SUPPLEMENTARY MATERIAL FOR "INFERENCE FOR EXTREMAL REGRESSION WITH DEPENDENT HEAVY-TAILED DATA"

BY ABDELAATI DAOUIA[1,a], GILLES STUPFLER[2,b] AND ANTOINE USSEGLIO-CARLEVE[3,c]

[1]*Toulouse School of Economics, University of Toulouse Capitole, France,* [a]*abdelaati.daouia@tse-fr.eu*

[2]*Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France,* [b]*gilles.stupfler@univ-angers.fr*

[3]*Avignon Université, LMA UPR 2151, 84000 Avignon, France,* [c]*antoine.usseglio-carleve@univ-avignon.fr*

This supplementary material document contains further details about our technical conditions and an expanded discussion of the rates of pointwise convergence of our estimators. We then provide the proofs of all theoretical results in the main paper and a full analysis of our worked-out regression examples, preceded by auxiliary results and their proofs. We finally provide further details about our bias and variance correction procedures, and extra finite-sample results.

Throughout we denote by $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$ the positive and negative parts of a real number x. For a function f on $\mathbb{R}^p$, $\nabla f(\boldsymbol{x})$, $Jf(\boldsymbol{x})$ and $Hf(\boldsymbol{x})$ stand respectively for its gradient vector, Jacobian matrix, and Hessian matrix at the point $\boldsymbol{x}$. For a function $f = f(\boldsymbol{x}, \boldsymbol{y})$ on $\mathbb{R}^p \times \mathbb{R}^q$, $\nabla_{\boldsymbol{x}} f$ and $H_{\boldsymbol{x}} f$ denote its partial gradient vector and Hessian matrix with respect to $\boldsymbol{x}$ (*i.e.* the first p components of its gradient vector and the submatrix made of the first p rows and columns of its Hessian matrix, respectively). The symbols $\mathbf{0}_p$ and $\mathbf{1}_p$ denote vectors in $\mathbb{R}^p$ with all components equal to 0 and 1, respectively. The symbol $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^p$ and we abuse notation to let it denote the corresponding matrix norm, *i.e.* the spectral norm.

APPENDIX A: MATHEMATICAL CONCEPTS AND PROOFS

A.1. Further details about mixing conditions. The $\alpha-$mixing (or strong mixing) assumption is conveniently expressed as follows: let, for any two positive integers $a \le b \le +\infty$, $\mathcal{F}_a^b = \sigma(\{(\boldsymbol{X}_j, Y_j), a \le j \le b\})$ be the $\sigma-$algebra generated by $\{(\boldsymbol{X}_j, Y_j), a \le j \le b\}$, and say that $((\boldsymbol{X}_t, Y_t))_{t\ge 1}$ is $\alpha-$mixing if and only if $\alpha(n) \to 0$ as $n \to \infty$, where

$$\alpha(n) = \sup_{k\ge 1} \sup_{A\in\mathcal{F}_1^k} \sup_{B\in\mathcal{F}_{k+n}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

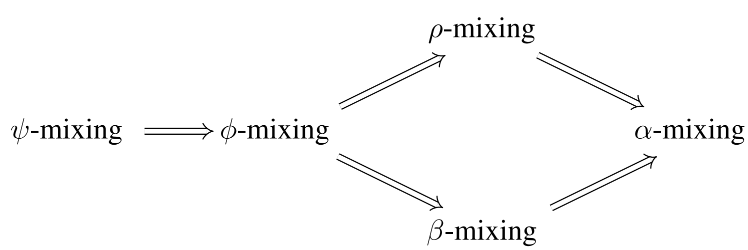
The $\alpha-$mixing assumption is satisfied in a large amount of classical models including nonlinear autoregressive processes [20, Section 2.4], nonlinear ARCH processes [see 37] and multivariate ARMA and GARCH models [see 8]. It has been widely used in standard regression, see for instance [3], [36], and more recently [10] in a high-dimensional setup. [41] develops a general theory of strong mixing stochastic processes and provides an extensive bibliography. The $\beta-$, $\rho-$, $\phi-$ and $\psi-$mixing coefficients of the series $((\boldsymbol{X}_t, Y_t))_{t\ge 1}$, meanwhile, are

MSC 2010 subject classifications: Primary 62G32; secondary 62G05, 62G08, 62G15, 62G20, 62G30.
Keywords and phrases: Conditional quantiles, conditional expectiles, extreme value analysis, heavy tails, inference, mixing, nonparametric regression.

