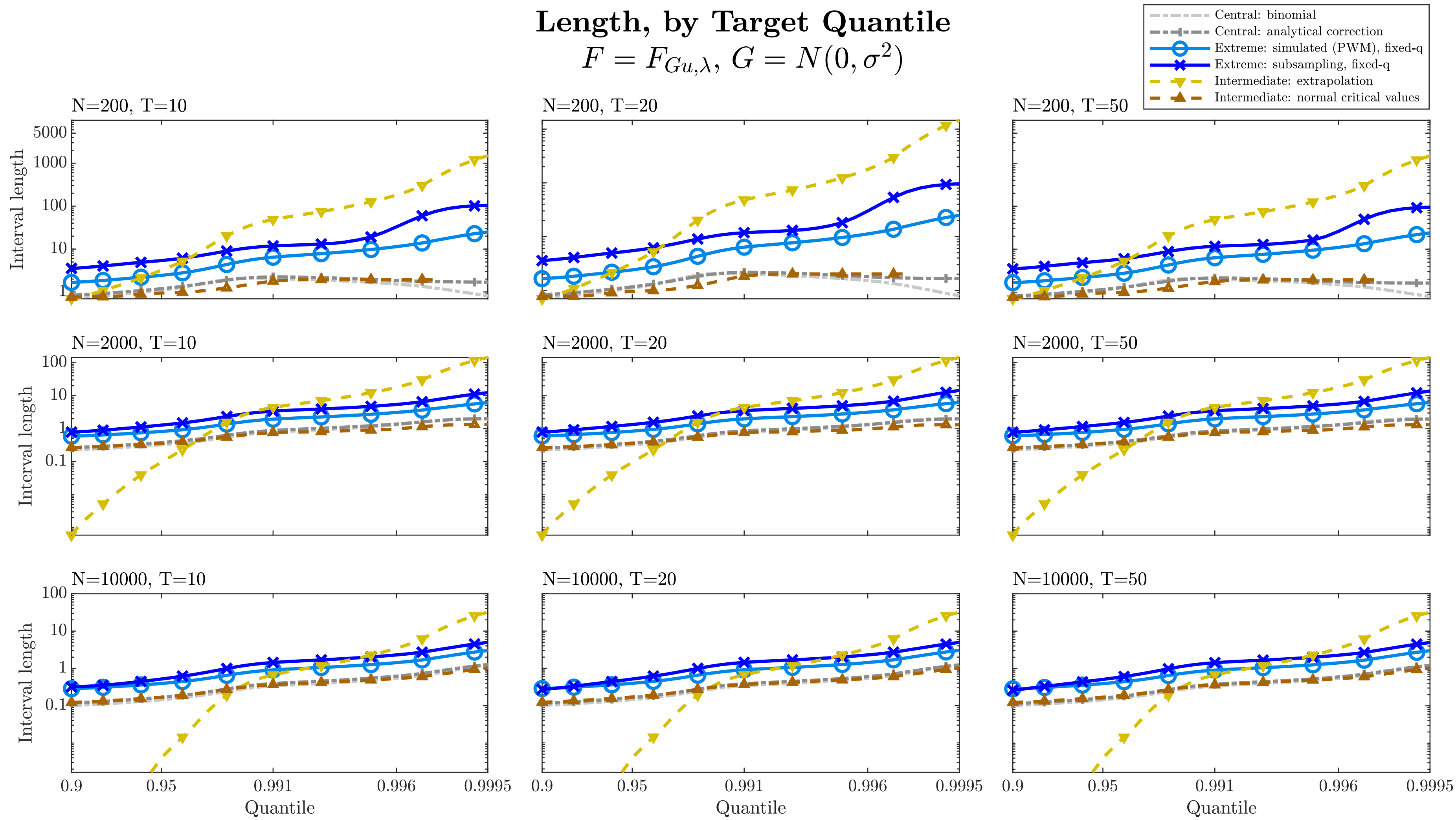

Figure OA.16: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

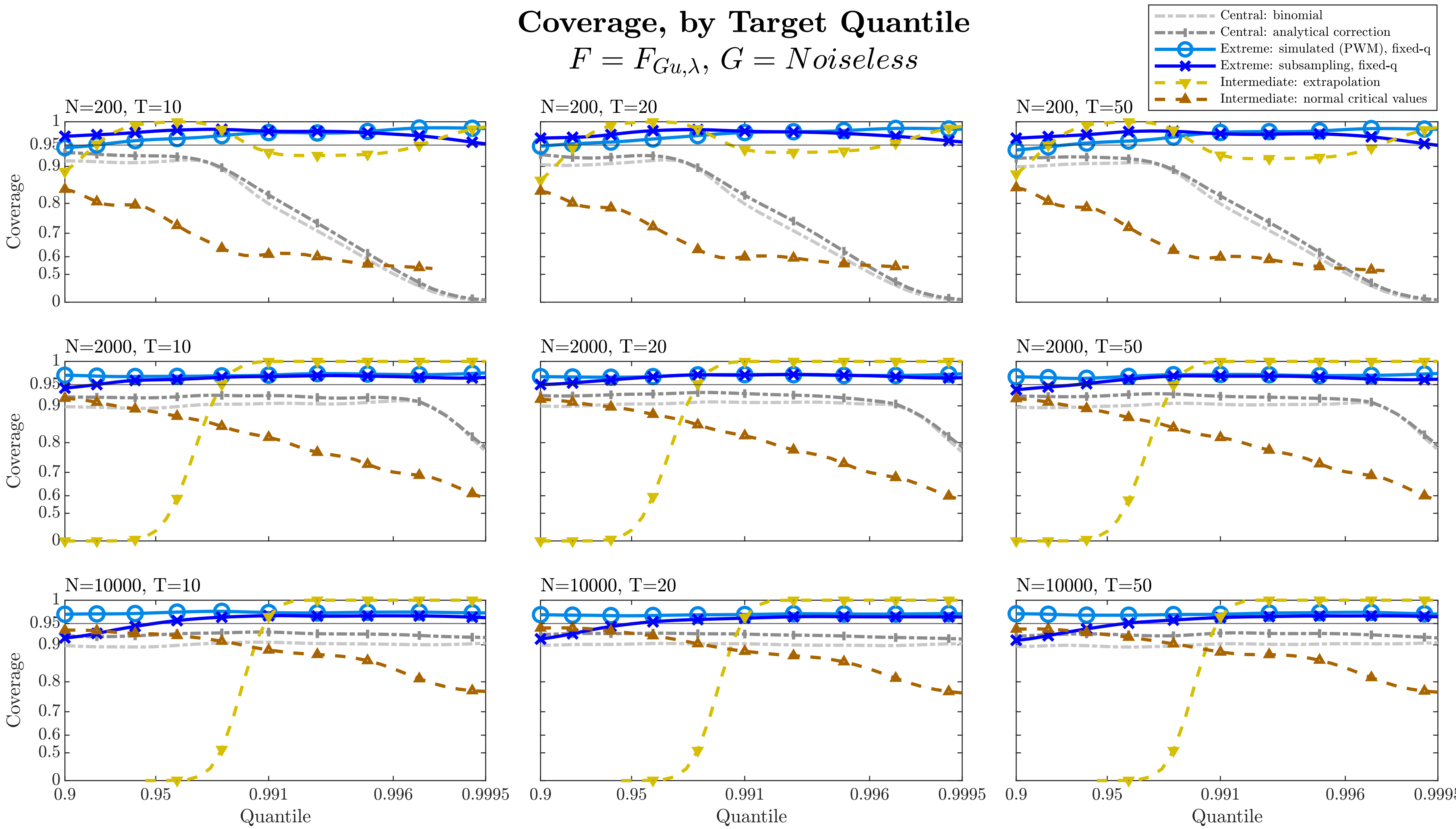

Figure OA.17: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

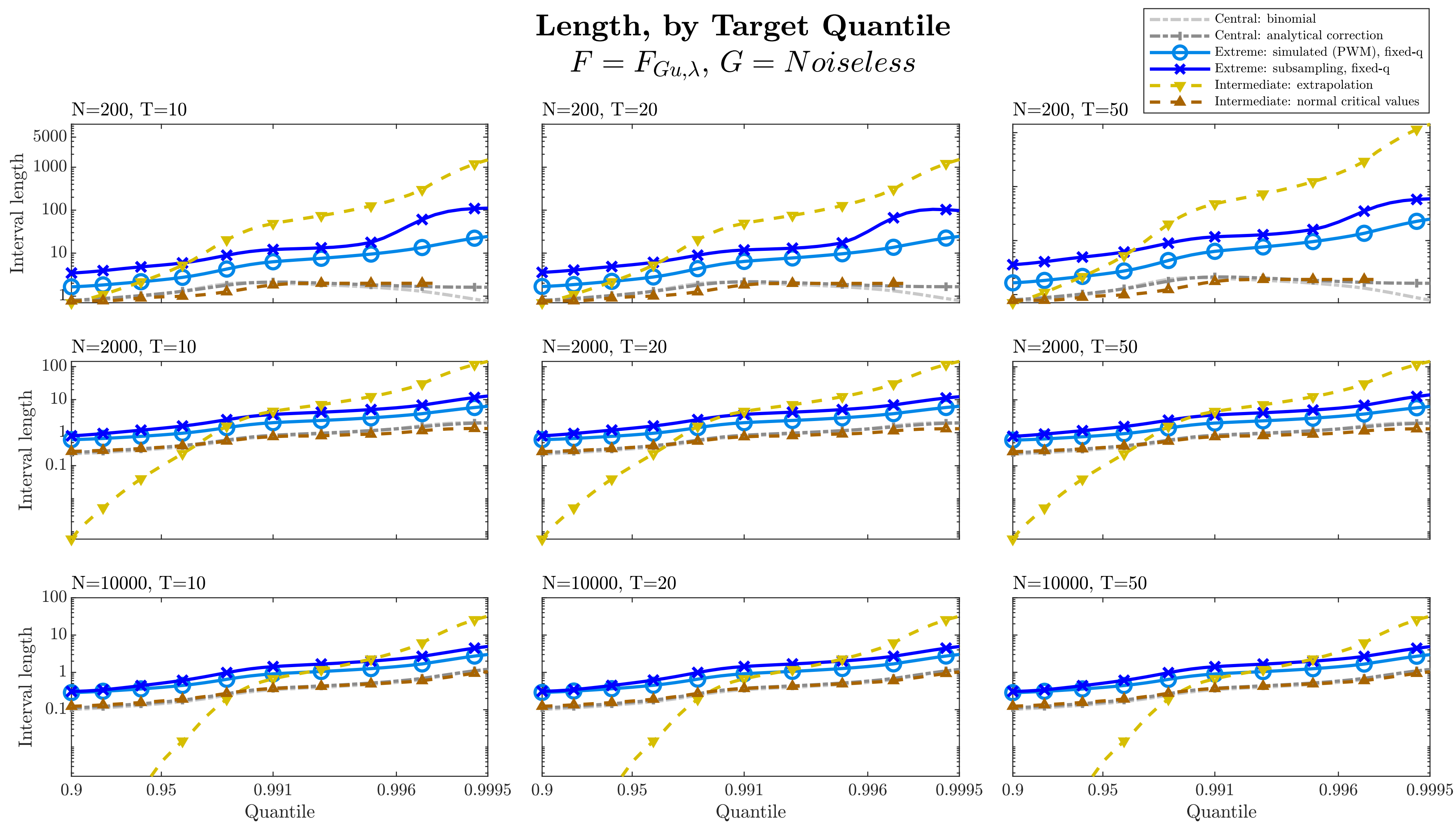

Figure OA.18: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

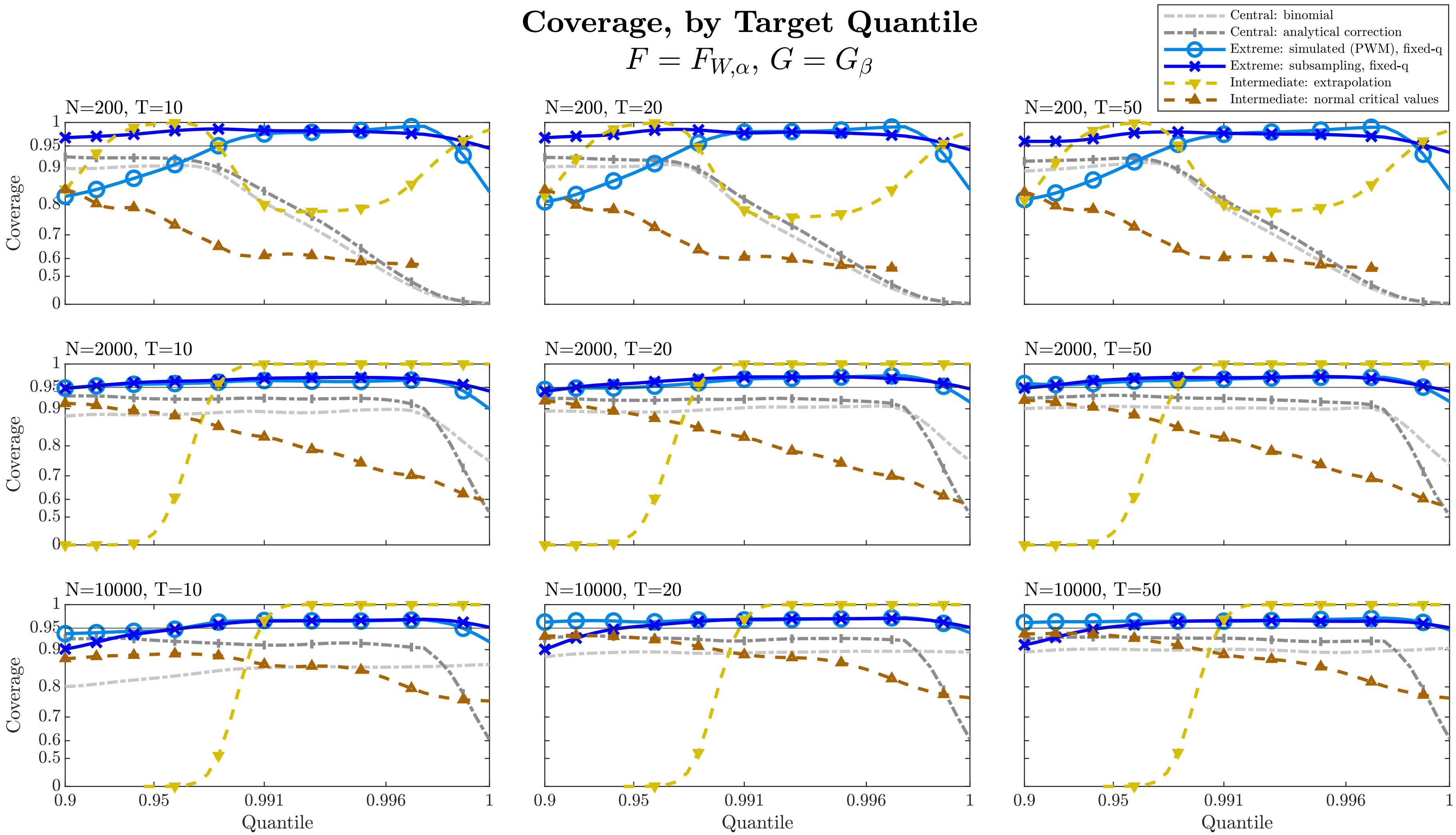

Figure OA.19: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim G_\beta, \beta = 8$, Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

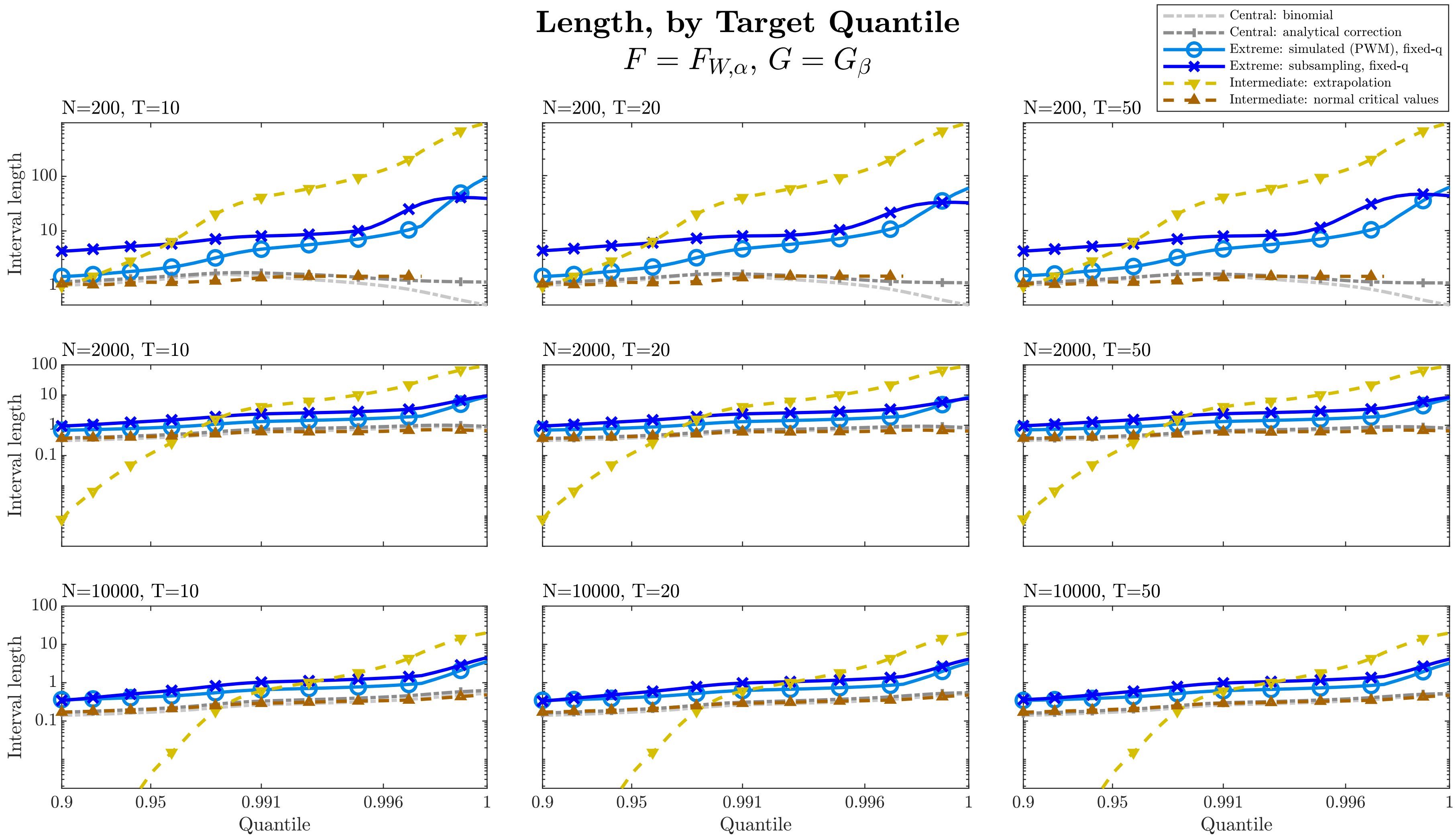

Figure OA.20: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, $u_{it} \sim G_\beta, \beta = 8$. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