respectively defined as

$$
\begin{aligned}
\beta(n) &= \sup_{k\geq 1} \mathbb{E}\left(\sup_{B\in\mathcal{F}_{k+n}^{\infty}} |\mathbb{P}(B|\mathcal{F}_1^k) - \mathbb{P}(B)| \right), \\
\rho(n) &= \sup_{k\geq 1} \sup_{U\in L^2(\mathcal{F}_1^k)} \sup_{V\in L^2(\mathcal{F}_{k+n}^{\infty})} |\operatorname{Corr}(U,V)|, \\
\phi(n) &= \sup_{k\geq 1} \sup_{\substack{A\in\mathcal{F}_1^k \\ \mathbb{P}(A)>0}} \sup_{B\in\mathcal{F}_{k+n}^{\infty}} |\mathbb{P}(B|A) - \mathbb{P}(B)| \\
\text{and } \psi(n) &= \sup_{k\geq 1} \sup_{\substack{A\in\mathcal{F}_1^k \\ \mathbb{P}(A)>0}} \sup_{\substack{B\in\mathcal{F}_{k+n}^{\infty} \\ \mathbb{P}(B)>0}} \left| \frac{\mathbb{P}(B|A)}{\mathbb{P}(B)} - 1 \right|.
\end{aligned}
$$

One then says that the stochastic process $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ is $\beta-$mixing (resp. $\rho-$mixing, $\phi-$mixing, $\psi-$mixing) if $\beta(n)\to 0$ (resp. $\rho(n)\to 0$, $\phi(n)\to 0$, $\psi(n)\to 0$) as $n\to\infty$. Nice surveys of mixing conditions are provided in [6] and [20]. The implications between these conditions can be represented as follows:



There is in general no converse implication to any of the above implications.

Particular examples of $\beta-$ and $\rho-$mixing processes are autoregressive processes whose innovations have absolutely continuous distributions satisfying certain regularity conditions. This fact will be used in Section 4.

A.2. Expanded discussion of technical conditions. We recall here conditions $\mathcal{K}$, $\mathcal{L}_g$, $\mathcal{L}_m$, $\mathcal{L}_\omega$, $\mathcal{B}_p$, $\mathcal{B}_m$, $\mathcal{B}_\Omega$, $\mathcal{H}_\delta$, $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$, $\mathcal{D}_\omega$, and we give further details about their rationale, interpretation, and position with respect to assumptions used in the literature.

Condition $\mathcal{K}$ The p.d.f. K is bounded with a support contained in the unit closed Euclidean ball.

This classical assumption in nonparametric estimation ensures in particular that only those observations $(\boldsymbol{X}_t, Y_t)$ such that $\boldsymbol{X}_t$ is close to $\boldsymbol{x}$ will be taken into account. It is satisfied by all standard compactly supported kernel functions, such as the uniform kernel over the unit ball, or (for $p=1$) the Epanechnikov, triangular and quartic kernels.

Condition $\mathcal{L}_g$ The p.d.f. g satisfies $g(\boldsymbol{x})>0$ and is Lipschitz continuous at $\boldsymbol{x}$: there exist $c,r>0$ such that for any $\boldsymbol{x}'\in B(\boldsymbol{x},r)$, $|g(\boldsymbol{x})-g(\boldsymbol{x}')|\leq c\,\|\boldsymbol{x}-\boldsymbol{x}'\|$.

Due to the local nature of nonparametric estimation, assumptions such as condition $\mathcal{L}_g$ on the local behavior of the marginal density of $\boldsymbol{X}$ are common in the regression literature, including in the extremal regression setup, see for example [14], [15], [22] and in the mixing framework, see for instance [43]. Their role is to guarantee that distributions at neighboring points $\boldsymbol{x}$ and $\boldsymbol{x}'$ are sufficiently close for nonparametric estimators to be asymptotically normal at standard rates of convergence.

Condition $\mathcal{B}_p$ There exists an integer $t_0 \geq 1$ such that

$$1 \leq t < t_0 \Rightarrow \lim_{r\to 0} r^{-p}\,\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x},r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x},r)) = 0$$
$$\text{and } \limsup_{r\to 0} \sup_{t\geq t_0} r^{-2p}\,\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x},r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x},r)) < \infty.$$

The purpose behind this condition is to ensure that similar values of the covariate cannot occur too often at neighboring time points. It is in particular satisfied (with $t_0 = 1$, meaning that the first half of the condition is empty and hence trivially holds) as soon as, for all $t \geq 1$, the random vector $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ has a joint p.d.f. g_t such that $\sup_{t\geq 1} g_t$ is bounded on $B(\boldsymbol{x}, r) \times B(\boldsymbol{x}, r)$ for some $r > 0$. The latter local boundedness condition has been considered in part of the nonparametric regression literature for strongly mixing data: see for instance [3], [11], [36] and [38]. An example of stochastic process that clearly does not satisfy this local boundedness condition, but does actually satisfy $\mathcal{B}_p$, is a causal and invertible AR(p) process (Y_t), for $p \geq 2$, with natural covariate $\boldsymbol{X}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\top \in \mathbb{R}^p$, see our list of examples in Section 4.

Assumptions $\mathcal{K}$, $\mathcal{L}_g$ and $\mathcal{B}_p$ are imposed in particular to control the asymptotic behavior of the Parzen-Rosenblatt estimator $\widehat{g}_n(\boldsymbol{x})$. In addition, our analysis of extreme conditional expectile estimators requires regularity assumptions about conditional moments.

Condition $\mathcal{H}_\delta$ One has $\gamma(\boldsymbol{x}) < 1/(2+\delta)$ and there exists $r > 0$ such that

$$\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \mathbb{E}(Y_-^{2+\delta} | \boldsymbol{X} = \boldsymbol{x}') < \infty.$$

The motivation for this condition is to guarantee a finite conditional moment of order $(2+\delta)$ in a neighborhood of $\boldsymbol{x}$; in the unconditional framework, Theorem 2 in [16] imposes the analogous condition $\mathbb{E}(Y_-^{2+\delta}) < \infty$. This assumption is a sensible requirement for conditional expectile estimation since the asymptotic normality of empirical smoothed conditional expectiles should intuitively require a bit more than a finite conditional variance for a Lyapunov-type central limit theorem to apply (recall that expectiles extend the mean as quantiles extend the median). A sufficient condition for $\mathcal{H}_\delta$ to hold in terms of the conditional p.d.f. of $\boldsymbol{X}$ given Y is established in Lemma A.1(i).

Condition $\mathcal{L}_m$ The response Y has a finite second moment given $\boldsymbol{X} = \boldsymbol{x}$, and the conditional mean functions $\mathbb{E}(Y|\boldsymbol{X} = \cdot)$ and $\mathbb{E}(Y^2|\boldsymbol{X} = \cdot)$ are Lipschitz continuous at $\boldsymbol{x}$: there exist $c, r > 0$ such that

$$\forall \boldsymbol{x}' \in B(\boldsymbol{x}, r),\ |\mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}) - \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}')| \leq c\,\|\boldsymbol{x} - \boldsymbol{x}'\|$$
$$\text{and } |\mathbb{E}(Y^2|\boldsymbol{X} = \boldsymbol{x}) - \mathbb{E}(Y^2|\boldsymbol{X} = \boldsymbol{x}')| \leq c\,\|\boldsymbol{x} - \boldsymbol{x}'\|.$$

Condition $\mathcal{B}_m$ There exists $r > 0$ such that

$$\sup_{t\geq 1} \sup_{\boldsymbol{x}_1,\boldsymbol{x}_{t+1}\in B(\boldsymbol{x},r)} \mathbb{E}(Y_1^2 + Y_{t+1}^2 | \boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_{t+1} = \boldsymbol{x}_{t+1}) < \infty.$$

Conditions $\mathcal{L}_m$ and $\mathcal{B}_m$ ensure that the nonparametric estimator of the regression mean will converge at a standard rate. An alternative option to condition $\mathcal{B}_m$ consists in assuming a finite unconditional second moment of Y_1 and putting all the requirements on the conditional joint p.d.f. of $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ given (Y_1, Y_{t+1}): this approach is followed by [37] and [48]. We provide a link between these two viewpoints in our Lemma A.1(ii).

Finally, to control the variation in conditional extreme value behavior and the dependence between conditional extremes at different points in time and the covariate space, we introduce, for any $z > 1$, the quantities

$$\omega_h(z|\boldsymbol{x}) = \sup_{y \geq z} \sup_{\boldsymbol{x}' \in B(\boldsymbol{x},h)} \frac{1}{\log(y)} \left| \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} \right|$$

$$\text{and } \Omega_h(z|\boldsymbol{x}) = \sup_{t \geq 1} \sup_{y,y' \geq z} \sup_{\boldsymbol{x}',\boldsymbol{x}'' \in B(\boldsymbol{x},h)} \frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y' | \boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}}.$$

The first supremum $\omega_h(z|\boldsymbol{x})$ quantifies the gap between marginal conditional extremes in a neighborhood of $\boldsymbol{x}$. This quantity has already been introduced and used in the literature on conditional extremes, see *e.g.* [25, 26], [27] and [44, 45]. Intuitively, if the focus is on the conditional extremes of Y, say $Y \geq z = y_n \to \infty$, based on observations $\boldsymbol{X}_t \in B(\boldsymbol{x}, h_n)$, then $\omega_{h_n}(y_n|\boldsymbol{x})$ should be small to make consistent estimation of conditional extremes of Y given $\boldsymbol{X} = \boldsymbol{x}$ possible. This in fact true under a formal, stronger Lipschitz-type assumption in the spirit of condition $\mathcal{L}_g$ that will be satisfied in all our worked-out examples in Section 4.

Condition $\mathcal{L}_\omega$ There exists $r > 0$ such that

$$\limsup_{y \to \infty} \sup_{\substack{\boldsymbol{x}' \in B(\boldsymbol{x},r) \\ \boldsymbol{x}' \neq \boldsymbol{x}}} \frac{1}{\|\boldsymbol{x}' - \boldsymbol{x}\|} \left| \frac{1}{\log(y)} \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} \right| < \infty.$$

Condition $\mathcal{L}_\omega$ is a Lipschitz assumption on the log-tail probability $\log \overline{F}(y|\cdot)$, in an appropriate, uniform sense in y large enough. Under this assumption, it is immediate that $\omega_{h_n}(y_n|\boldsymbol{x}) = O(h_n) \to 0$ for any $y_n \to \infty$ and $h_n \to 0$.