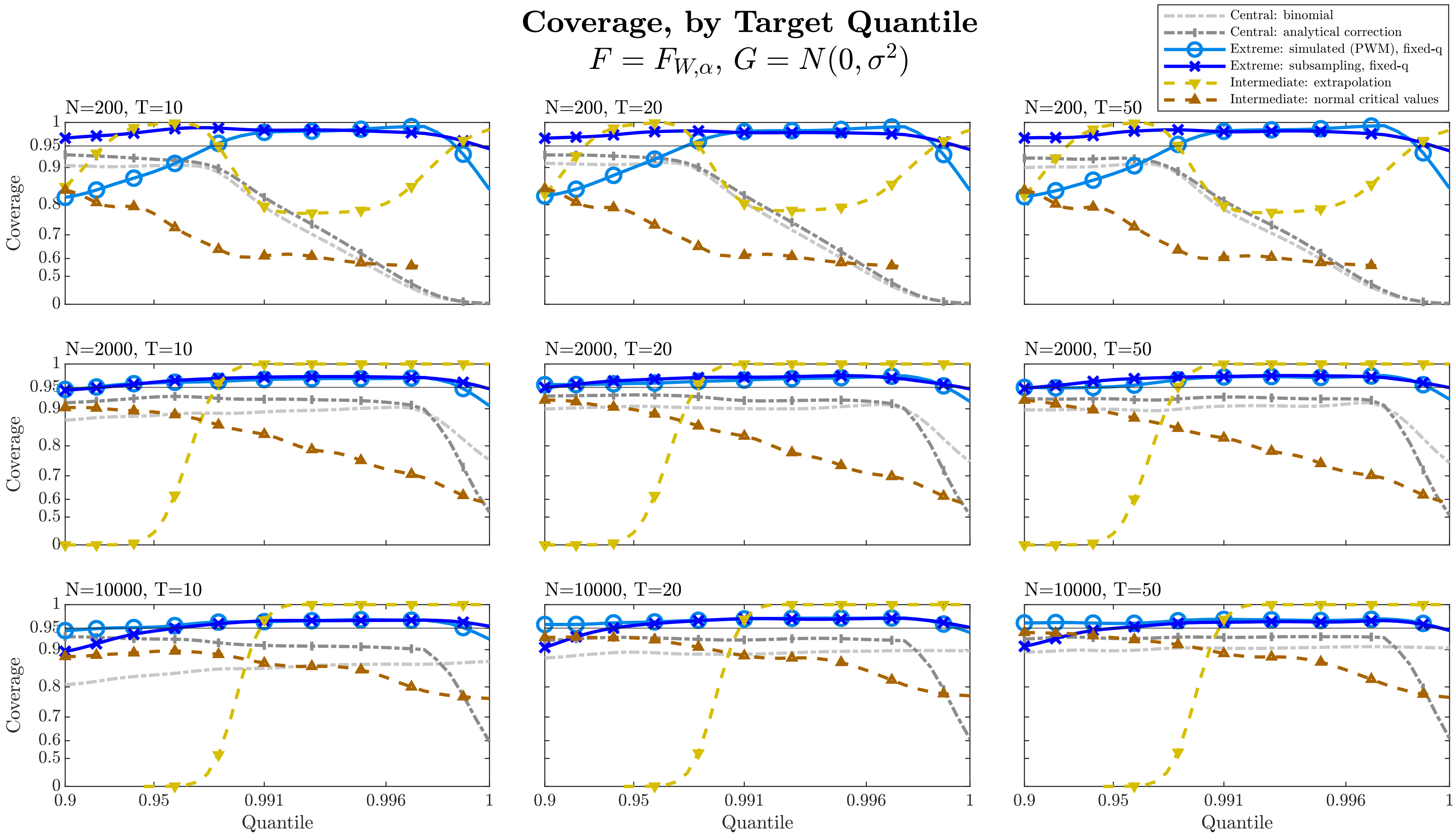

Figure OA.21: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

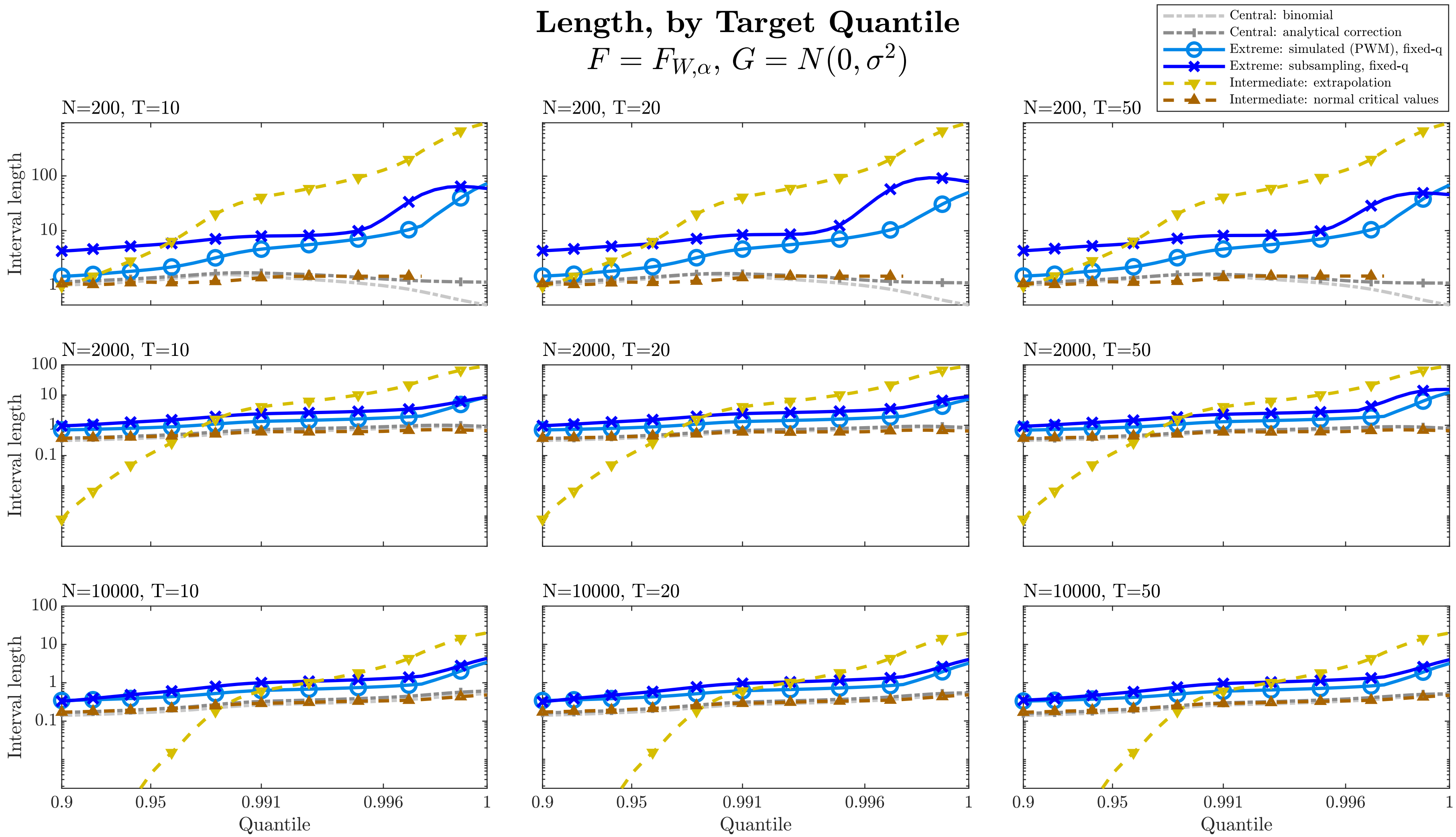

Figure OA.22: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

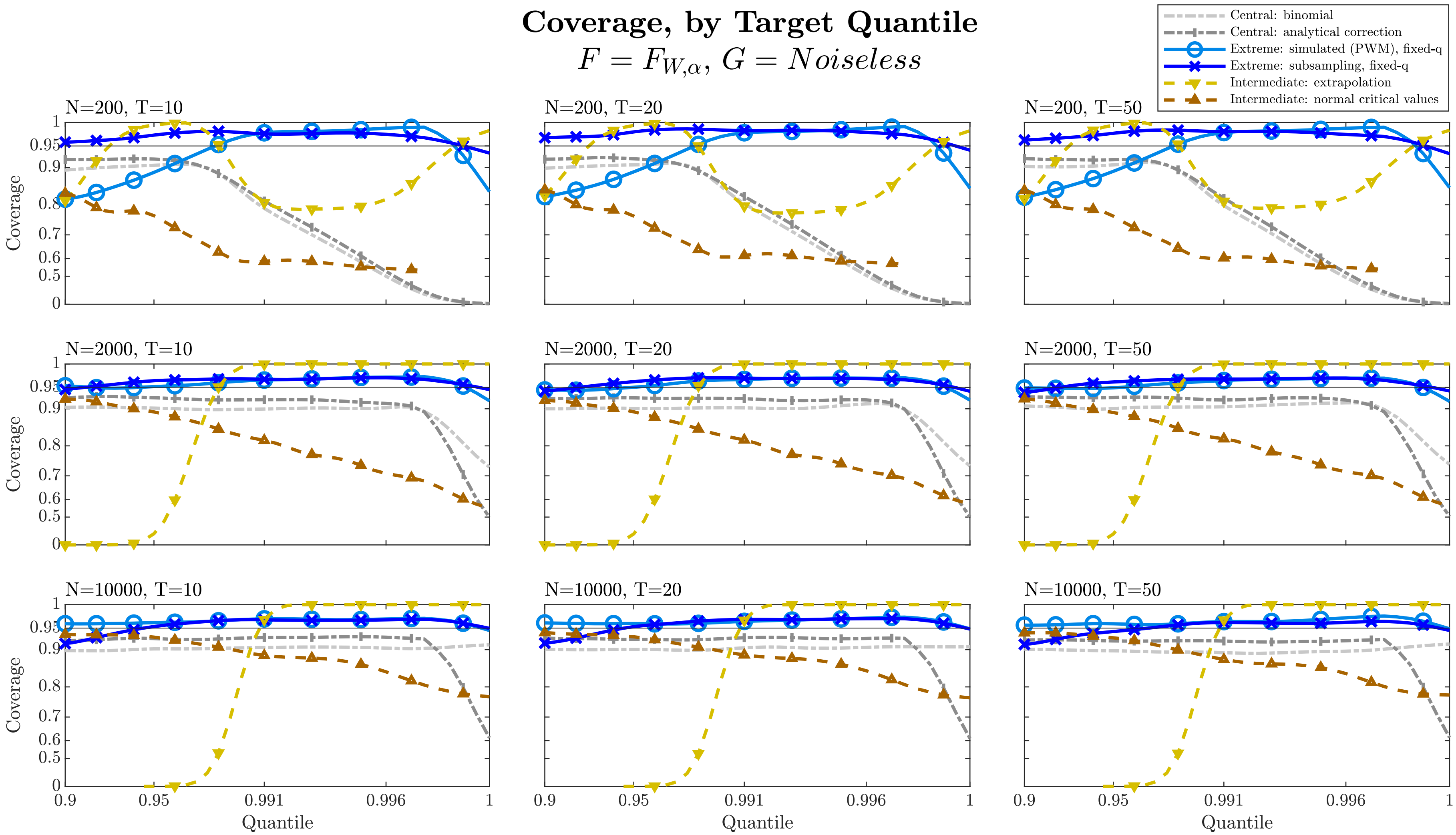

Figure OA.23: Coverage of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

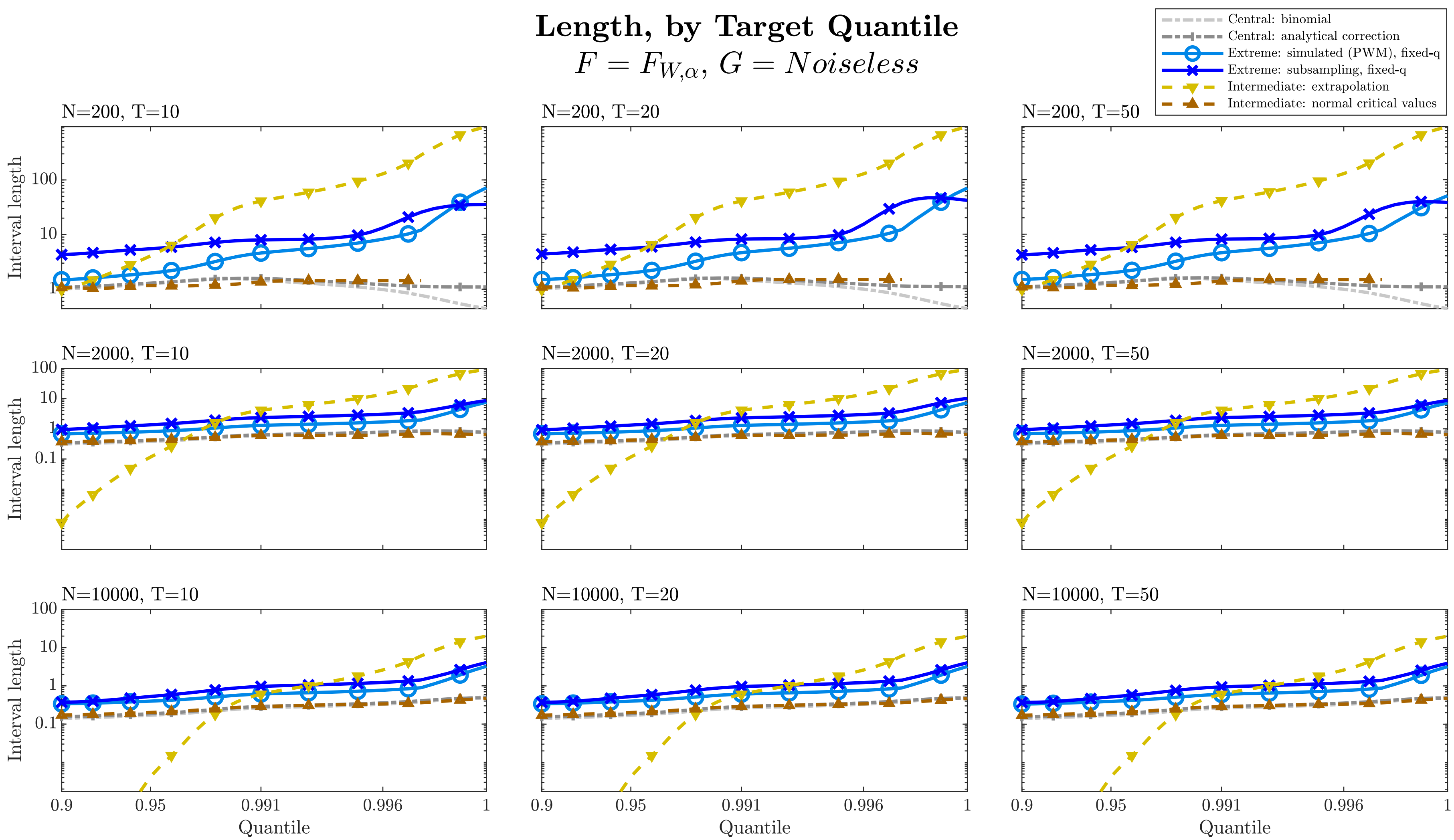

Figure OA.24: Length of a 95% confidence interval based on different approximations, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, noiseless data. Notes: (1) nonlinear x and y-axes; (2) intermediate CIs cannot be constructed for some quantiles (remark 10)

OA.4.2 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:

- For CIs with simulated critical values:
  - 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 ($N \lesssim 2000$), 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 $r = 0$.
- For CIs with subsampled critical values:
  - If $(1-\tau)N \leq 100$, we suggest $q = 2$ in smaller cross-sections ($N \lesssim 2000$); larger values may be used in larger cross-sections. For r, we again suggest matching $\vartheta_{N-r,N,T}$ 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 sections 5 and OA.4.1. 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 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.25-OA.32. Specifically, figures OA.25-OA.28 report the coverage and length properties for a range of values of q. Throughout, we match r to the sample quantiles. Figures OA.29-OA.32 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 that the central-order CI with analytical correction (compare figs. OA.30 and OA.32 to fig. OA.2). 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.25). 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.29). 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.25). 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 3.1 than theorem 3.3. 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.29).

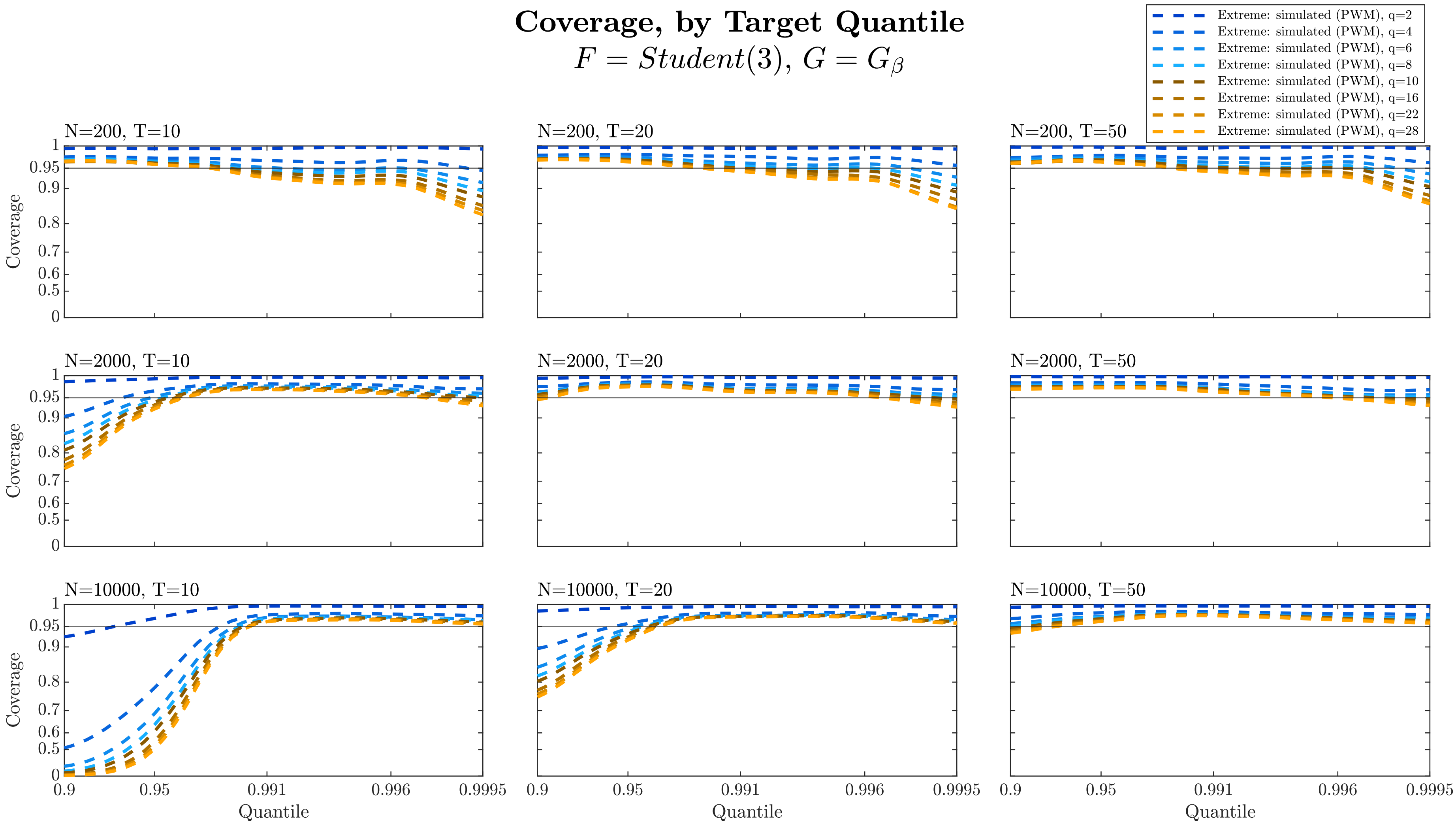

Figure OA.25: Impact of choice of denominator parameter q on coverage of a 95% confidence interval based on an extreme value approximation with simulated critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

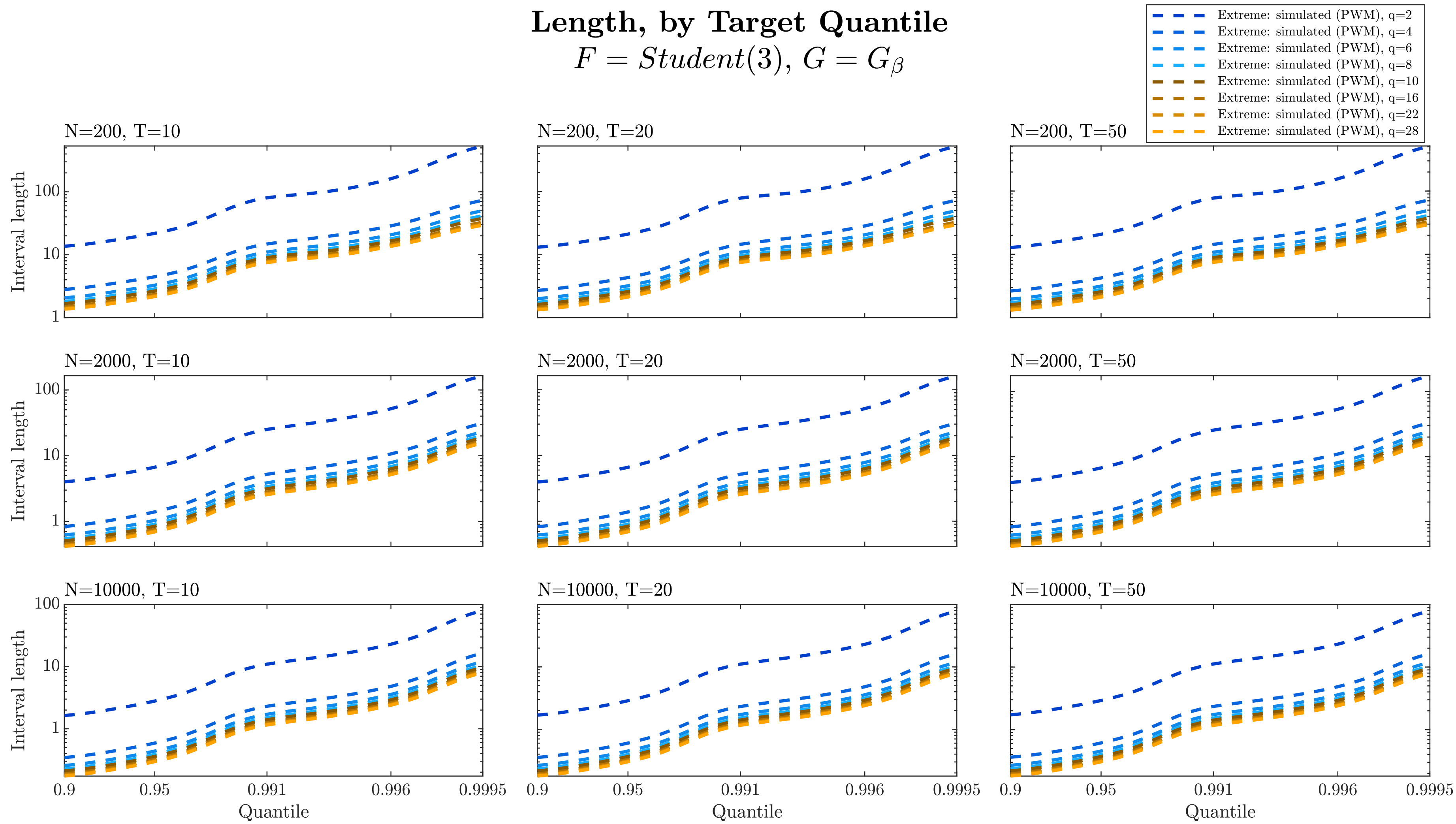

Figure OA.26: Impact of choice of denominator parameter q on length of a 95% confidence interval based on an extreme value approximation with simulated critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