The second quantity $\Omega_h(z|\boldsymbol{x})$ evaluates the degree of clustering in the joint conditional extremes of (Y_1, Y_{t+1}). It is indeed instructive to note that when $(\boldsymbol{X}_t, Y_t) = (\boldsymbol{X}, Y)$ for every t, then

$$\frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y' | \boldsymbol{X}_1 = \boldsymbol{x}, \boldsymbol{X}_{t+1} = \boldsymbol{x})}{\sqrt{\overline{F}(y|\boldsymbol{x})\overline{F}(y'|\boldsymbol{x})}} = \frac{\overline{F}(\max(y,y')|\boldsymbol{x})}{\sqrt{\overline{F}(y|\boldsymbol{x})\overline{F}(y'|\boldsymbol{x})}} \leq 1$$

with equality when $y = y'$, via the Cauchy-Schwarz inequality $\mathbb{P}(A \cap B) \leq \sqrt{\mathbb{P}(A)\mathbb{P}(B)}$. At the opposite, when $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is i.i.d., then the left-hand side above clearly converges to 0 as $y, y' \to \infty$. The denominator in $\Omega_h(z|\boldsymbol{x})$ thereby quantifies a kind of “worst-case scenario” for conditional extreme dependence across time, and it is reasonable to assume that $\Omega_h(z|\boldsymbol{x})$ is bounded when $h \to 0$ and $z \to \infty$, which corresponds to the assumption that a joint conditional extreme value of (Y_1, Y_{t+1}) is not much more likely than a marginal conditional extreme of Y_1, uniformly across time and locally uniformly across the covariate space. This is formalized as

Condition $\mathcal{B}_\Omega$ There exist $h, z > 0$ such that $\Omega_h(z|\boldsymbol{x}) < \infty$.

This condition should be considered as a weak assumption compared with the existence of a conditional tail copula as assumed in *e.g.* [18] and [21] in the unconditional setting. Lemma A.1(iii) provides a general conditional independence framework for $\mathcal{B}_\Omega$ to hold. The problem of extreme conditional expectile estimation requires an extra assumption about conditional tail heaviness (in which δ is a positive number).

The aforementioned conditions will ensure the pointwise asymptotic normality of our estimators at rates of convergence that have hitherto been standard in the conditional extreme value framework. Achieving optimal rates of convergence requires, similarly to classical nonparametric estimation, stronger regularity conditions: when estimating, for instance, the p.d.f. g in $\mathbb{R}^p$ with $\widehat{g}_n(\boldsymbol{x})$ using a symmetric p.d.f. K as kernel, it is well-known that the optimal rate of convergence $n^{-2/(p+4)}$ is obtained by solving the bias-variance tradeoff if g is twice differentiable at $\boldsymbol{x}$. This motivates the following additional assumptions.

Condition $\mathcal{KS}$ The p.d.f. K is bounded and symmetric (*i.e.* $K(\boldsymbol{u}) = K(-\boldsymbol{u})$) with a support contained in the unit closed Euclidean ball.

Condition $\mathcal{D}_g$ The p.d.f. g satisfies $g(\boldsymbol{x}) > 0$, is continuously differentiable in a neighborhood of $\boldsymbol{x}$ and its gradient is Lipschitz continuous at $\boldsymbol{x}$.

Condition $\mathcal{D}_m$ The response Y has a finite second moment given $\boldsymbol{X} = \boldsymbol{x}$, and the conditional mean functions $\mathbb{E}(Y|\boldsymbol{X} = \cdot)$ and $\mathbb{E}(Y^2|\boldsymbol{X} = \cdot)$ are continuously differentiable in a neighborhood of $\boldsymbol{x}$ and have Lipschitz continuous gradients at $\boldsymbol{x}$.

Condition $\mathcal{D}_\omega$ For y large enough, the function $\overline{F}(y|\cdot)$ is differentiable at $\boldsymbol{x}$, the function $y \mapsto \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y)$ has a limit $\boldsymbol{\mu}(\boldsymbol{x}) \in \mathbb{R}^p$ as $y \to \infty$, and there exists $r > 0$ with

$$\limsup_{y\to\infty} \sup_{\substack{\boldsymbol{x}'\in B(\boldsymbol{x},r)\\ \boldsymbol{x}'\neq\boldsymbol{x}}} \frac{1}{\|\boldsymbol{x}'-\boldsymbol{x}\|^2} \left| \frac{1}{\log(y)} \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} - (\boldsymbol{x}'-\boldsymbol{x})^\top \frac{\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})}{\log(y)} \right| < \infty.$$