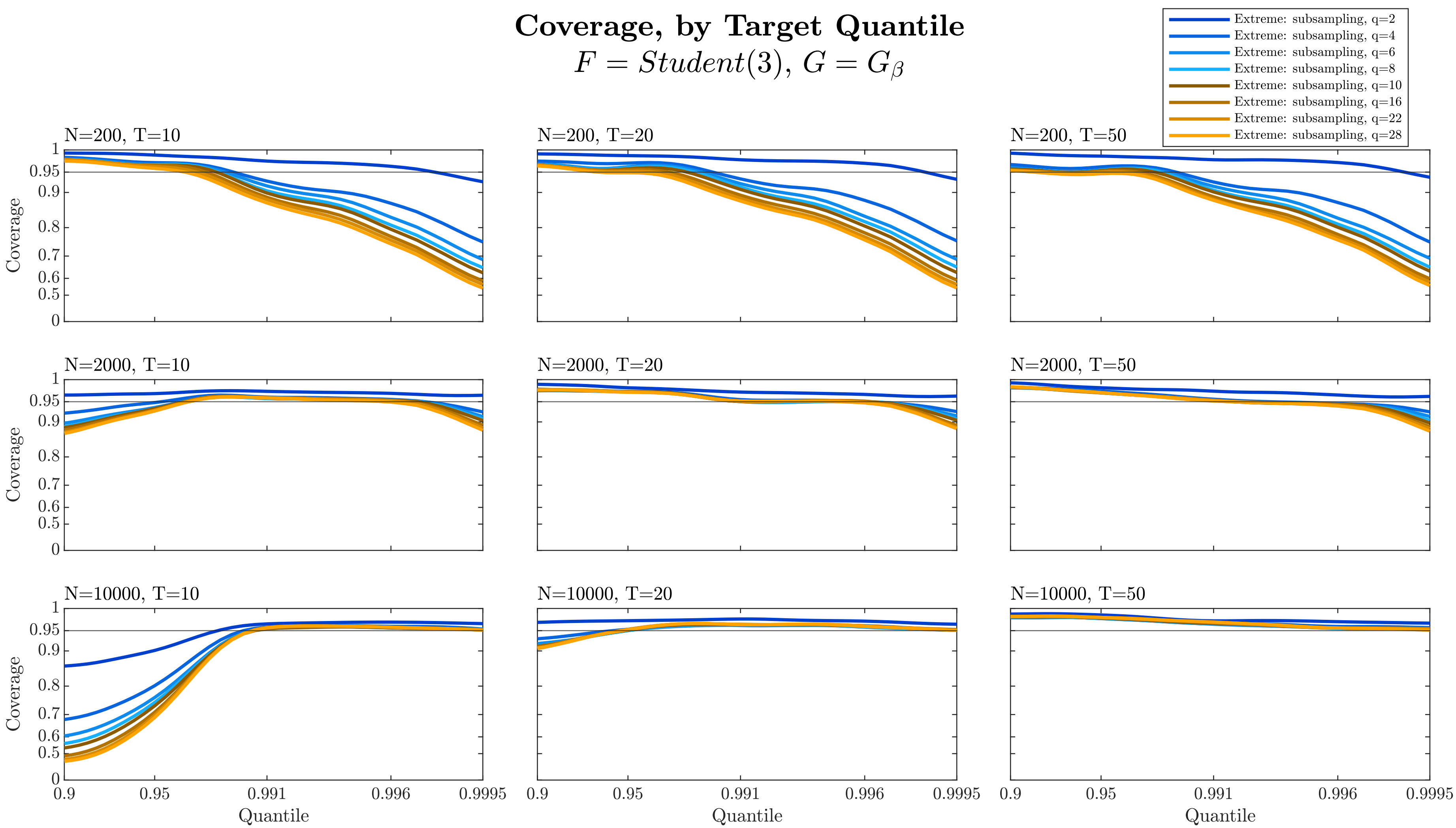

Figure OA.27: Impact of choice of denominator parameter q on coverage of a 95% confidence interval based on extreme value approximations with subsampled critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

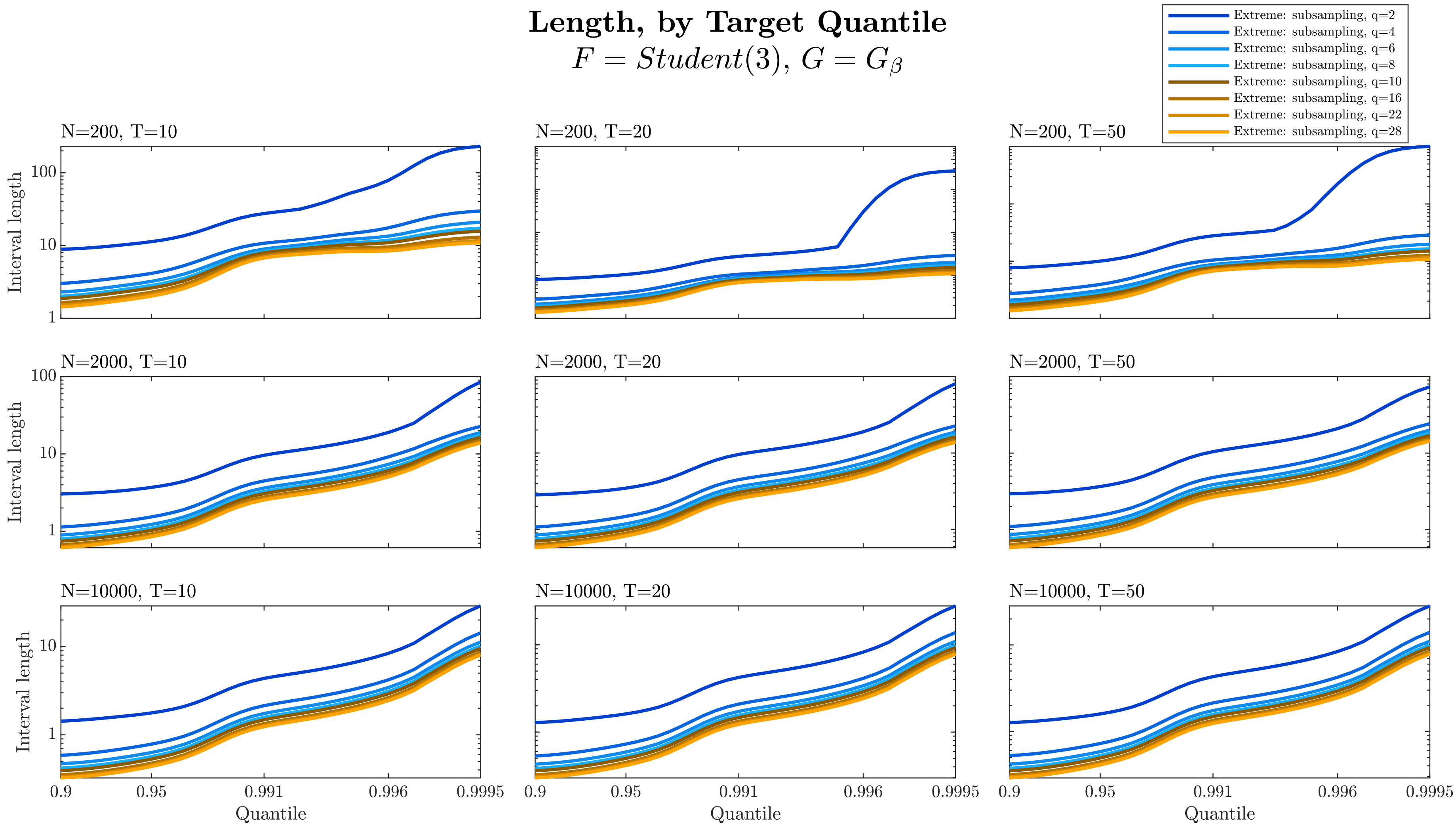

Figure OA.28: Impact of choice of denominator parameter q on length of a 95% confidence interval based on extreme value approximations with subsampled critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- y-axes. Throughout, r is chosen to match the sample τth quantile.

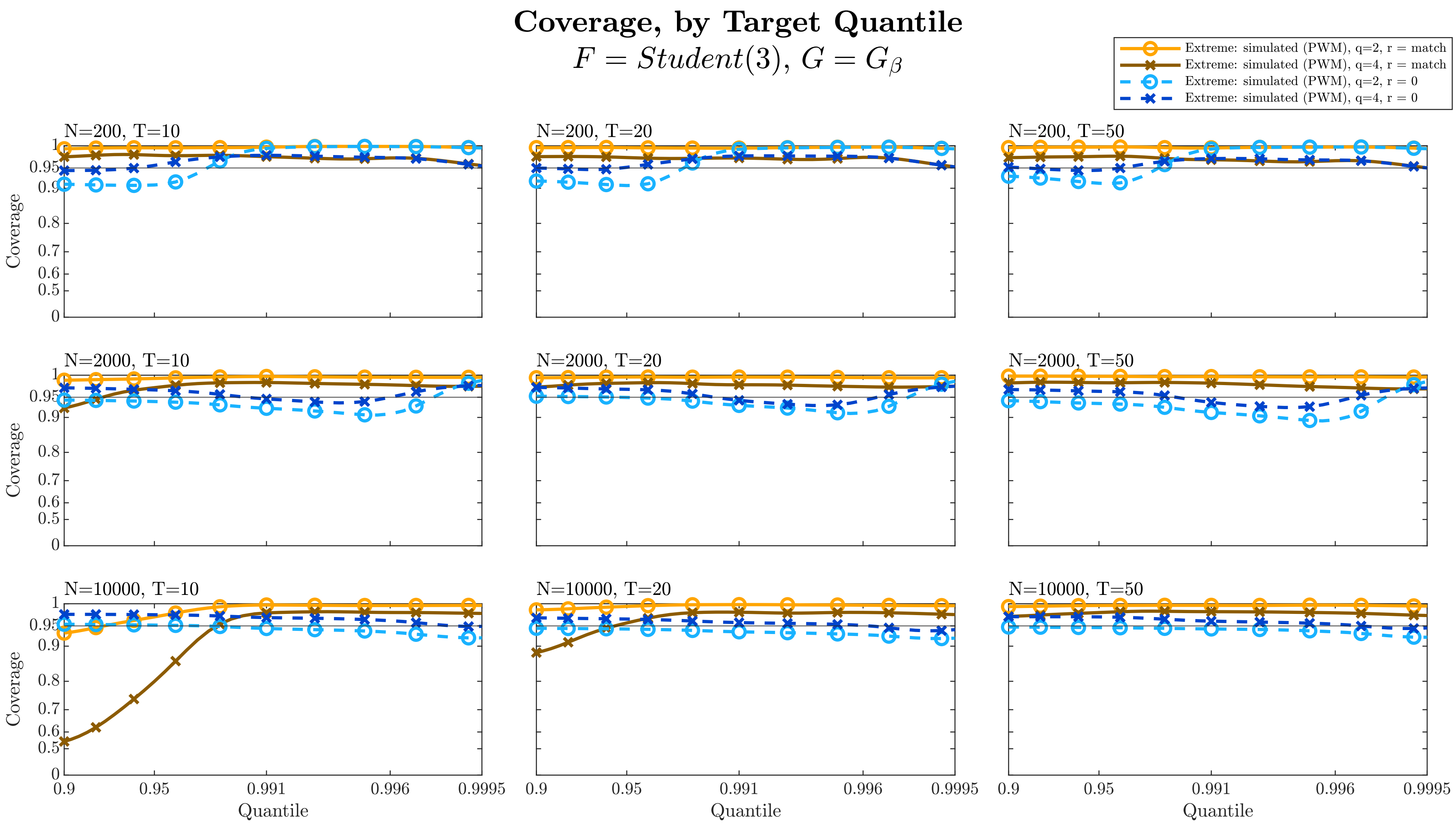

Figure OA.29: Impact of tuning parameter choice on length of a 95% confidence interval based on extreme value approximations with simulated critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- and y-axes.

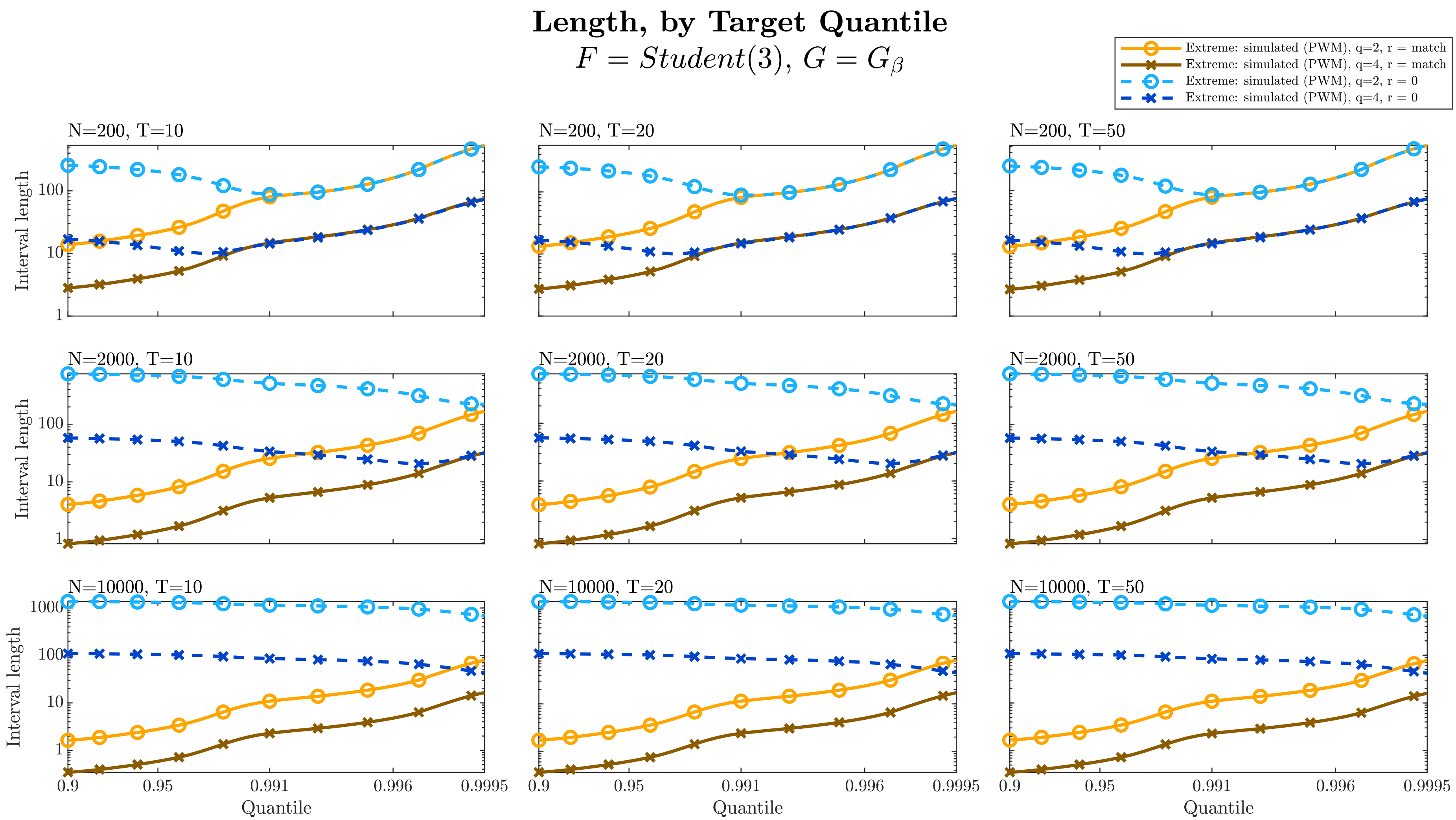

Figure OA.30: Impact of tuning parameter choice on length of a 95% confidence interval based on extreme value approximations with simulated critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- and y-axes.

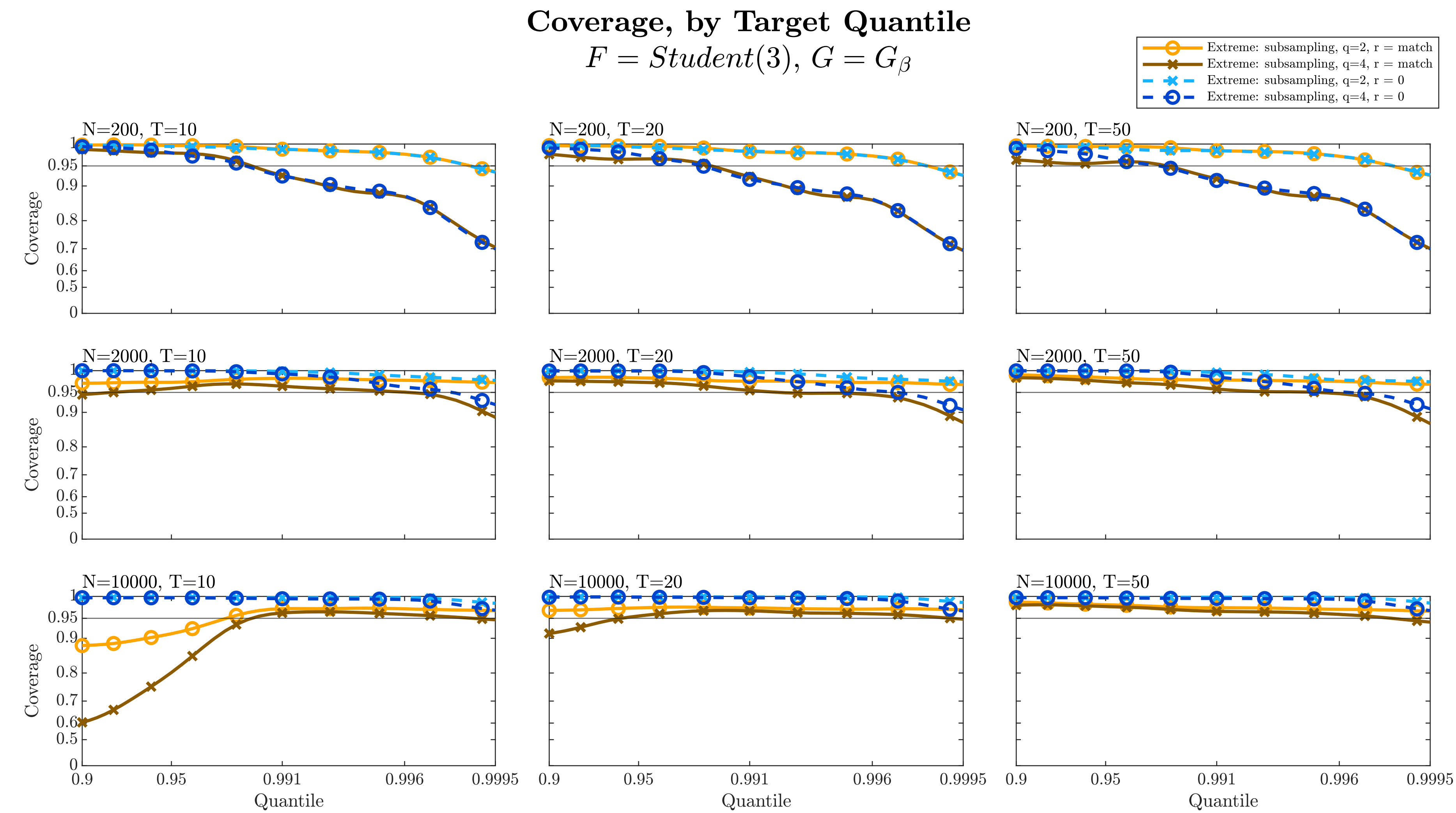

Figure OA.31: Impact of tuning parameter choice on length of a 95% confidence interval based on extreme value approximations with subsampled critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- and y-axes.

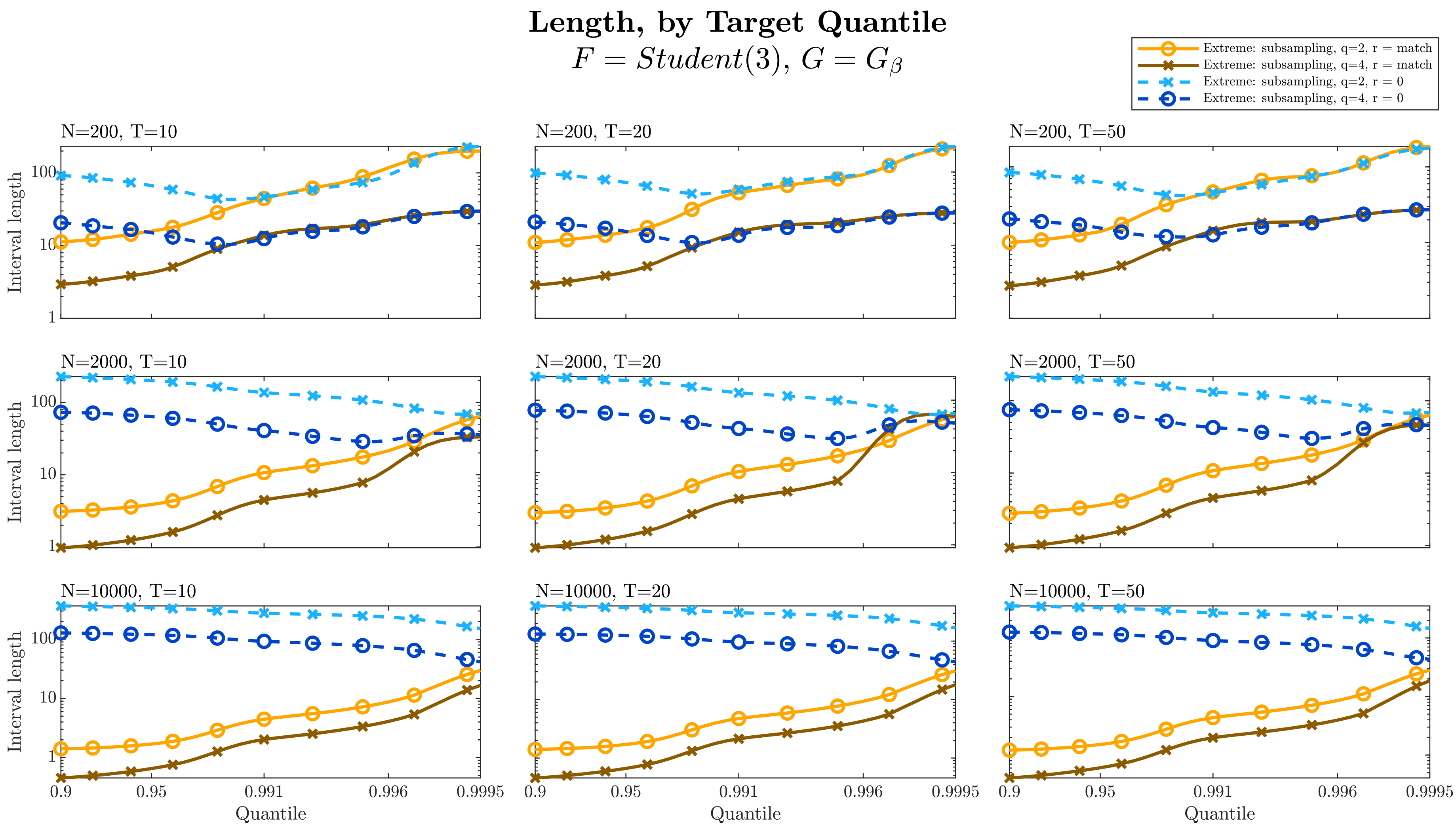

Figure OA.32: Impact of tuning parameter choice on length of a 95% confidence interval based on extreme value approximations with subsampled critical values, $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$. Note: nonlinear x- and y-axes.

OA.4.3 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 work in the setup of section 5 in the main text and consider all the specifications for (F, G) and (N, T) of section OA.4.1.

Figures OA.33-OA.44 summarize 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 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.

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 $N = 2000$ and $N = 10000$ (e.g. fig. OA.33). For $N = 200$, the two corrections perform worse than the benchmark above the 0.995th quantile for heavier-tailed distributions (those with the EV index $\gamma > 0$ — $F_{Fr,\kappa}, t_3$).

We highlight two points of interest that arise in noiseless settings (e.g. figure OA.35). First, the analytical correction of Jochmans and Weidner (2024) also leads to improvements in the MSE even if no estimation noise is present. Moreover, the improvement is uniform across the quantiles considered. The corrected estimator is also not dominated by any of the other estimators considered, although the extrapolation estimator and the two extreme-order corrections offer stronger gains in some regions. Second, the "intermediate" estimator also outperforms the benchmark estimator above the 0.95th quantile. It differs from the benchmark estimator is that it directly uses the "closest from the left" order statistic $\theta_{N-\lceil\tau N\rceil,N}$ (or $\vartheta_{N-\lceil\tau N\rceil,N,T}$) instead of interpolating the two nearest order statistics. The results of figs. OA.33-OA.44 then show that it is preferable to not interpolate order statistics above the 0.95th quantile, regardless of sample size. Below that threshold the benchmark estimator is slightly more efficient, and should be preferred.

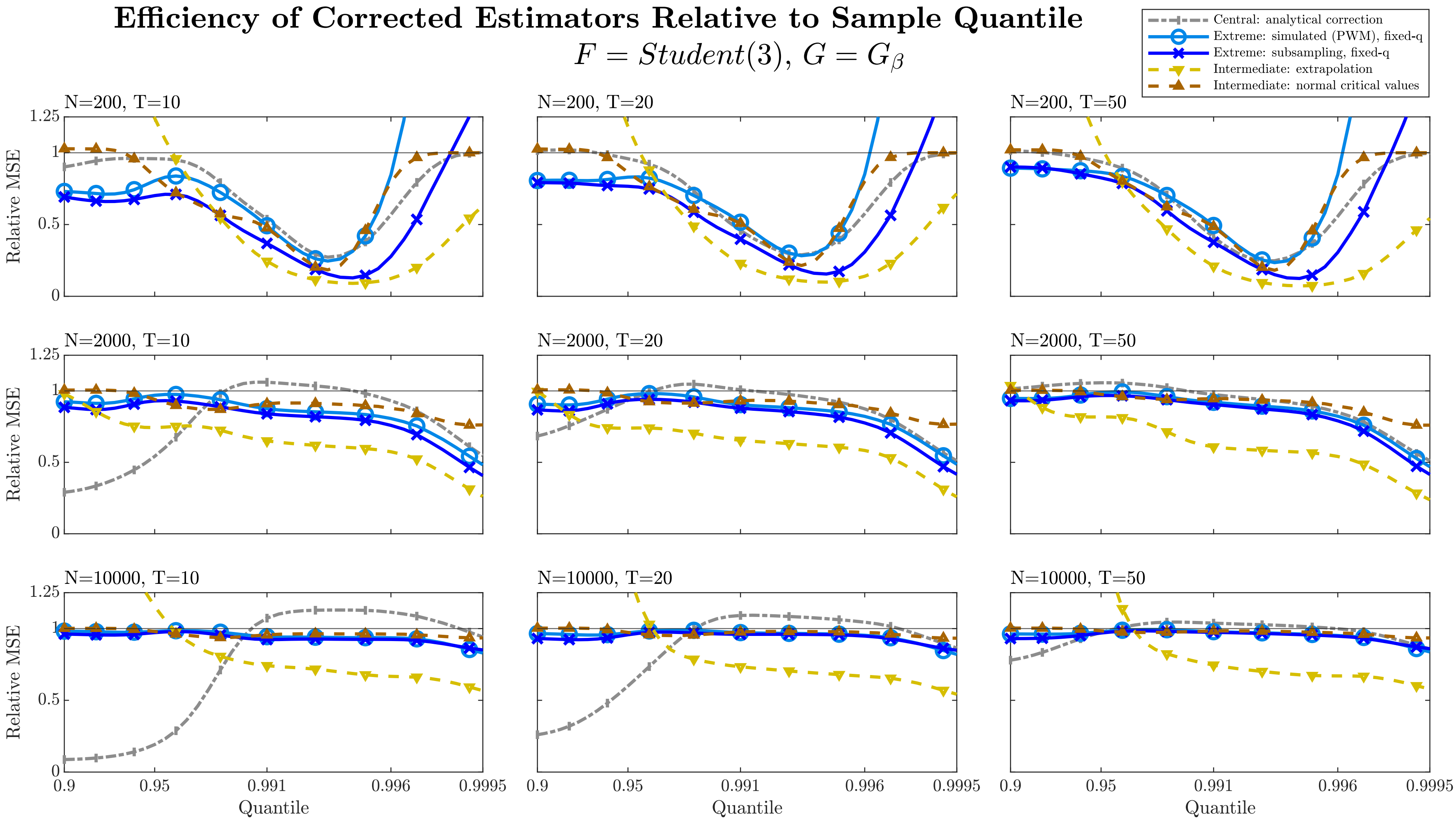

Figure OA.33: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$

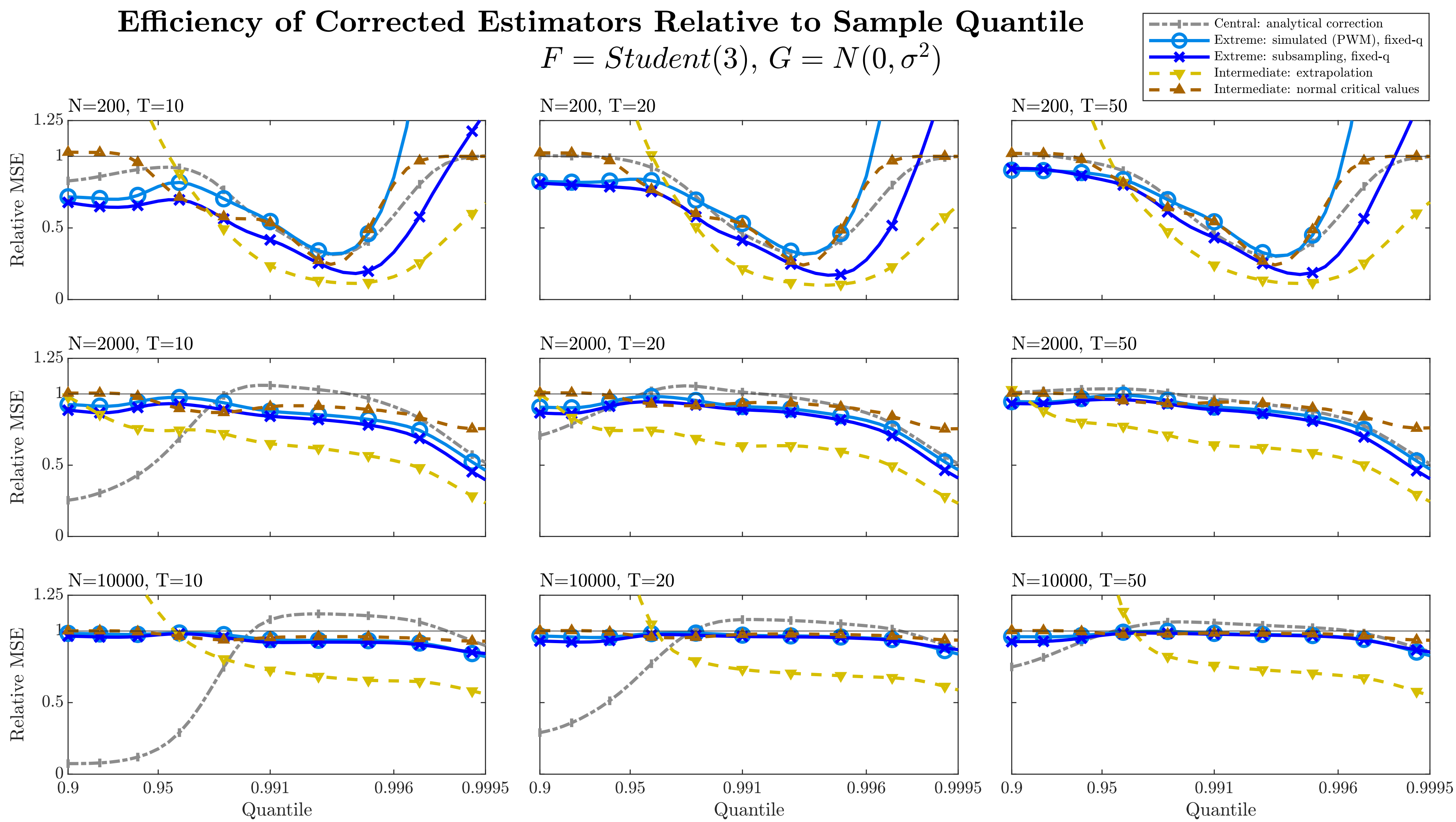

Figure OA.34: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients.

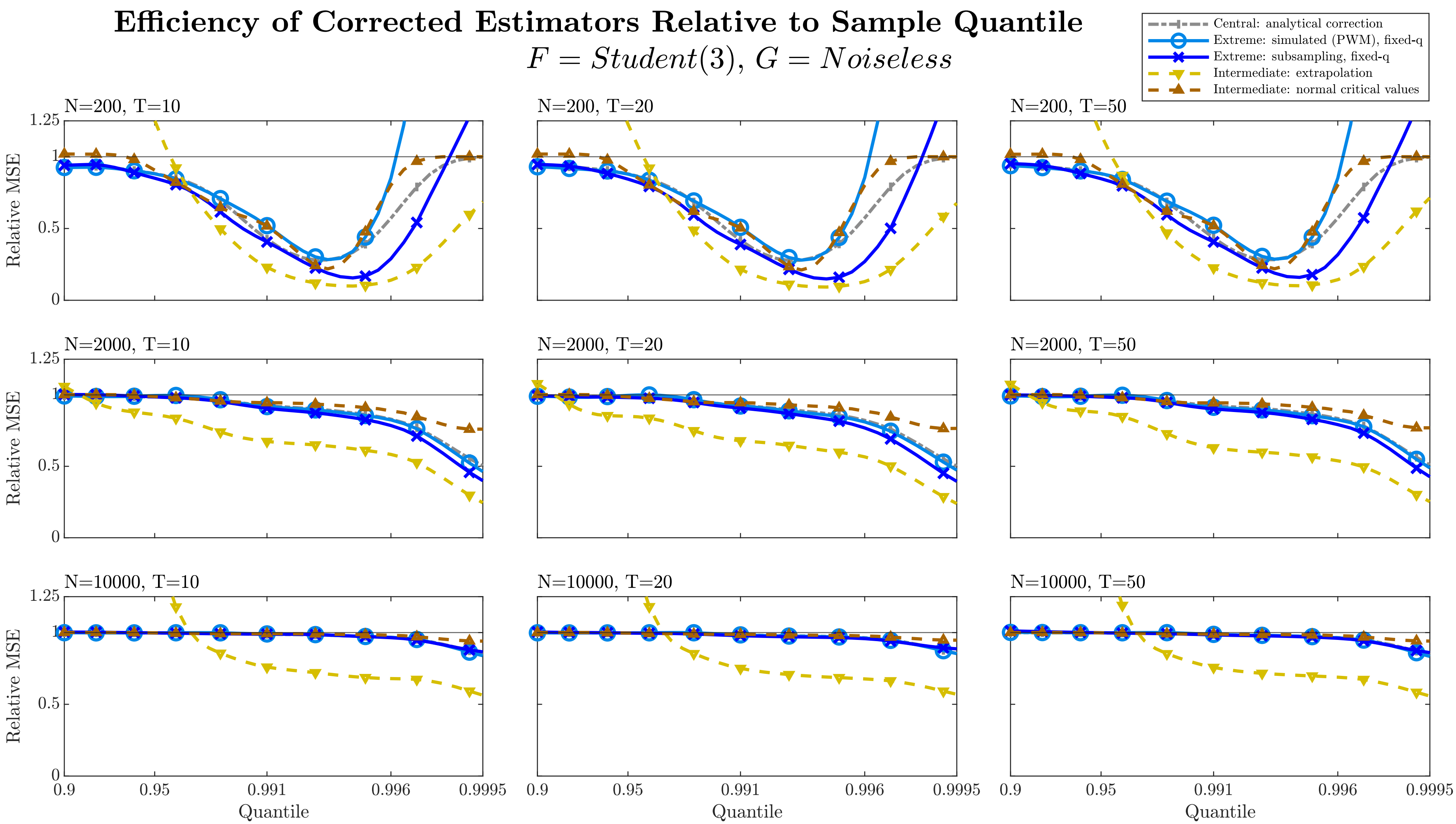

Figure OA.35: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, noiseless data.

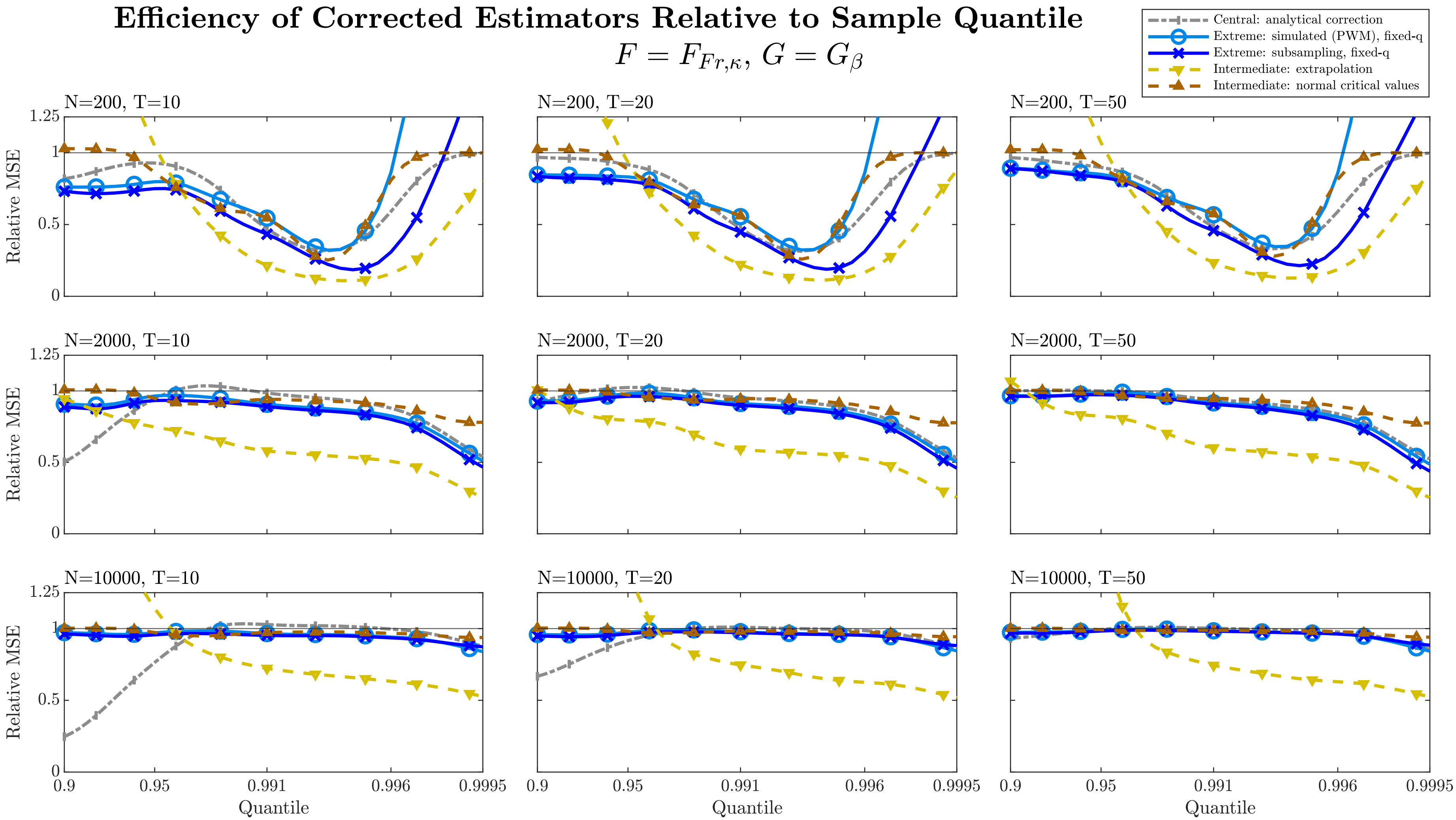

Figure OA.36: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim t_3$, $u_{it} \sim G_\beta, \beta = 8$