Conditions $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$ and $\mathcal{D}_\omega$ are stronger versions of conditions $\mathcal{K}$, $\mathcal{L}_g$, $\mathcal{L}_m$ and $\mathcal{L}_\omega$, respectively. Condition $\mathcal{KS}$ is satisfied by any function K of the form $K(\boldsymbol{u}) = p^{-p/2}\prod_{j=1}^p \kappa(p^{-1/2}u_j)$ (a product of independent p.d.f.s on $\mathbb{R}$) or, when $p \geq 2$, the isotropic kernel $K(\boldsymbol{u}) = \kappa(\|\boldsymbol{u}\|)/(s_p\int_0^1 r^{p-1}\kappa(r)\,dr)$, where s_p is the surface of the unit hypersphere in $\mathbb{R}^p$, if κ is any bounded symmetric kernel on $\mathbb{R}$ with support $[-1,1]$, such as the uniform, Epanechnikov, triangular or quartic kernel. In condition $\mathcal{D}_\omega$, the assumption that $\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y)$ converges as $y \to \infty$ is motivated by the fact that, in the setup of conditional heavy tails,

$$\frac{\log \overline{F}(y|\boldsymbol{x})}{\log(y)} = -\frac{1}{\gamma(\boldsymbol{x})} + \frac{\log L(y|\boldsymbol{x})}{\log(y)}$$

where $L(\cdot|\boldsymbol{x})$ is a slowly varying function. In particular, $\log L(y|\boldsymbol{x})/\log(y) \to 0$ as $y \to \infty$, see Proposition 1.3.6(i) on p.16 of [2]. The assumption translates into supposing that this convergence also holds when taking the gradient with respect to $\boldsymbol{x}$, *i.e.* the function $L(\cdot|\boldsymbol{x})$ does not vary too wildly in $\boldsymbol{x}$ when y is large. Condition $\mathcal{D}_\omega$ will hold as soon as the heavy tails assumption is satisfied in a neighborhood V of $\boldsymbol{x}$, with in addition γ twice continuously differentiable on V and the existence of $y_0 \geq 0$ such that $\log L(y|\cdot)/\log(y)$ has a uniformly bounded Hessian matrix on $[y_0,\infty) \times V$, see Lemma A.4. The finite limit of $\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y)$ as $y \to \infty$ will then be $\boldsymbol{\mu}(\boldsymbol{x}) = \nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x}) \in \mathbb{R}^p$. In summary, while condition $\mathcal{L}_\omega$ will readily be checked by showing that the tail conditional probability is continuously differentiable with respect to the covariate value in an appropriate sense, condition $\mathcal{D}_\omega$ essentially asks for it to be twice continuously differentiable in the same way.

A.3. On rates of pointwise convergence. Theorems 2.1, 3.1 and 3.2, and therefore Theorems 2.3 and 3.4 (and also Theorems 2.2 and 3.3 if $\widehat{\gamma} = \widehat{\gamma}^{(J)}$ and $\widehat{\gamma}^E$, respectively), hold under weaker bias assumptions than the corresponding results of [15] and [27] in the i.i.d. case, although this naturally comes at the cost of the reinforced regularity conditions $\mathcal{D}_g$ and $\mathcal{D}_\omega$ (and $\mathcal{D}_m$ for expectile estimation). An important implication is that the rates of convergence of the estimators $\widehat{q}_n(\tau_n|\boldsymbol{x})$, $\widehat{e}_n(\tau_n|\boldsymbol{x})$ and $\breve{e}_n(\tau_n|\boldsymbol{x})$ are, under reasonably general assumptions, faster than what had been reported so far in the conditional extreme value literature. To illustrate this, suppose that $h_n = C_1 n^{-h}$ and $\tau_n = 1 - C_2 n^{-\tau}$, with $C_1, C_2 > 0$ and $h, \tau > 0$, and that the $\alpha-$mixing coefficients of the data $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ decay geometrically fast, so that possible choices of l_n and r_n are $l_n = \lfloor C \log n \rfloor$ and $r_n = \lfloor \log^2(n) \rfloor$ for $C > 0$ large enough (where $\lfloor\cdot\rfloor$ denotes the floor function), and then the convergence $r_n(r_n/\sqrt{nh_n^p(1-\tau_n)})^\delta \to 0$ automatically holds as soon as the numbers $h, \tau > 0$ are such that $nh_n^p(1-\tau_n) \to \infty$, *i.e.* $1 - ph - \tau > 0$. In the setting where $A(t|\boldsymbol{x}) \propto t^{\rho(\boldsymbol{x})}$, satisfied in a wide range of heavy-tailed models used in extreme value practice [see *e.g.* Table 2.1 on p.59 of 1], it is then straightforward to see that, under the conditions $nh_n^{p+2}(1-\tau_n)\log^2(1-\tau_n) \to 0$ and $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(1)$, the rate of convergence $1/\sqrt{nh_n^p(1-\tau_n)}$ is optimal when the exponents h and τ solve the maximization problem

$$\max\{1-ph-\tau\} \text{ s.t. } 1-(p+2)h-\tau \le 0,\ 1-ph-(1-2\rho(\boldsymbol{x}))\tau \le 0.$$

The solution is $h = -\rho(\boldsymbol{x})/(1-(p+2)\rho(\boldsymbol{x}))$ and $\tau = 1/(1-(p+2)\rho(\boldsymbol{x}))$, with corresponding convergence rate $1/\sqrt{nh_n^p(1-\tau_n)} = n^{\rho(\boldsymbol{x})/(1-(p+2)\rho(\boldsymbol{x}))}$. By contrast, when condition $nh_n^{p+2}(1-\tau_n)\log^2(1-\tau_n) \to 0$ is replaced by $\sqrt{nh_n^p(1-\tau_n)} \times h_n^2\log^2(1-\tau_n) \to \Delta \in [0,\infty)$, the optimal choices of h and τ should solve

$$\max\{1-ph-\tau\} \text{ s.t. } 1-(p+4)h-\tau \le 0,\ 1-ph-(1-2\rho(\boldsymbol{x}))\tau \le 0.$$

These optimal choices become $h = -\rho(\boldsymbol{x})/(2-(p+4)\rho(\boldsymbol{x}))$ and $\tau = 2/(2-(p+4)\rho(\boldsymbol{x}))$, yielding an optimal rate of convergence $1/\sqrt{nh_n^p(1-\tau_n)} = n^{2\rho(\boldsymbol{x})/(2-(p+4)\rho(\boldsymbol{x}))}$. In this setting, it is interesting to note that $p=0$ yields the optimal convergence rate $n^{\rho(\boldsymbol{x})/(1-2\rho(\boldsymbol{x}))}$ of unconditional extreme value estimators in heavy-tailed models [see *e.g.* 19, p.77], and the case $\rho(\boldsymbol{x}) \to -\infty$, corresponding to the ideal but unrealistic case when all the Y_t such that $\boldsymbol{X}_t \in B(\boldsymbol{x}, h_n)$ can be used, yields the optimal convergence rate $n^{-2/(p+4)}$, *i.e.* the optimal convergence rate of nonparametric estimators of a twice continuously differentiable *central* conditional quantile, see [9].

A.4. Auxiliary results and their proofs. The first lemma discusses the validity of some of our assumptions under criteria on certain conditional densities of the process $((\boldsymbol{X}_t, Y_t))_{t\ge1}$ when they exist.

LEMMA A.1. *Let $((\boldsymbol{X}_t, Y_t))_{t\ge1}$ be a stationary sequence of copies of a random vector $(\boldsymbol{X}, Y)$.*

(i) *Assume that the random vector $(\boldsymbol{X}, Y)$ has a joint p.d.f. f, and let g be the p.d.f. of $\boldsymbol{X}$ and $f_{\boldsymbol{X}|Y}(\cdot|y)$ be the p.d.f. of $\boldsymbol{X}$ given $Y = y$. Suppose that there is a neighborhood V of $\boldsymbol{x}$ with*

$$\inf_{\boldsymbol{x}'\in V} g(\boldsymbol{x}') > 0 \ \text{ and } \ \sup_{\boldsymbol{x}'\in V}\sup_{y\in\mathbb{R}} f_{\boldsymbol{X}|Y}(\boldsymbol{x}'|y) < \infty.$$

Then one has, for any $p > 0$,

$$\mathbb{E}(|Y|^p) < \infty \Rightarrow \sup_{\boldsymbol{x}'\in V}\mathbb{E}(|Y|^p|\boldsymbol{X} = \boldsymbol{x}') < \infty.$$

(ii) *Assume that for any $t \ge 1$, the random vector $(\boldsymbol{X}_1, Y_1, \boldsymbol{X}_{t+1}, Y_{t+1})$ has a joint p.d.f. $f_{\boldsymbol{X}_1,Y_1,\boldsymbol{X}_{t+1},Y_{t+1}}$. Let g_t be the joint p.d.f. of $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ and $f_{\boldsymbol{X}_1,\boldsymbol{X}_{t+1}|Y_1,Y_{t+1}}(\cdot,\cdot|y_1,y_{t+1})$ be the joint p.d.f. of $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ given $\{Y_1 = y_1, Y_{t+1} = y_{t+1}\}$. Suppose that there is a neighborhood U of $(\boldsymbol{x}, \boldsymbol{x})$ with*

$$\inf_{t\ge1}\inf_{(\boldsymbol{x}_1,\boldsymbol{x}_{t+1})\in U} g_t(\boldsymbol{x}_1, \boldsymbol{x}_{t+1}) > 0$$

$$\text{and } \sup_{t\ge1}\sup_{(\boldsymbol{x}_1,\boldsymbol{x}_{t+1})\in U}\sup_{y_1,y_{t+1}\in\mathbb{R}} f_{\boldsymbol{X}_1,\boldsymbol{X}_{t+1}|Y_1,Y_{t+1}}(\boldsymbol{x}_1, \boldsymbol{x}_{t+1}|y_1, y_{t+1}) < \infty.$$