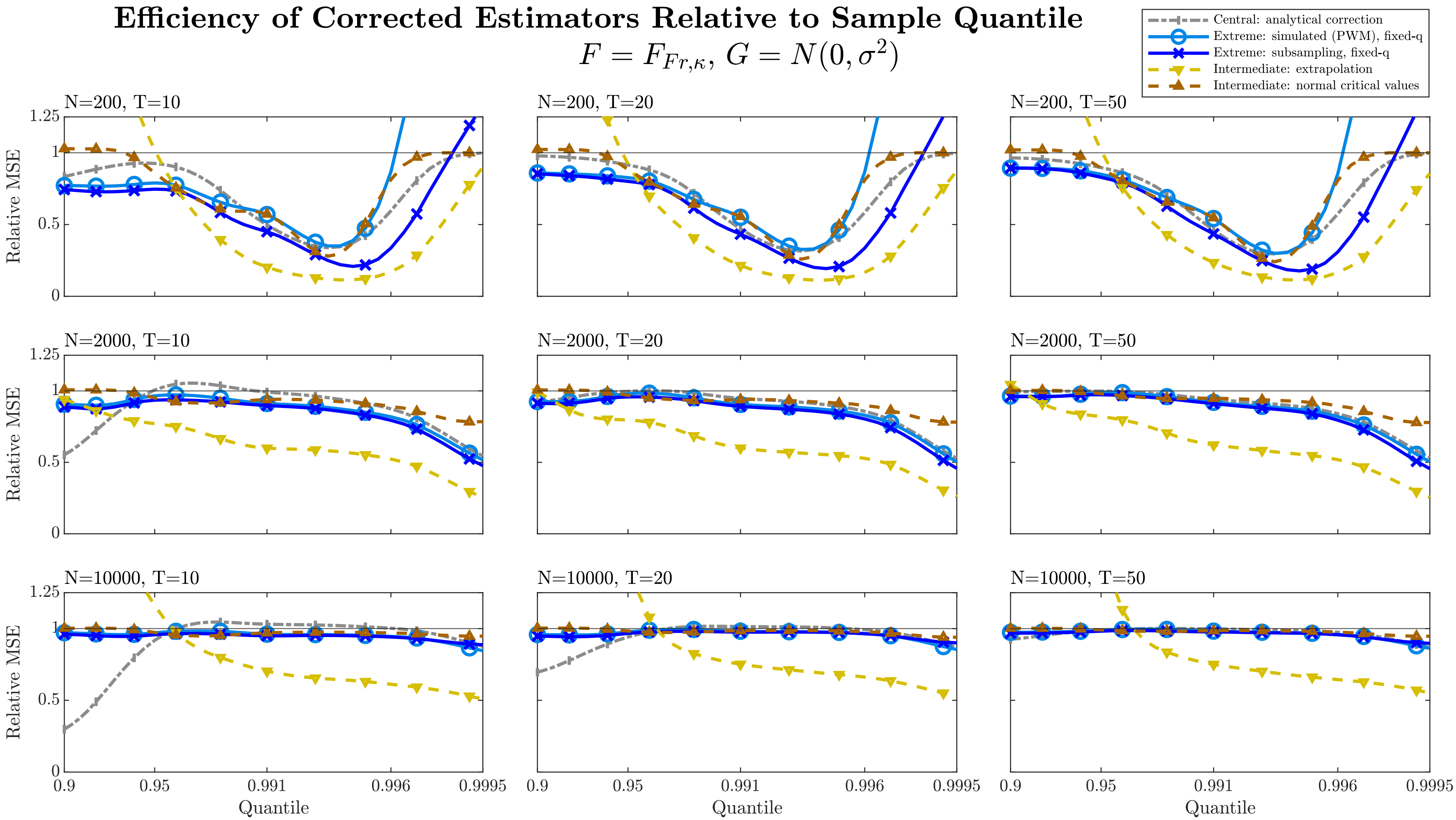

Figure OA.37: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients.

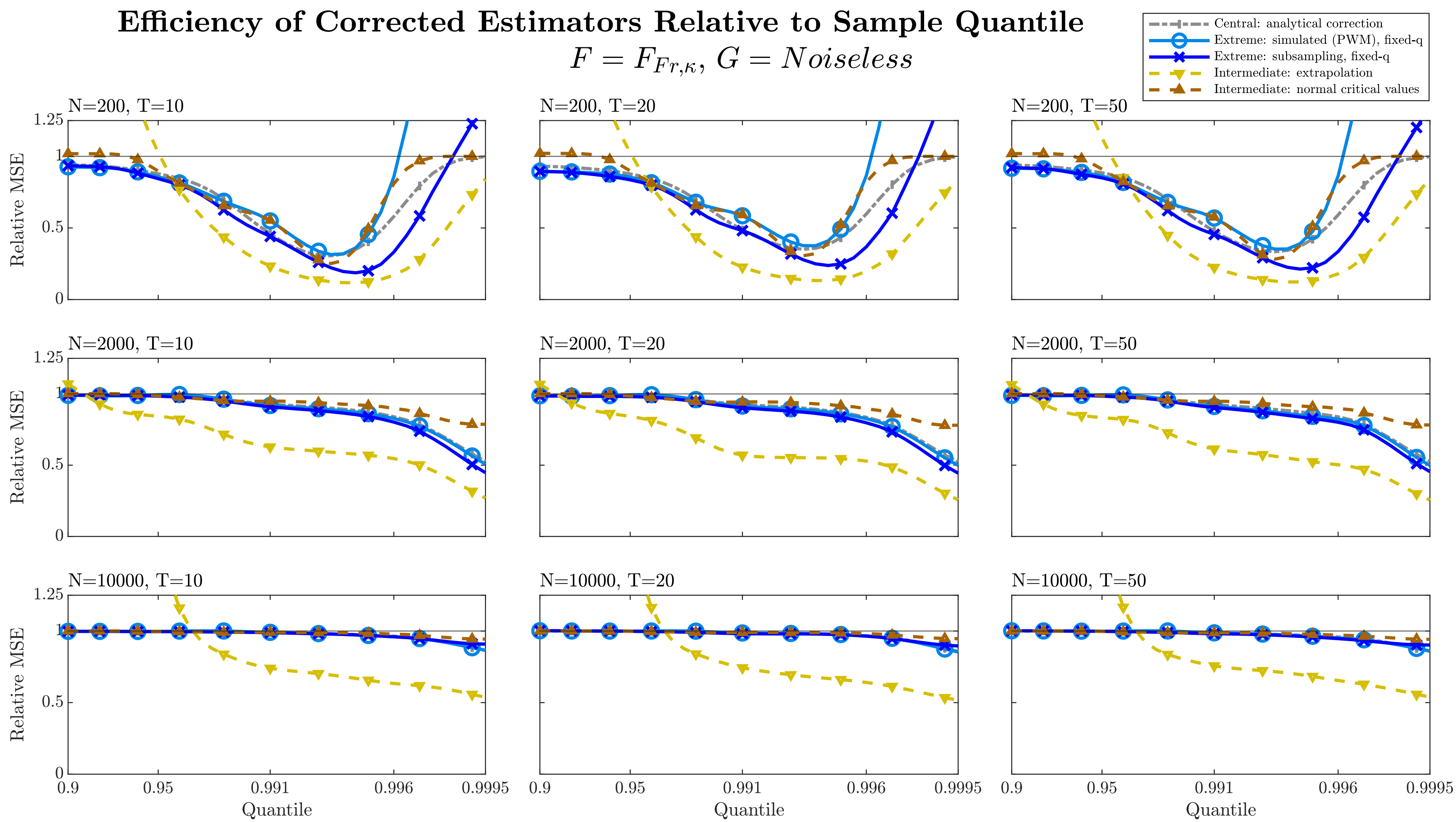

Figure OA.38: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Fr,4}$, noiseless data.

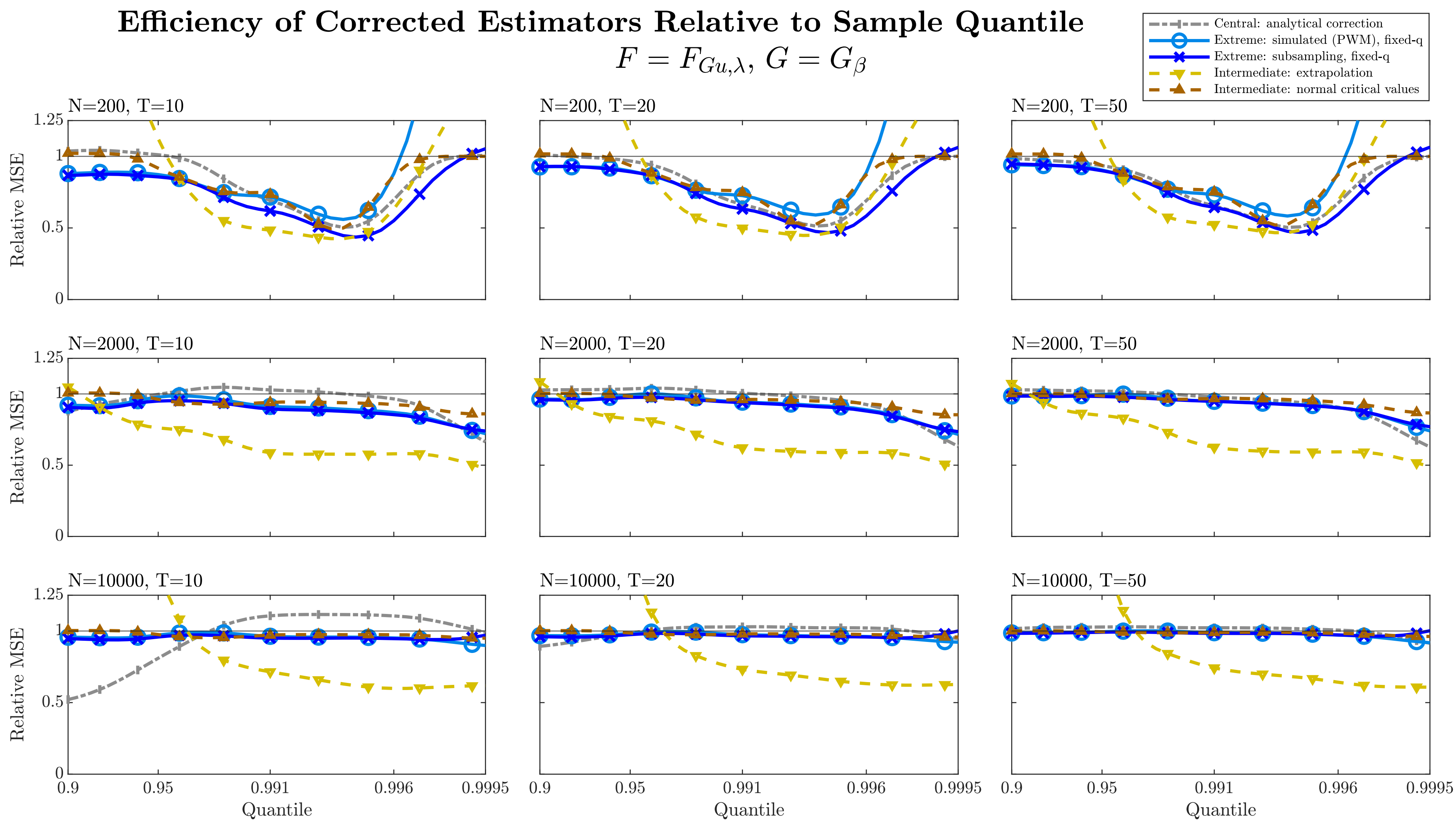

Figure OA.39: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim G_\beta, \beta = 8$

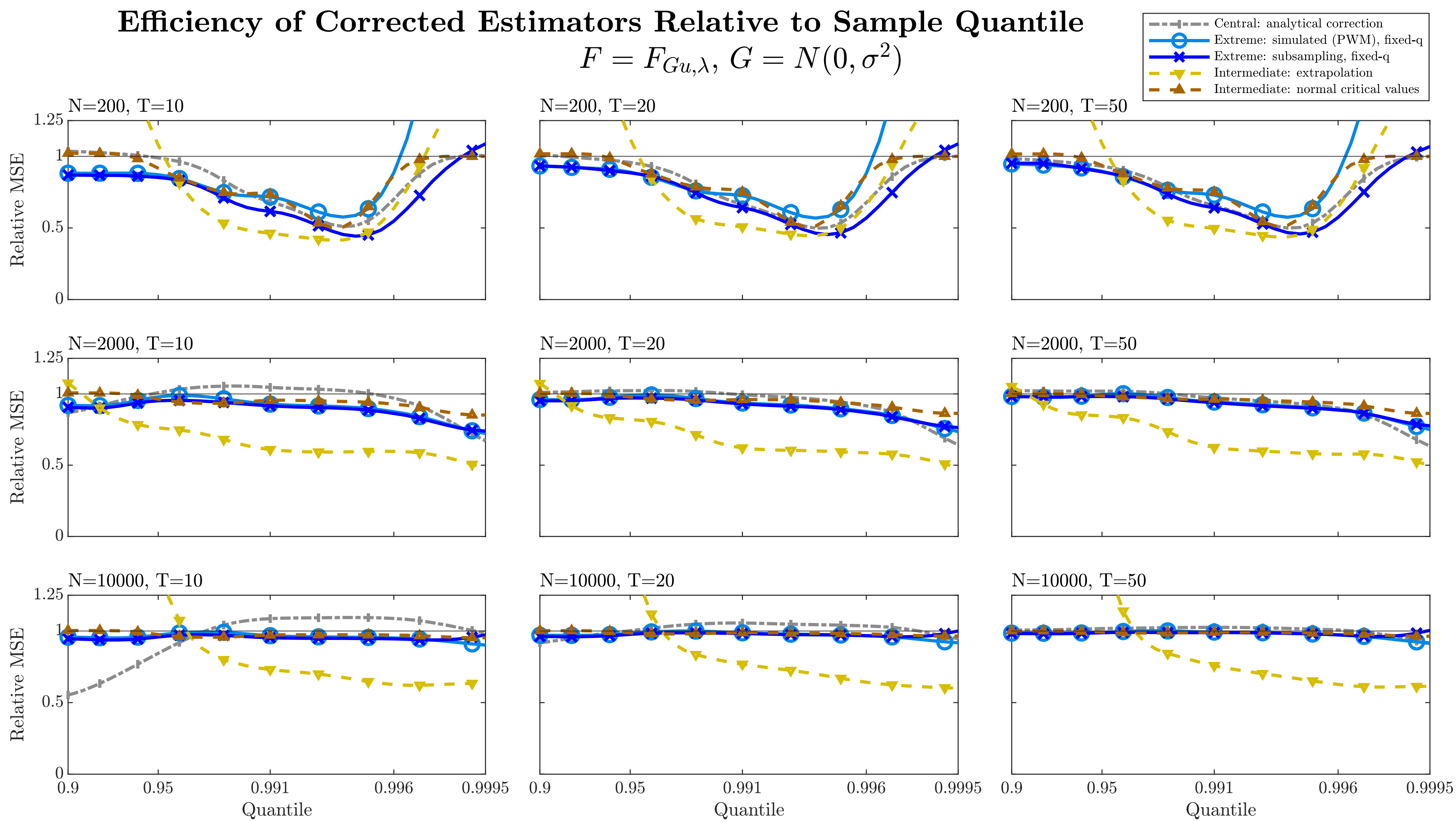

Figure OA.40: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients.

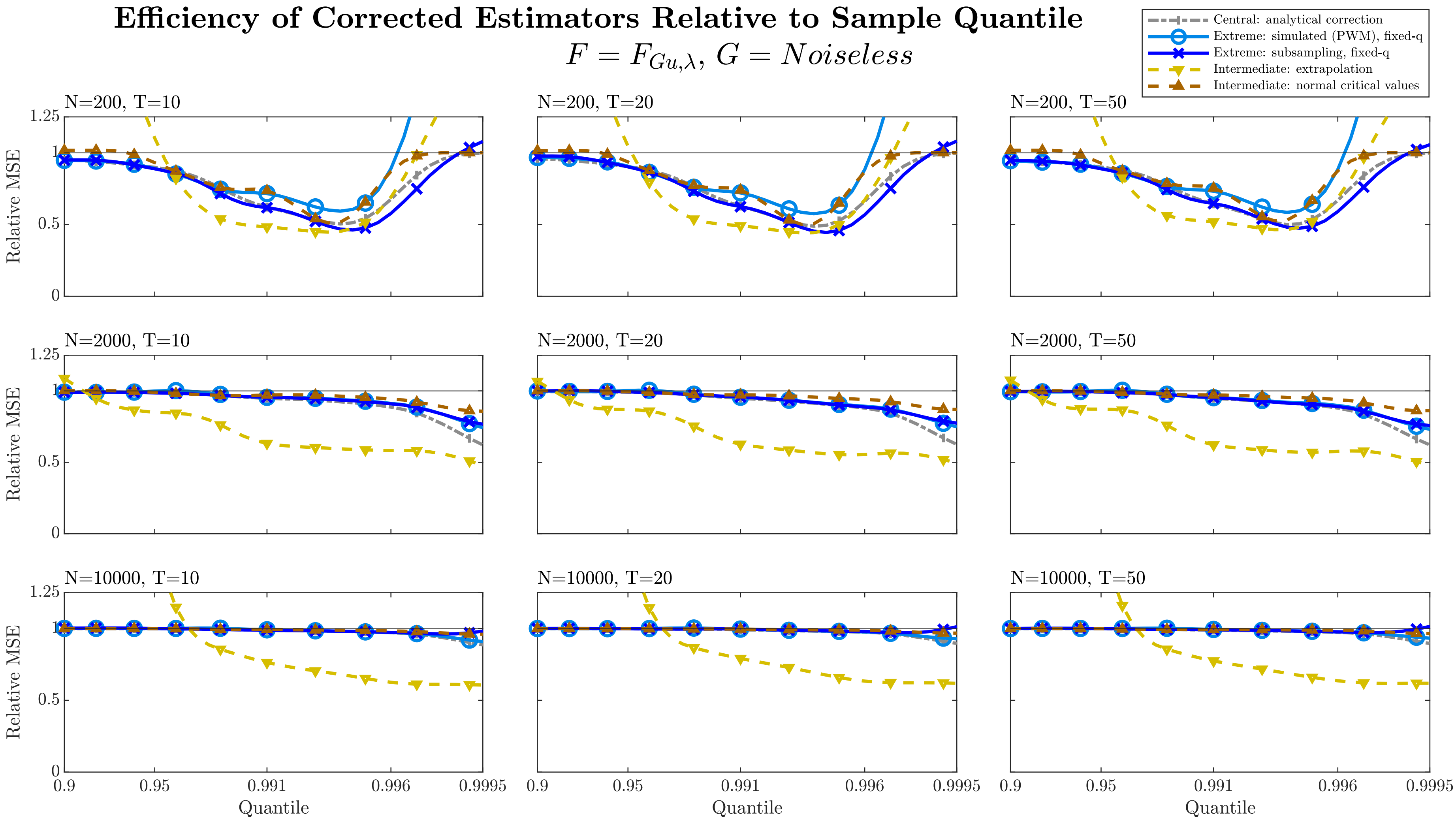

Figure OA.41: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{Gu,1}$, noiseless data.

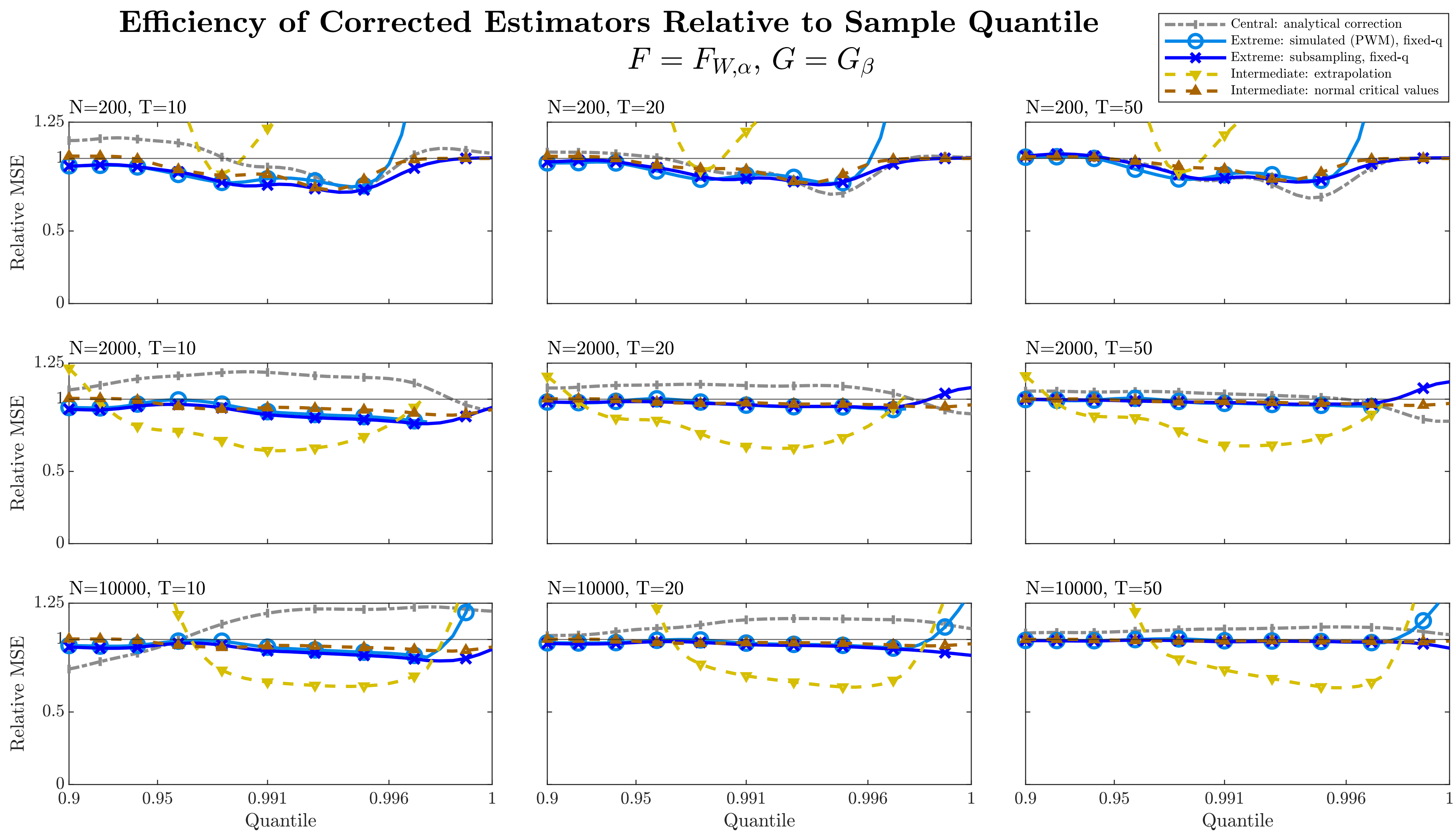

Figure OA.42: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, $u_{it} \sim G_\beta, \beta = 8$

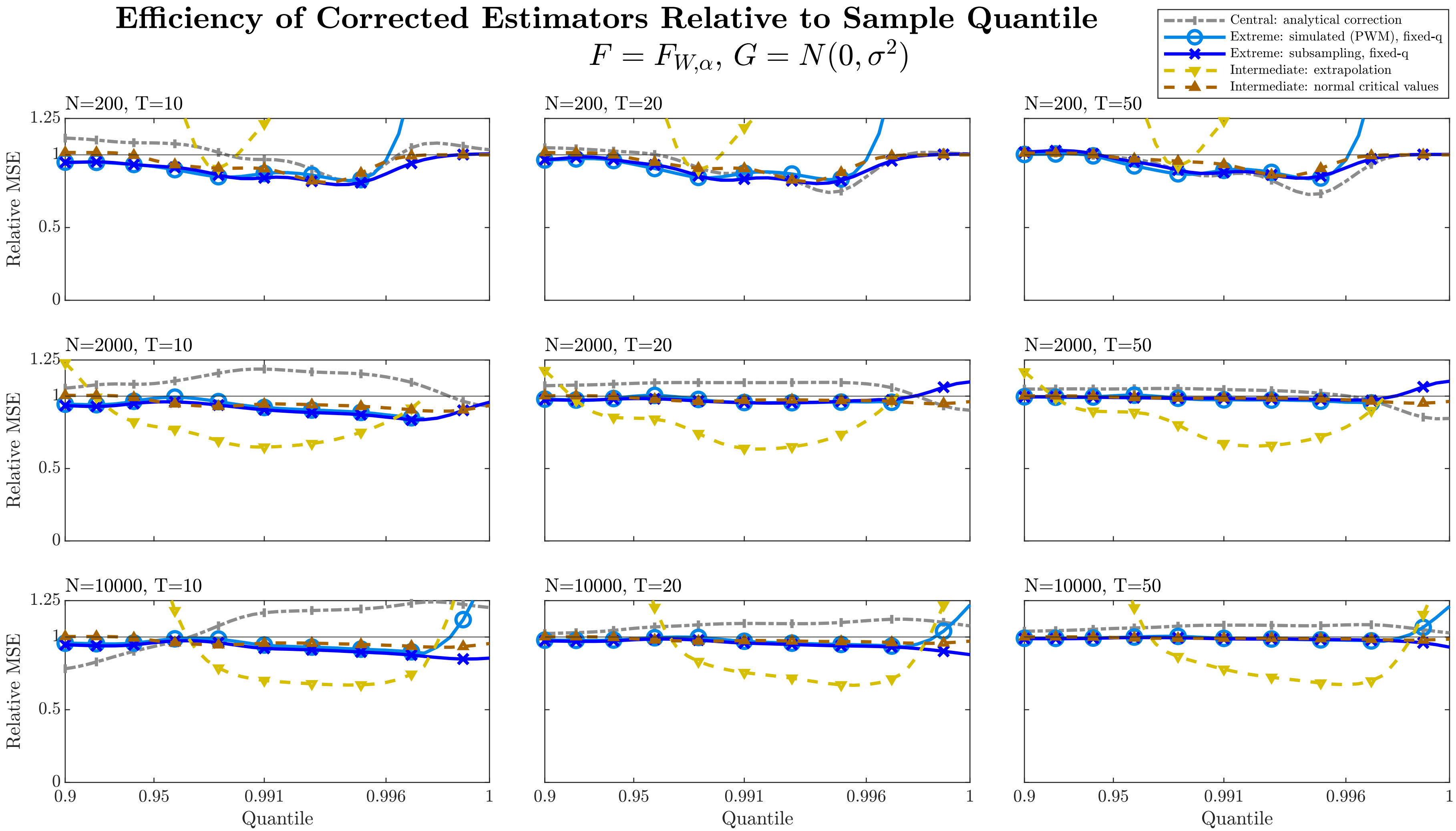

Figure OA.43: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, $u_{it} \sim N(0, \sigma^2)$, σ^2 chosen to match variance of coefficients.

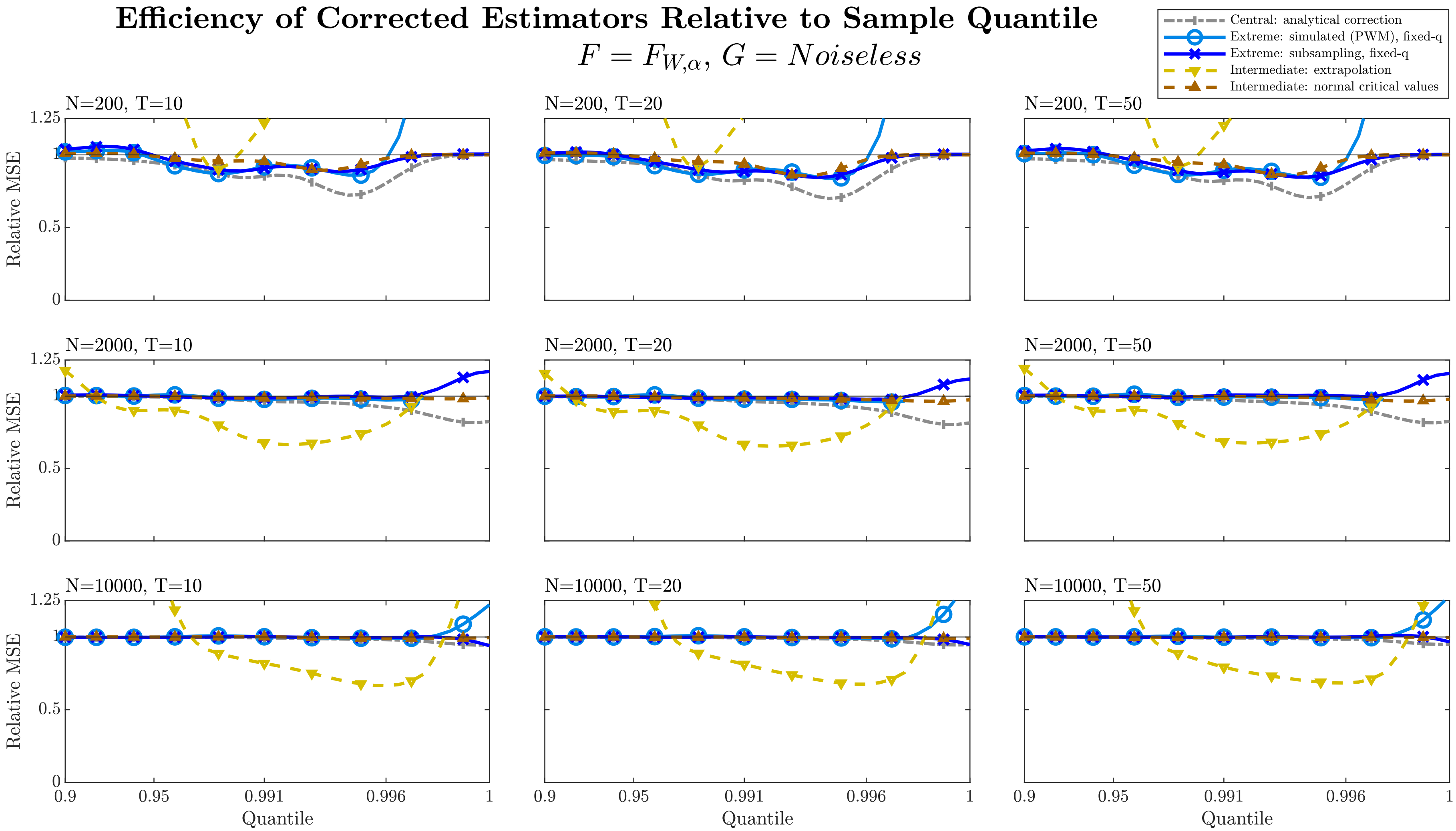

Figure OA.44: MSE of various corrected estimators relative to the MSE of sample quantiles. $N = 200, 2000, 10000$, $T = 10, 20, 50$. $\theta \sim F_{W,4}$, noiseless data.

OA.4.4 Quality of Intermediate Approximation

The confidence intervals based on theorem 4.5 begin to show adequate performance only for the largest sample size of $N = 10000$ in the simulation studies of section 5 and OA.4.1 (for example, see figure OA.5). Furthermore, the coverage is only adequate if $(1-\tau)N \geq 100$, where τ is the target quantile (intuitively, if there are at least a 100 order statistics to the right of the sample τth quantile).

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.45, 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.

The simulation results demonstrate the above two convergence issues. First, the first two rows of fig. OA.45 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.45.

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).

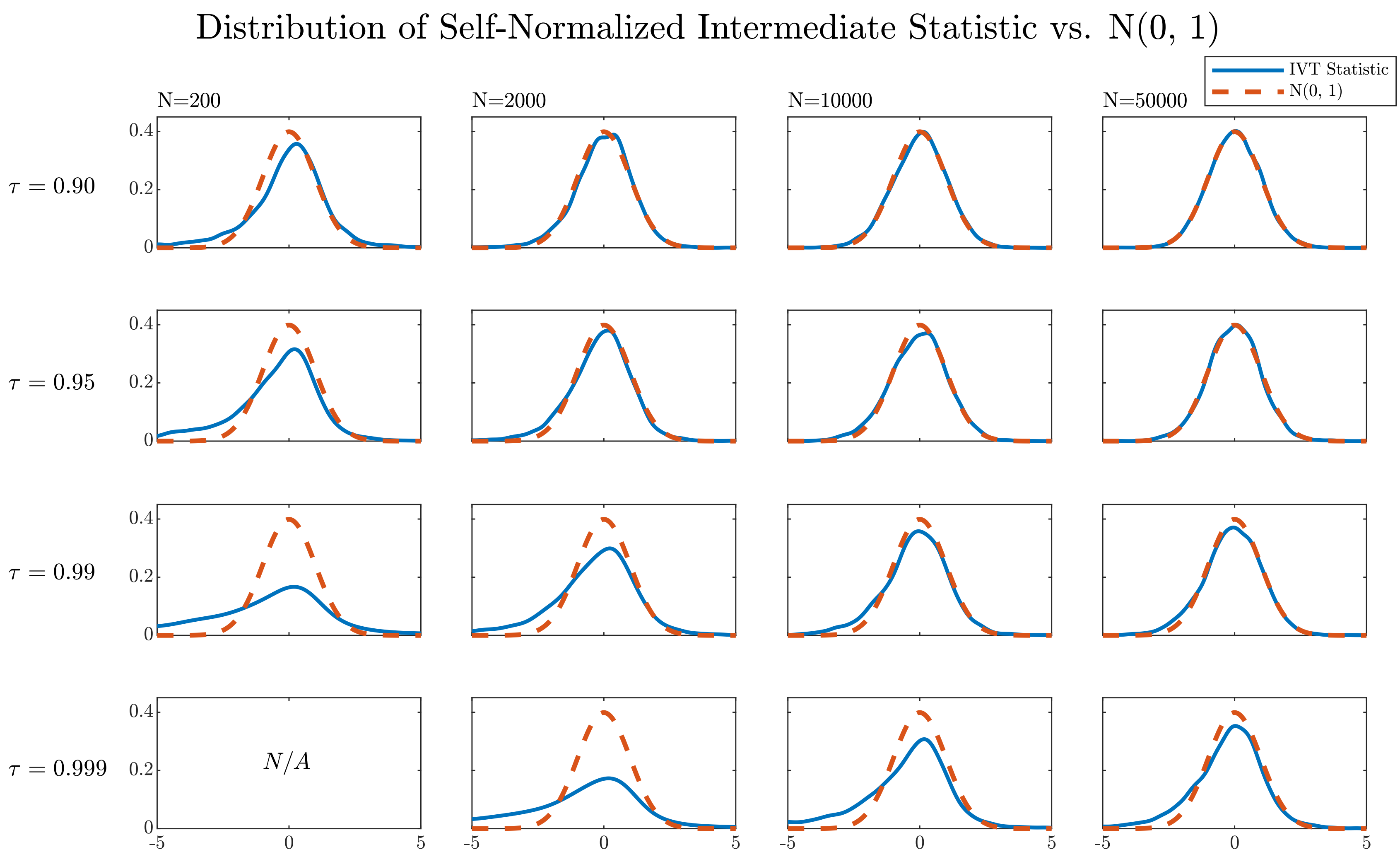

Figure OA.45: Distribution of the self-normalized intermediate order statistic of theorem 4.5 vs. the limit $N(0,1)$ distribution. $\theta \sim t_3$, noiseless data. Rows index target quantiles, column index cross-sectional sample sizes N.

OA.5 Supplemental Materials for the Empirical Application

This section presents additional estimation 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).

We group the expanded results in two groups. First, figures OA.46-OA.51 expand on figure 3 in the main text. Each figure presents confidence intervals estimated by one of the methods discussed in section 5. We refer to section 5 for the definition and the implementation details of the methods. As in the main text, we shade the area where the rule of thumb of section 4 suggests using extreme-order approximations. For comparability, all CIs are reported on the same y-axis. Second, figures OA.52-OA.57 expand on figure 4. As with the first group, each figure corresponds to a single inference method.

The results broadly match those reported in section 6. In no case are the confidence intervals for AMD and BMD disjoint, regardless of whether a mean-variance standardization is applied. This finding lends additional credence to the key assumption of Combes et al. (2012). We refer to section 6 for a full discussion.

We also remark that the confidence intervals follow the predictions of the simulation study in section 5 and OA.4. In particular, the length of the "textbook" extrapolation confidence interval grows dramatically as the target quantile approaches 0 or 1. As a result, these confidence intervals are longer by several orders of magnitude than the extreme order CIs for the very tail quantiles.