Then, for any $p > 0$,

$$\mathbb{E}(|Y|^p) < \infty \Rightarrow \sup_{t\ge1}\sup_{(\boldsymbol{x}_1,\boldsymbol{x}_{t+1})\in U}\mathbb{E}(|Y_1|^p + |Y_{t+1}|^p|\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_{t+1} = \boldsymbol{x}_{t+1}) < \infty.$$

(iii) *Suppose that there exists $t_0 \geq 0$ such that, for $t > t_0$, Y_{t+1} is conditionally independent of $(\boldsymbol{X}_1, Y_1)$ given $\boldsymbol{X}_{t+1}$, and $(\boldsymbol{X}_1, Y_1, \boldsymbol{X}_{t+1})$ has a joint p.d.f $f_{\boldsymbol{X}_1,Y_1,\boldsymbol{X}_{t+1}}$. Let $f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\cdot|\boldsymbol{x}_1)$ be the conditional p.d.f. of $\boldsymbol{X}_{t+1}$ given $\{\boldsymbol{X}_1 = \boldsymbol{x}_1\}$, and $f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1,Y_1}(\cdot|\boldsymbol{x}_1, y_1)$ be the conditional p.d.f. of $\boldsymbol{X}_{t+1}$ given $\{\boldsymbol{X}_1 = \boldsymbol{x}_1, Y_1 = y_1\}$. Suppose also that there exist y_0 large enough and $r > 0$ such that*

$$\inf_{t>t_0} \inf_{\boldsymbol{x}_1,\boldsymbol{x}_{t+1}\in B(\boldsymbol{x},r)} f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1) > 0$$

$$\text{and} \sup_{t>t_0} \sup_{\boldsymbol{x}_1,\boldsymbol{x}_{t+1}\in B(\boldsymbol{x},r)} \sup_{y_1\geq y_0} f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1,Y_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1, y_1) < \infty.$$

Then there is a finite positive constant c such that

$$\sup_{t>t_0} \sup_{\substack{y,y'\geq y_0\\ \boldsymbol{x}',\boldsymbol{x}''\in B(\boldsymbol{x},r)}} \frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y'|\boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}} \leq c \times \sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \overline{F}(y_0|\boldsymbol{x}').$$

PROOF. The proofs of (i) and (ii) are similar; we only prove (ii). Take an open neighborhood U of $(\boldsymbol{x}_1, \boldsymbol{x}_{t+1})$ as in the statement of the result. Then, for any $(\boldsymbol{x}_1, \boldsymbol{x}_{t+1}) \in U$,

$$\begin{aligned}
&\mathbb{E}(|Y_1|^p + |Y_{t+1}|^p|\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_{t+1} = \boldsymbol{x}_{t+1}) \\
&= \int_{\mathbb{R}^2} (|y_1|^p + |y_{t+1}|^p) \frac{f_{\boldsymbol{X}_1,Y_1,\boldsymbol{X}_{t+1},Y_{t+1}}(\boldsymbol{x}_1, y_1, \boldsymbol{x}_{t+1}, y_{t+1})}{g_t(\boldsymbol{x}_1, \boldsymbol{x}_{t+1})} \, dy_1 \, dy_{t+1} \\
&\leq \frac{1}{g_t(\boldsymbol{x}_1, \boldsymbol{x}_{t+1})} \times \sup_{y_1,y_{t+1}\in\mathbb{R}} f_{\boldsymbol{X}_1,\boldsymbol{X}_{t+1}|Y_1,Y_{t+1}}(\boldsymbol{x}_1, \boldsymbol{x}_{t+1}|y_1, y_{t+1}) \times 2\,\mathbb{E}(|Y|^p).
\end{aligned}$$

Take suprema with respect to $t \geq 1$ and $(\boldsymbol{x}_1, \boldsymbol{x}_{t+1}) \in U$ to conclude the proof of (ii).

To show (iii), use first the conditional independence assumption and the stationarity property to obtain

$$\begin{aligned}
&\sup_{t>t_0} \sup_{\substack{y,y'\geq y_0\\ \boldsymbol{x}',\boldsymbol{x}''\in B(\boldsymbol{x},r)}} \frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y'|\boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}} \\
&= \sup_{t>t_0} \sup_{\substack{y,y'\geq y_0\\ \boldsymbol{x}',\boldsymbol{x}''\in B(\boldsymbol{x},r)}} \frac{\mathbb{P}(Y_1 > y|\boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')\mathbb{P}(Y_{t+1} > y'|\boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}} \\
&= \sup_{\boldsymbol{x}''\in B(\boldsymbol{x},r)} \sqrt{\overline{F}(y_0|\boldsymbol{x}'')} \left\{ \sup_{t>t_0} \sup_{\substack{y\geq y_0\\ \boldsymbol{x}'\in B(\boldsymbol{x},r)}} \frac{\mathbb{P}(Y_1 > y|\boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')}} \right\}.
\end{aligned}$$