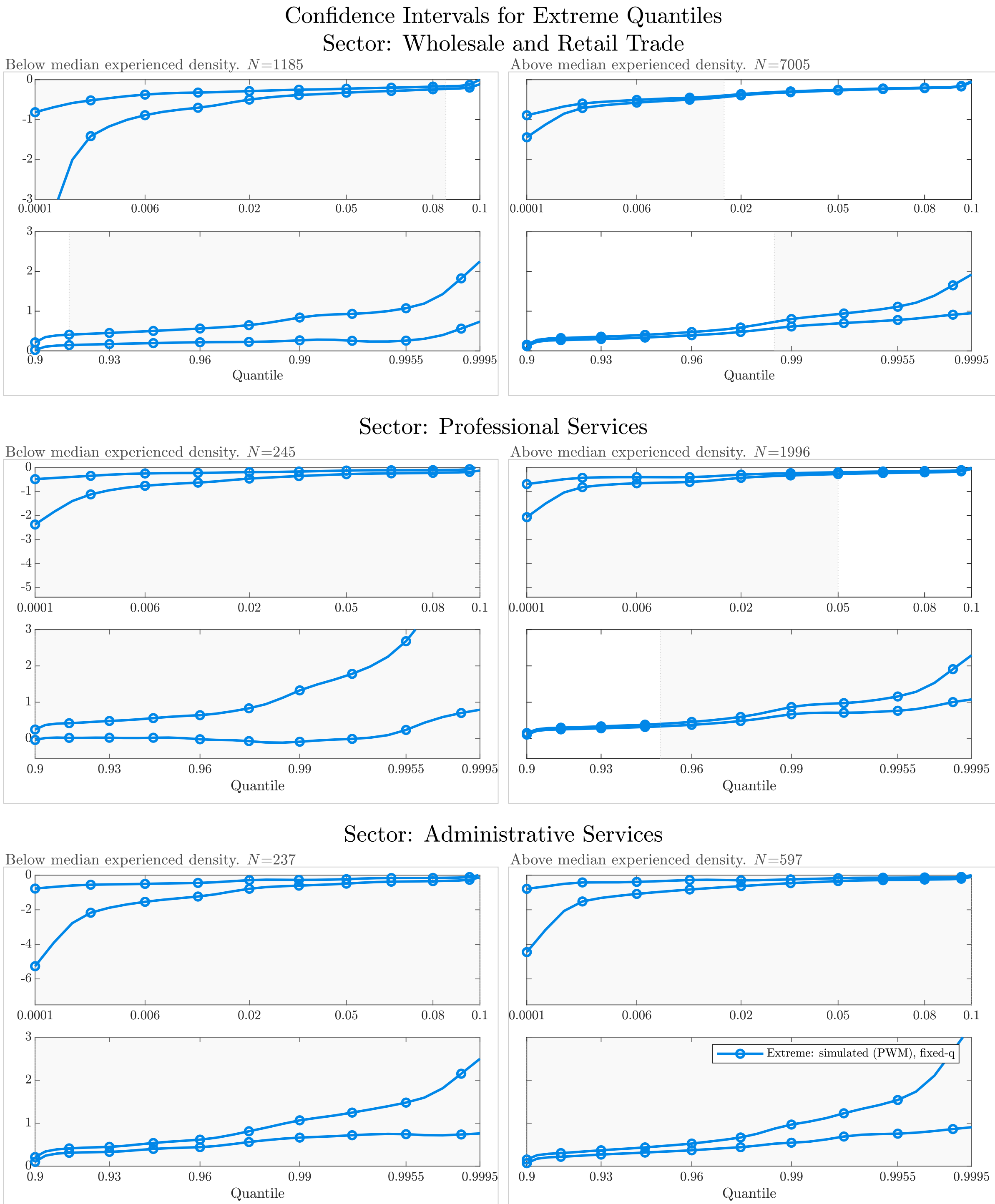

Figure OA.46: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

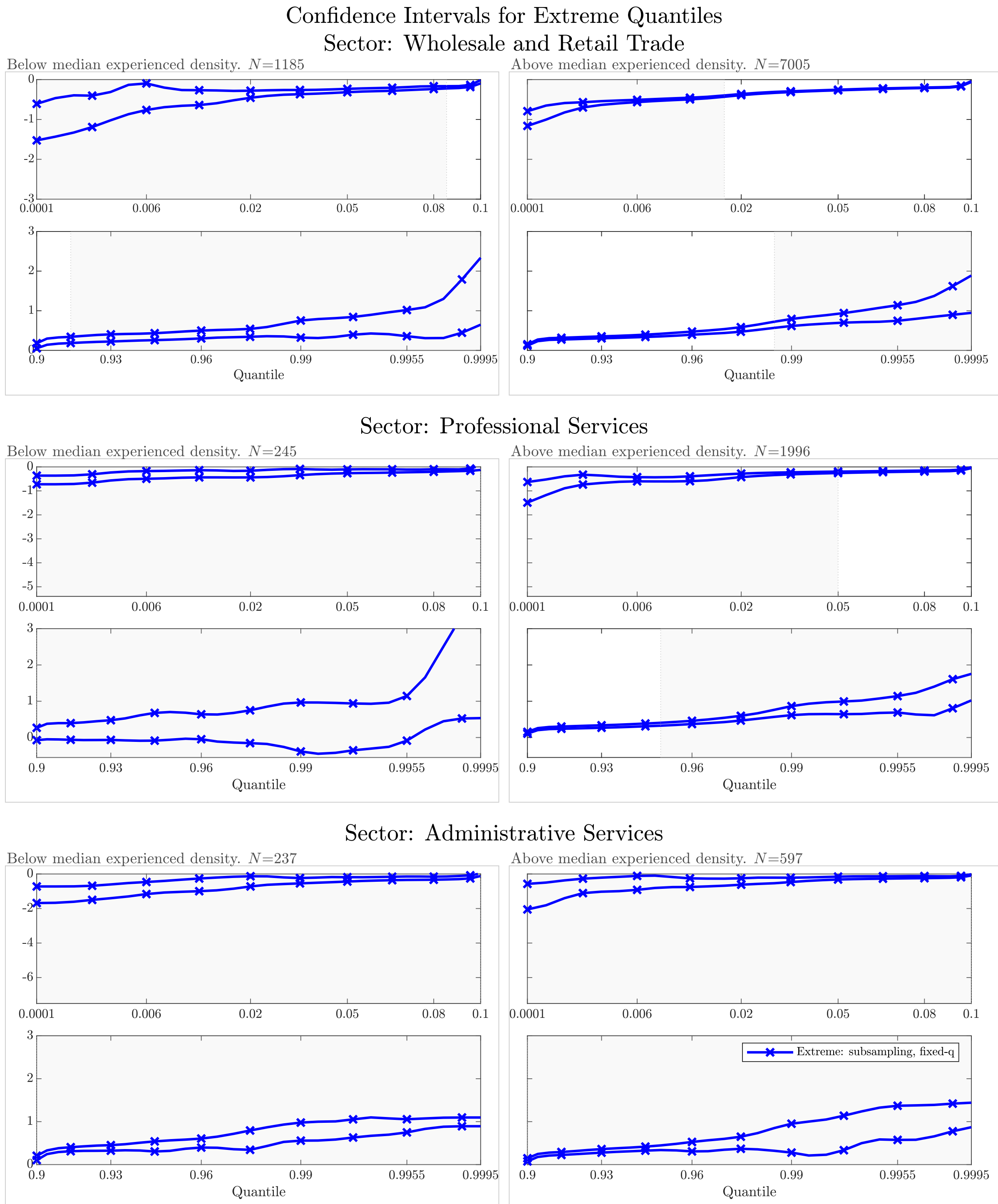

Figure OA.47: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

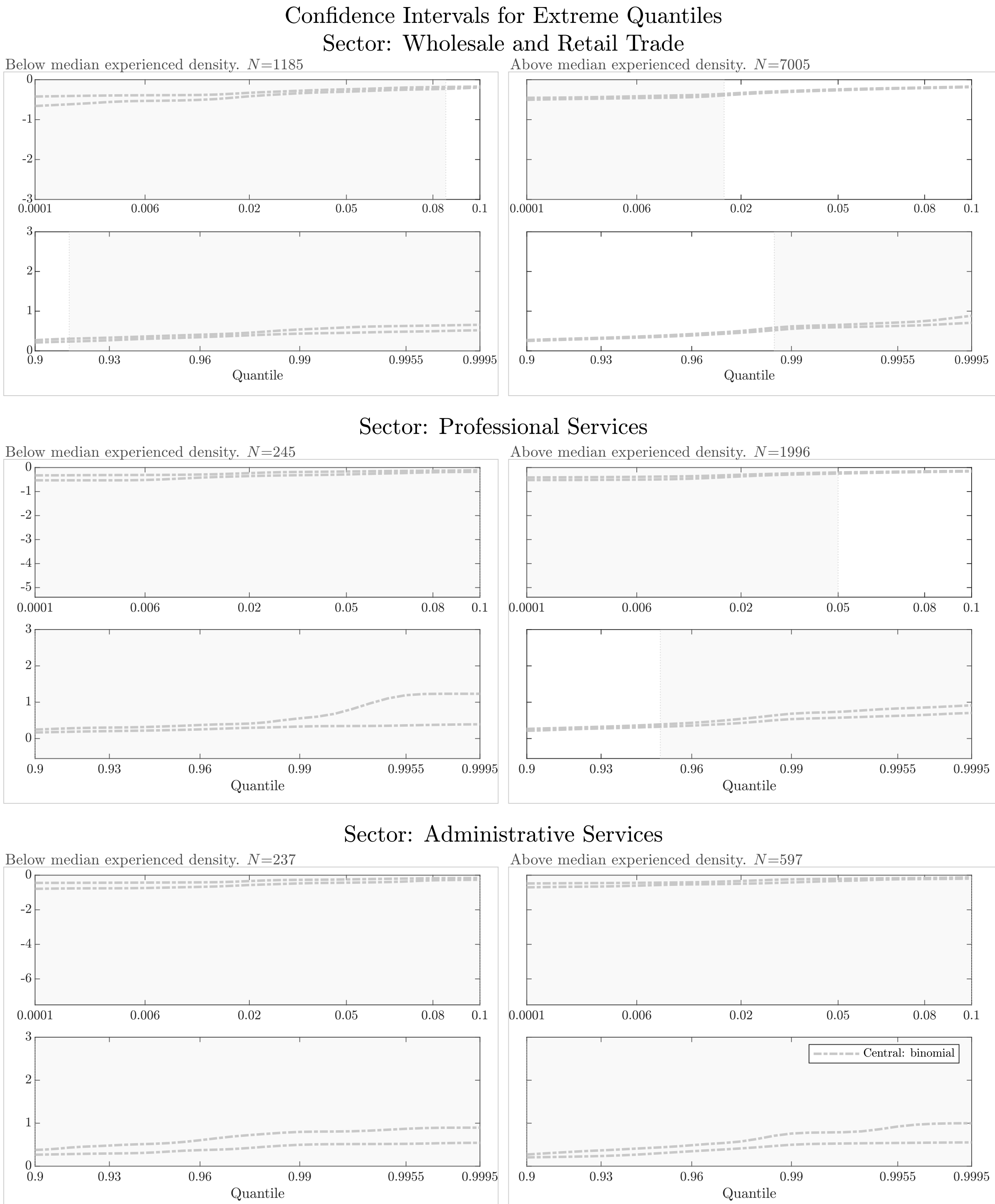

Figure OA.48: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

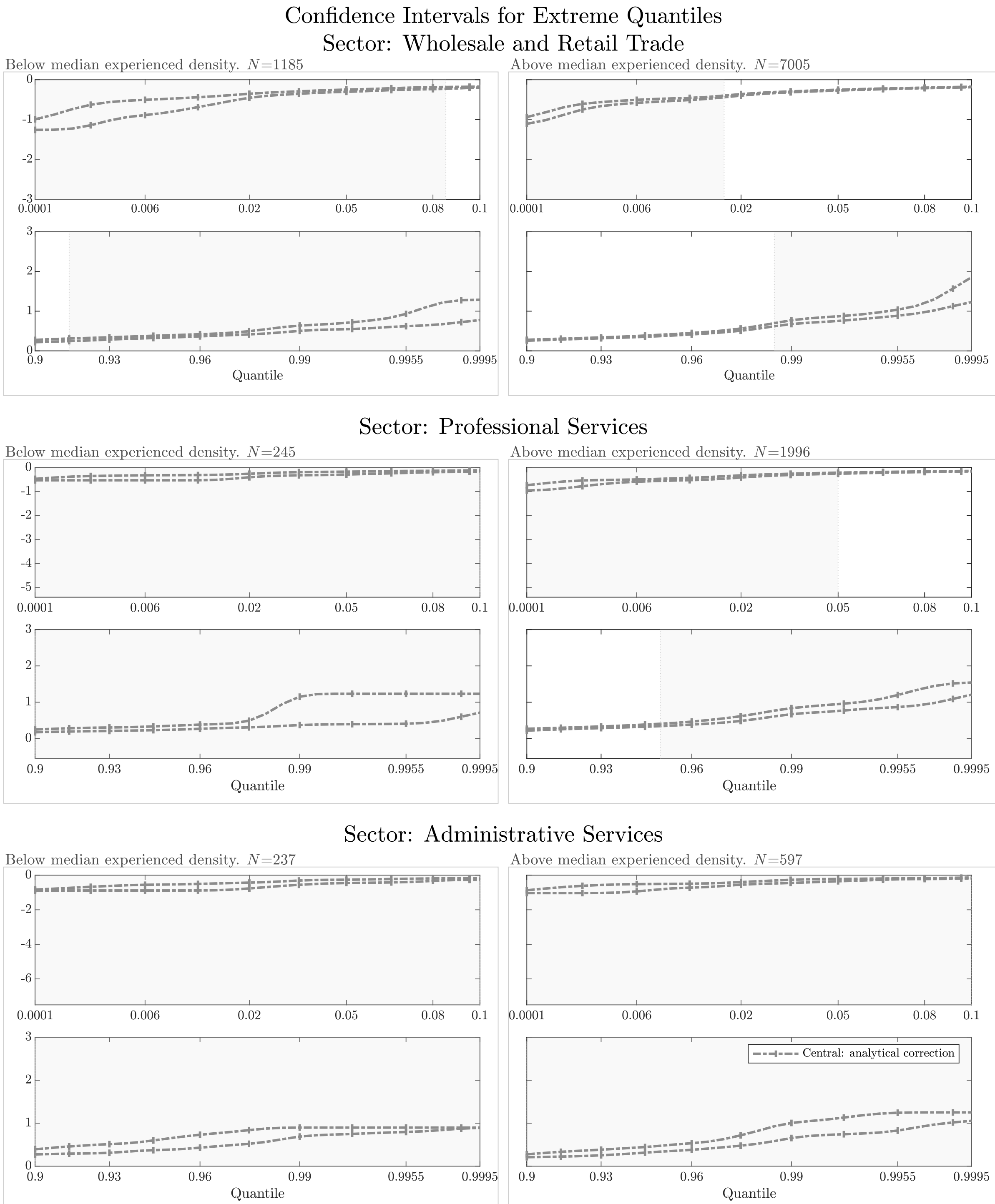

Figure OA.49: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

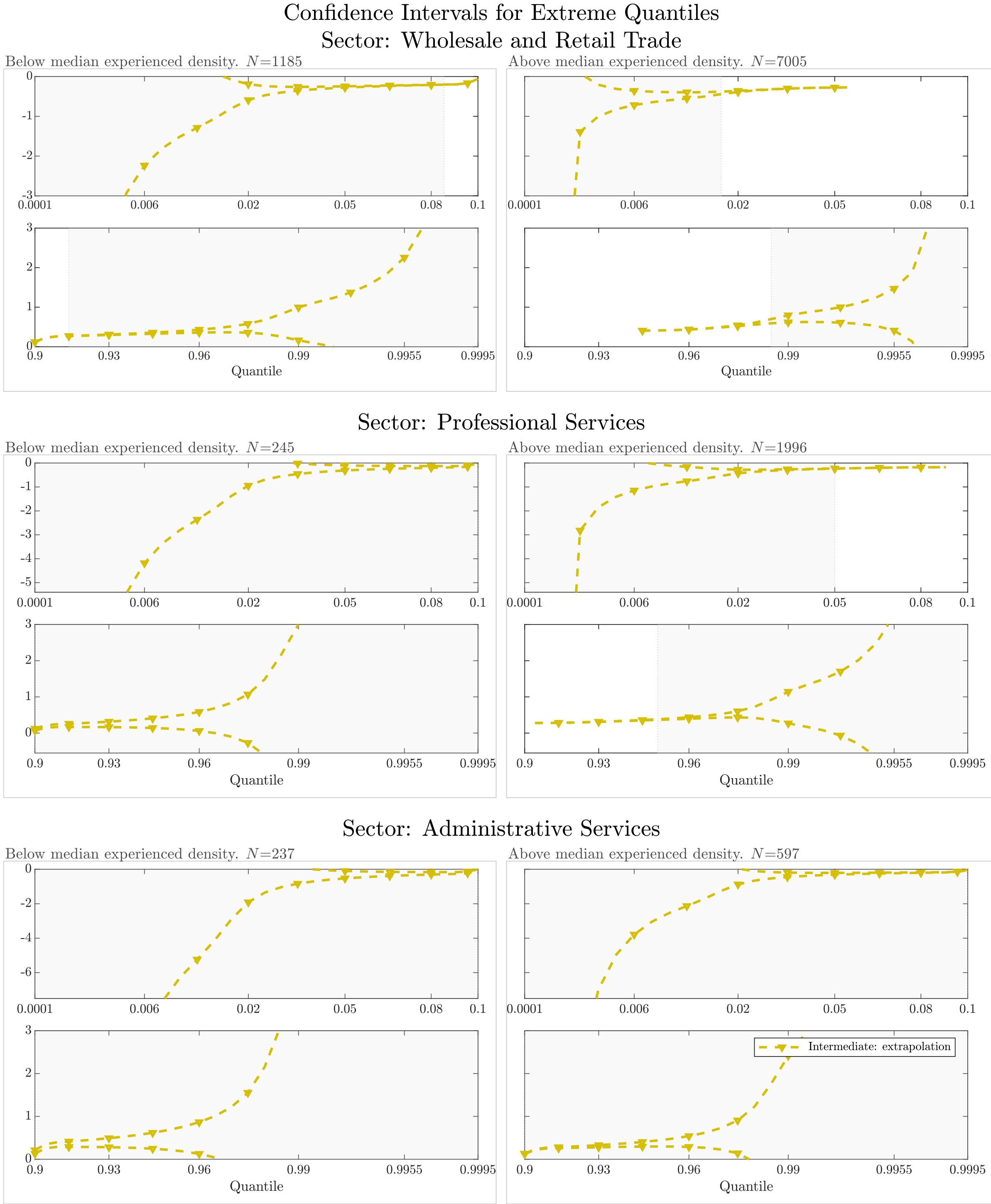

Figure OA.50: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

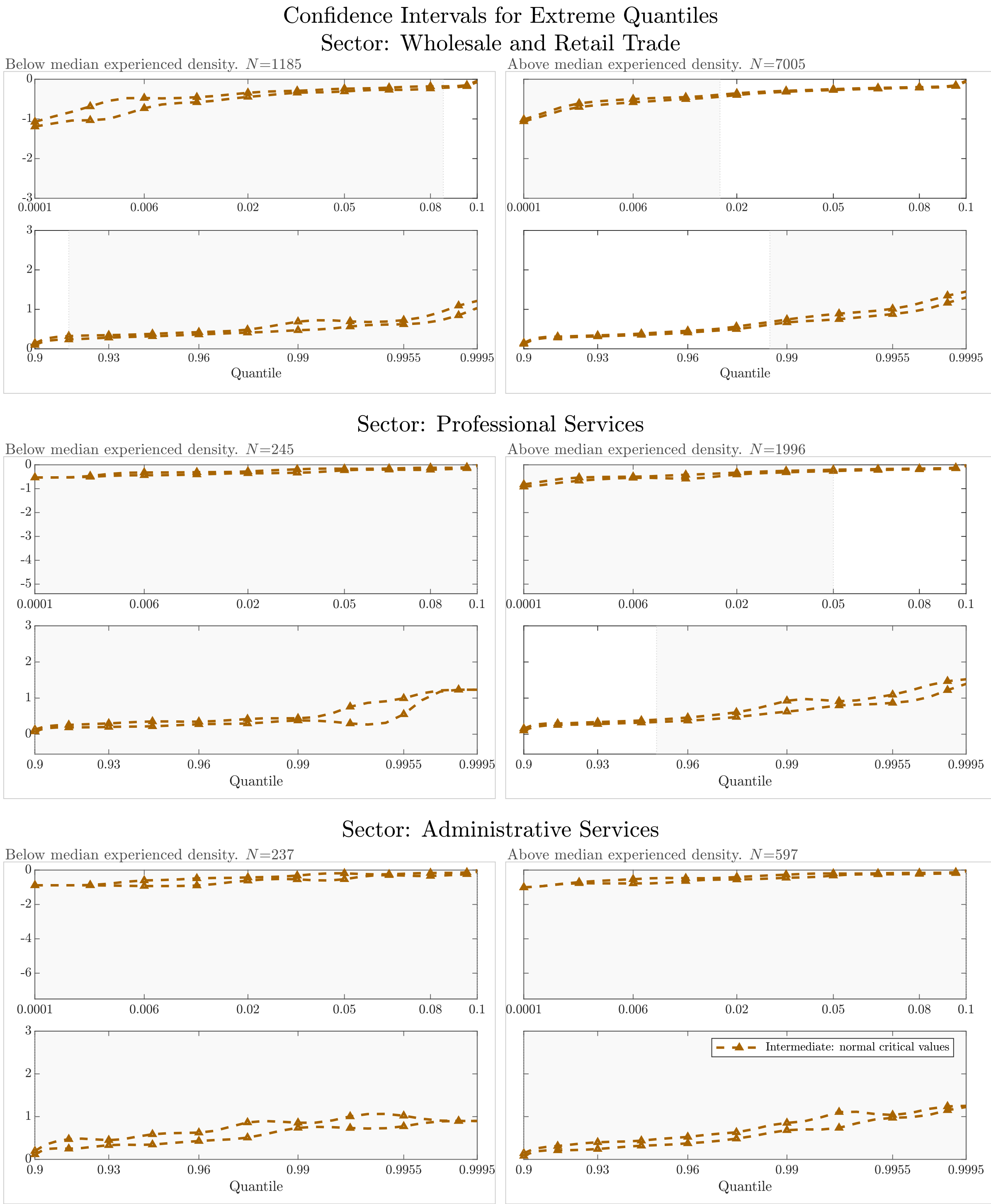

Figure OA.51: 95% confidence intervals for extreme quantiles of total factor productivity. Split by below and above median experience density (BME and AME, respectively); split by sectors. Shaded area: zone where the rule of thumb of section 4 suggests extreme-order approximations. For areas and sectors with $N \leq 1000$, all the depicted quantiles fall into this zone.

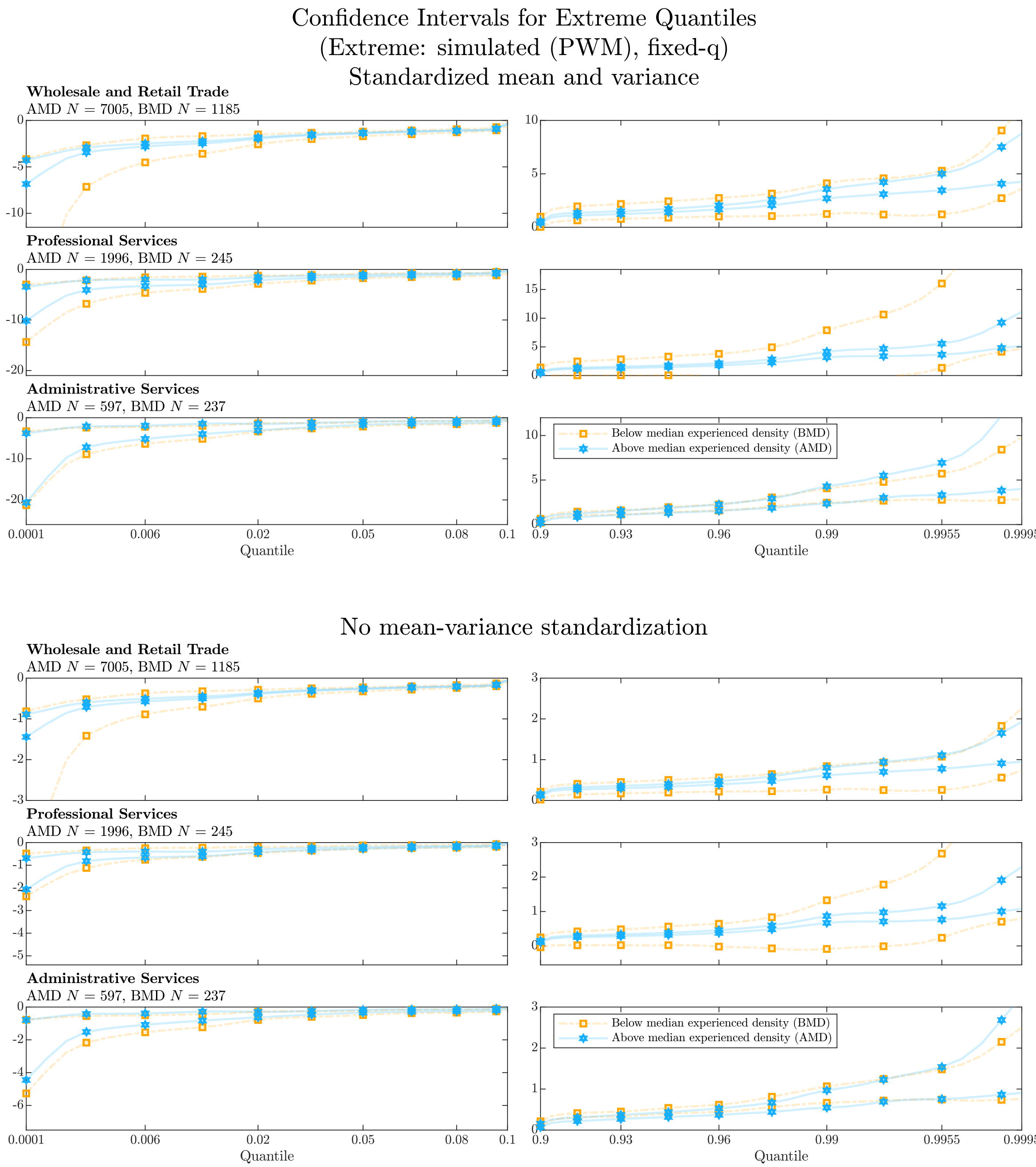

Figure OA.52: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

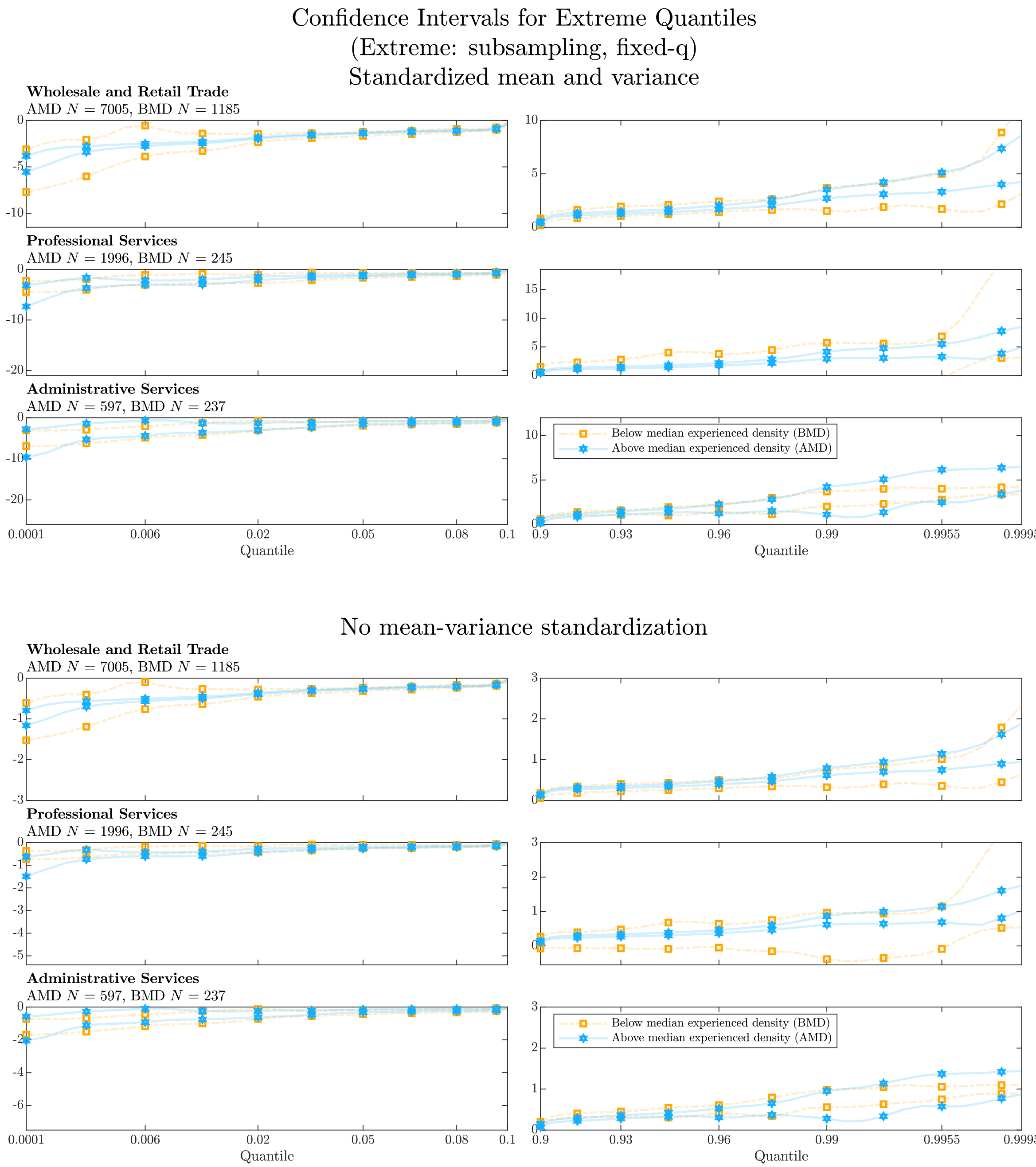

Figure OA.53: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

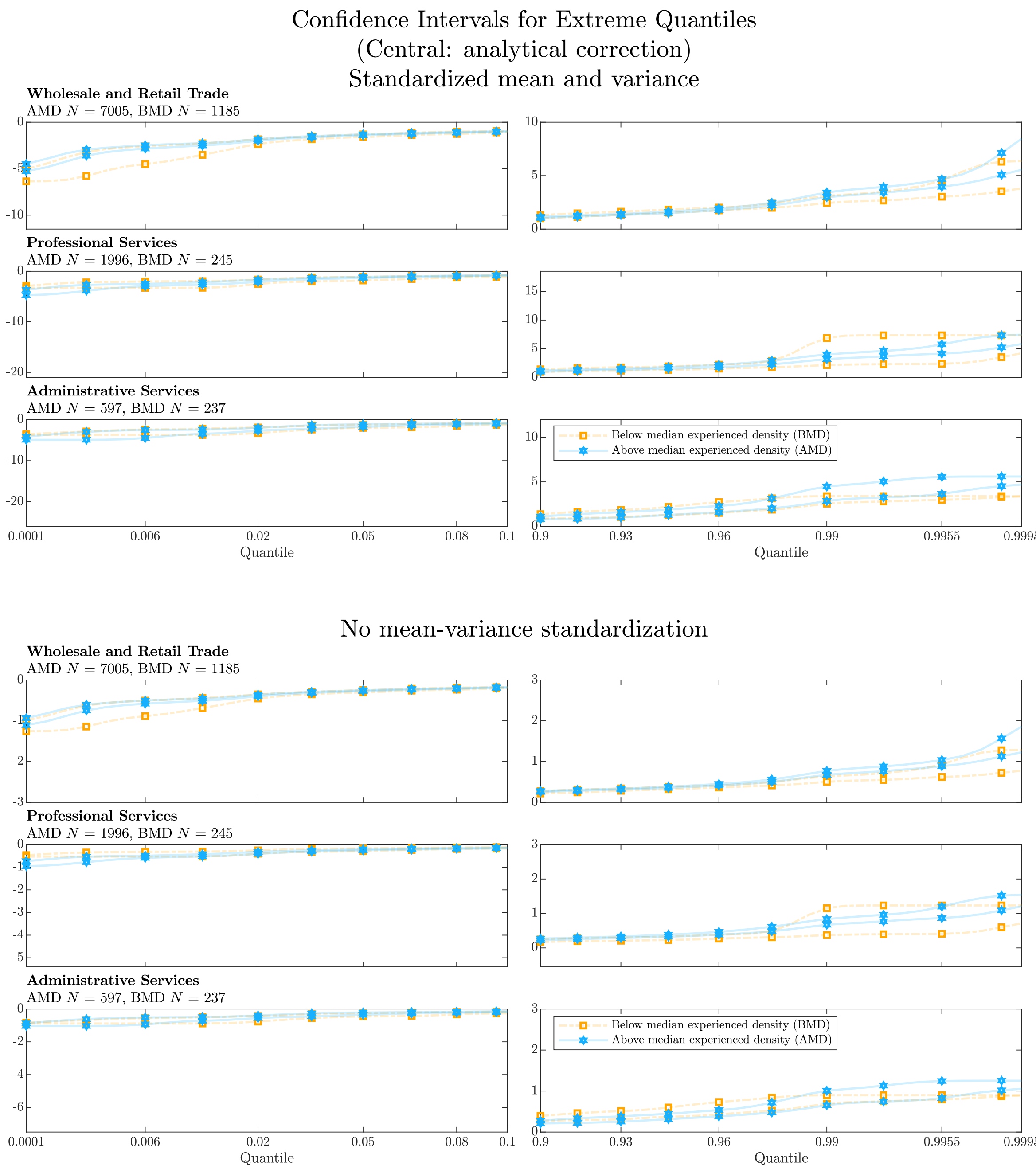

Figure OA.54: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

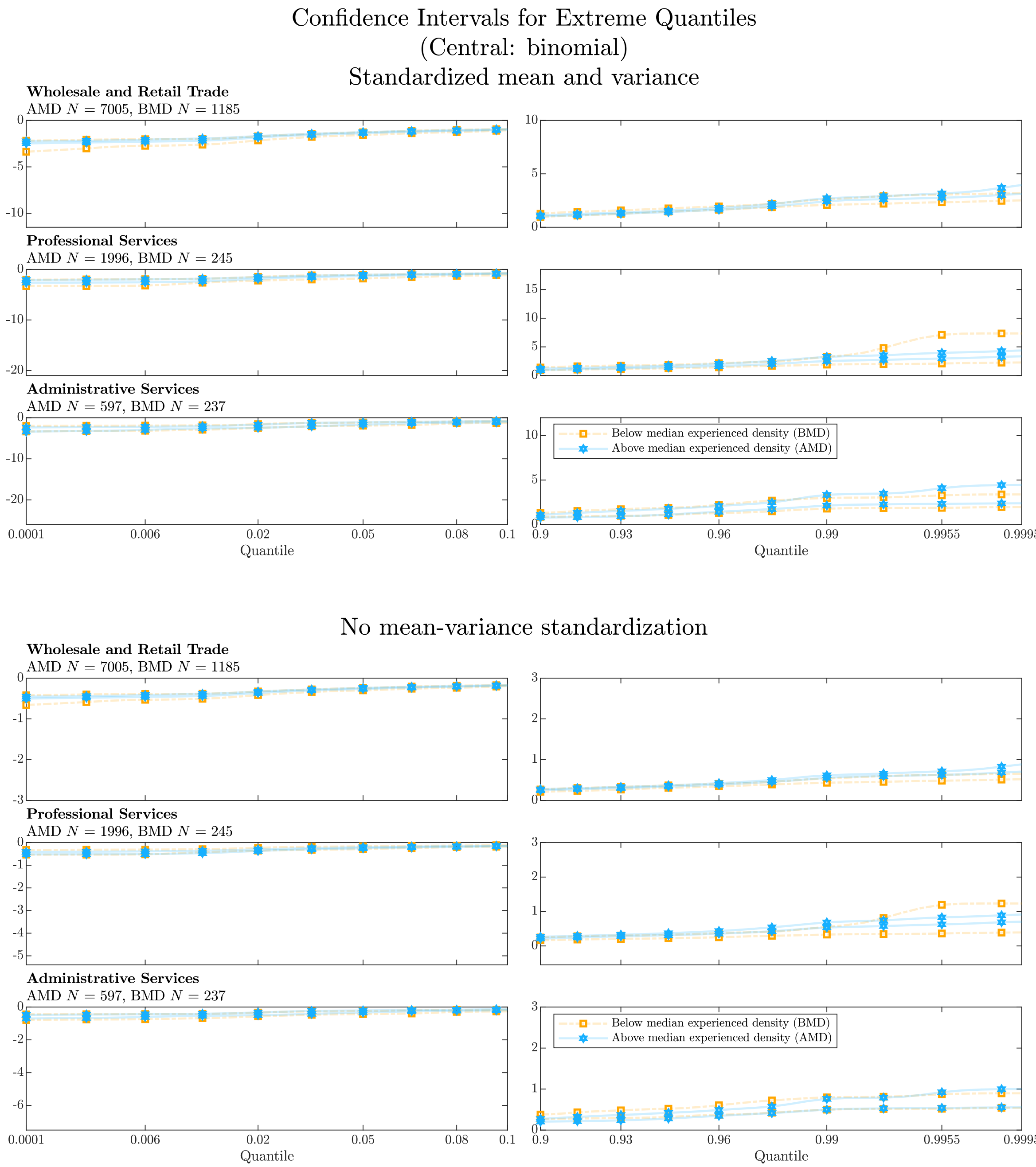

Figure OA.55: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

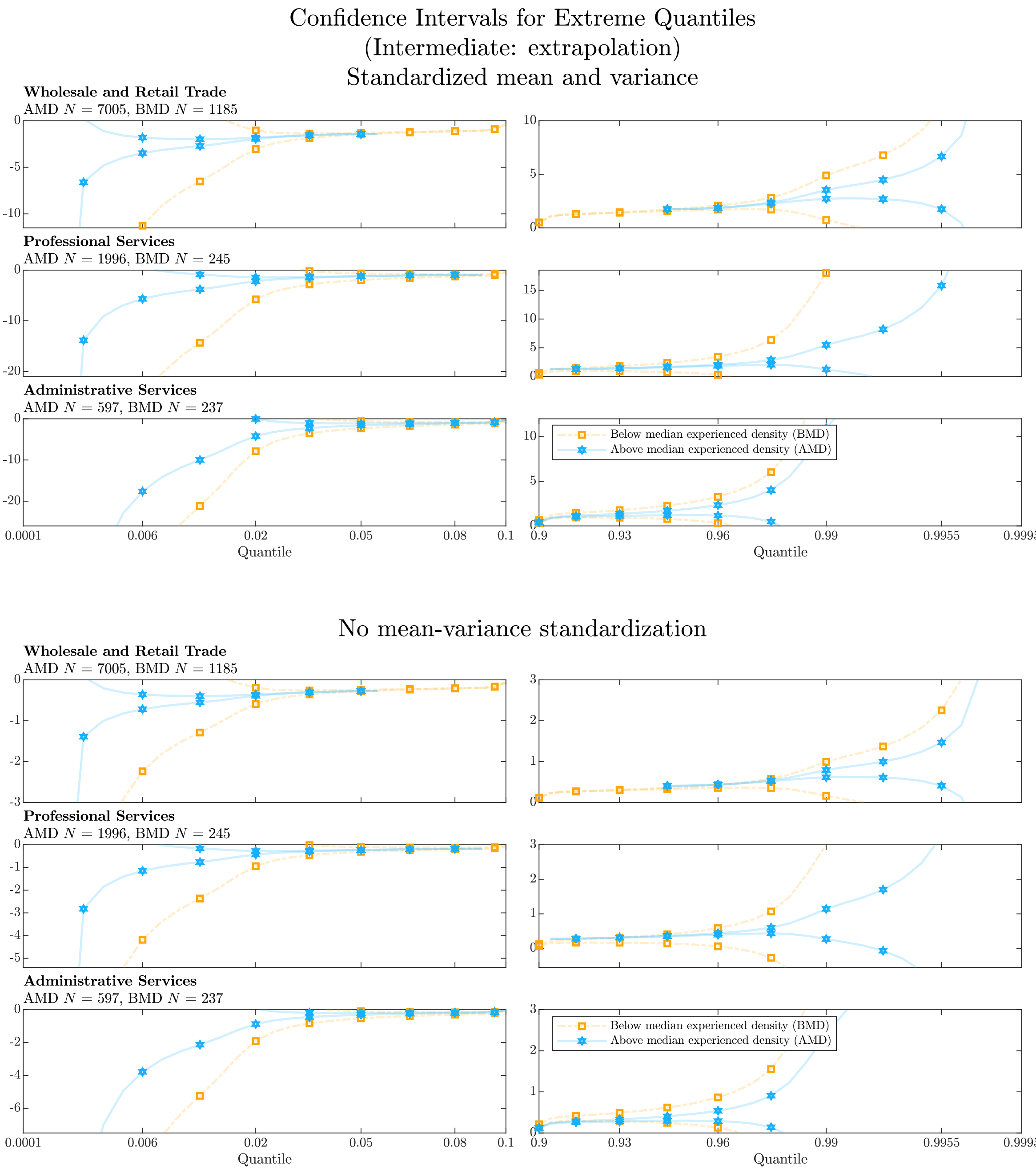

Figure OA.56: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

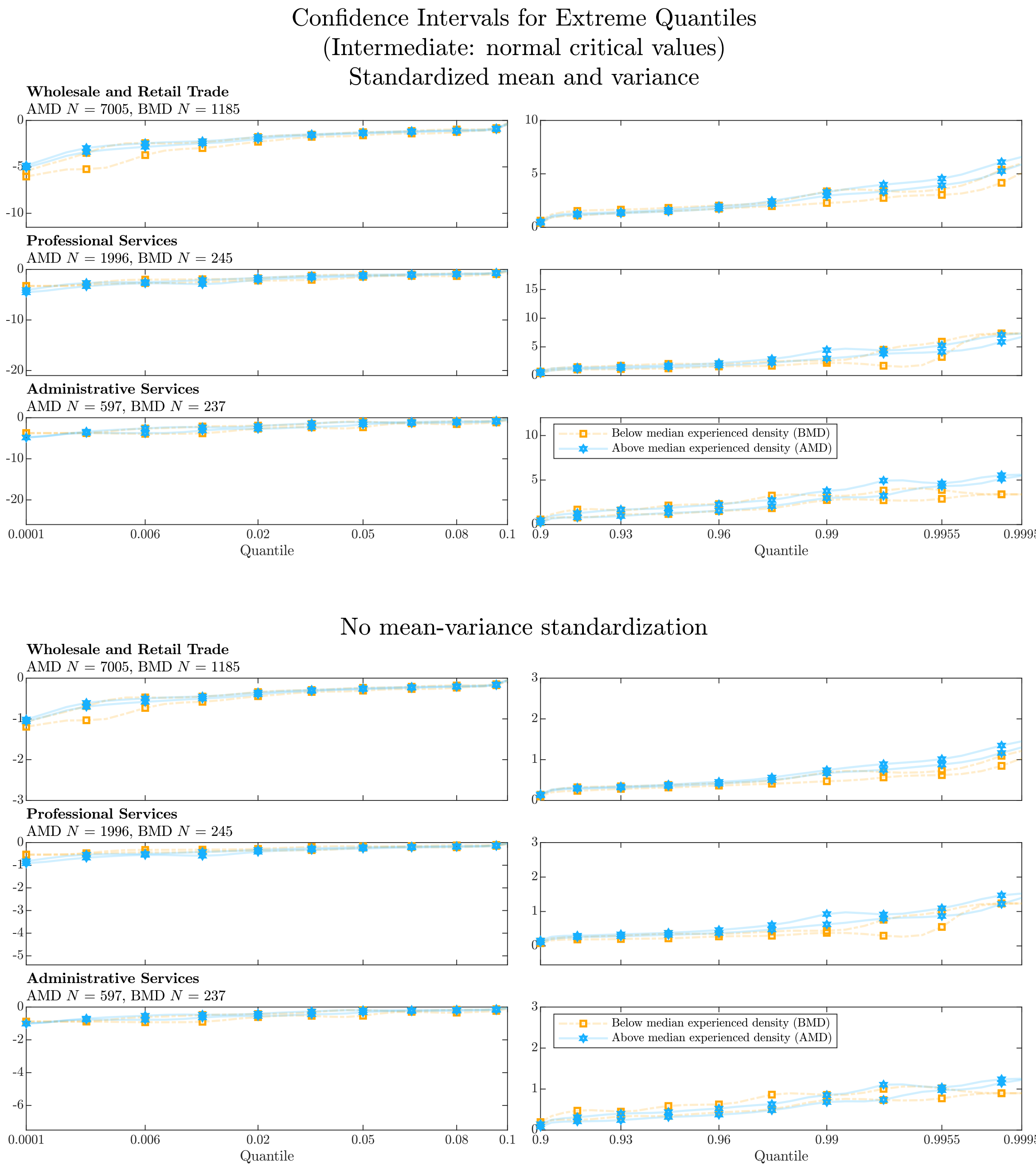

Figure OA.57: 95% confidence intervals for extreme quantiles of total factor productivity. Split by sector. Top panel: AMD and BMD data standardized to have the same mean and variance. Bottom panel: raw data.

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*, 40(1): 60–97, 2024. doi: 10.1017/S0266466622000378.

U. K. Müller and A. Norets. Credibility of Confidence Sets in Nonstandard Econometric Problems. *Econometrica*, 84(6):2183–2213, 2016. ISSN 0012-9682. doi: 10.3982/ecta14023.

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