Now, for any $y \geq y_0$ and $\boldsymbol{x}', \boldsymbol{x}'' \in B(\boldsymbol{x}, r)$,

$$\begin{aligned}
\mathbb{P}(Y_1 > y|\boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'') &= \int_y^\infty \frac{f_{\boldsymbol{X}_1,Y_1,\boldsymbol{X}_{t+1}}(\boldsymbol{x}', y_1, \boldsymbol{x}'')}{f_{\boldsymbol{X}_1,\boldsymbol{X}_{t+1}}(\boldsymbol{x}', \boldsymbol{x}'')} dy_1 \\
&= \int_y^\infty \frac{f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1,Y_1}(\boldsymbol{x}''|\boldsymbol{x}', y_1)}{f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}''|\boldsymbol{x}')} \times \frac{f_{\boldsymbol{X}_1,Y_1}(\boldsymbol{x}', y_1)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}')} dy_1 \\
&\leq \overline{F}(y|\boldsymbol{x}') \times \sup_{\substack{y_1\geq y_0\\ \boldsymbol{x}_1,\boldsymbol{x}_{t+1}\in B(\boldsymbol{x},r)}} \frac{f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1,Y_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1, y_1)}{f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1)}.
\end{aligned}$$

The result readily follows. □

Lemma A.3 and Proposition A.1 deal with the estimation of the p.d.f. and regression function. Before stating these results, we collect in Lemma A.2 below two elementary results of multivariate calculus that we will extensively use in our subsequent proofs and in the analysis of our examples in Section B.

LEMMA A.2. (i) *Let $f:\mathbb{R}^p\to\mathbb{R}$ be continuously differentiable in a neighborhood of $\boldsymbol{x}$ and assume that its gradient ∇f is Lipschitz continuous at $\boldsymbol{x}$, that is,*

$$\exists c,r>0,\ \forall \boldsymbol{x}'\in B(\boldsymbol{x},r),\ \|\nabla f(\boldsymbol{x}')-\nabla f(\boldsymbol{x})\|\le c\|\boldsymbol{x}'-\boldsymbol{x}\|.$$

Then

$$\sup_{\|\boldsymbol{u}\|\le 1}|f(\boldsymbol{x}-h\boldsymbol{u})-f(\boldsymbol{x})+h\boldsymbol{u}^\top\nabla f(\boldsymbol{x})|=O(h^2) \text{ as } h\to 0.$$

(ii) *Suppose that, for a function $\phi:(\boldsymbol{x},y)\in\mathbb{R}^p\times\mathbb{R}\mapsto\phi(\boldsymbol{x},y)\in\mathbb{R}$, there exist $r>0$, $\boldsymbol{x}\in\mathbb{R}^p$ and a nonempty set $\mathcal{Y}\subset\mathbb{R}$ such that, for any $y\in\mathcal{Y}$, $\phi(\cdot,y)$ is twice continuously differentiable on $B(\boldsymbol{x},r)$ and $\sup_{y\in\mathcal{Y}}\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)}\|H_{\boldsymbol{x}}\phi(\boldsymbol{x}',y)\|<\infty$. Then*

$$\sup_{y\in\mathcal{Y}}\sup_{\substack{\boldsymbol{x}'\in B(\boldsymbol{x},r)\\ \boldsymbol{x}'\neq\boldsymbol{x}}}\frac{1}{\|\boldsymbol{x}'-\boldsymbol{x}\|^2}|\phi(\boldsymbol{x}',y)-\phi(\boldsymbol{x},y)-(\boldsymbol{x}'-\boldsymbol{x})^\top\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},y)|<\infty.$$

PROOF. (i) This is a straightforward consequence of the identity

$$f(\boldsymbol{x}-h\boldsymbol{u})-f(\boldsymbol{x})=-h\boldsymbol{u}^\top\nabla f(\boldsymbol{x})-h\int_0^1\boldsymbol{u}^\top[\nabla f(\boldsymbol{x}-th\boldsymbol{u})-\nabla f(\boldsymbol{x})]\,dt$$

valid for $|h|\le r$ and $\|\boldsymbol{u}\|\le 1$.

(ii) Fix $y\in\mathcal{Y}$ and $\boldsymbol{x}'\in B(\boldsymbol{x},r)$ and write similarly

$$\phi(\boldsymbol{x}',y)-\phi(\boldsymbol{x},y)-(\boldsymbol{x}'-\boldsymbol{x})^\top\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},y)=(\boldsymbol{x}'-\boldsymbol{x})^\top\int_0^1[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}+t(\boldsymbol{x}'-\boldsymbol{x}),y)-\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},y)]\,dt.$$

By the mean value theorem, for any $t\in[0,1]$,

$$\|\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}+t(\boldsymbol{x}'-\boldsymbol{x}),y)-\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},y)\|\le t\|\boldsymbol{x}'-\boldsymbol{x}\|\times\sup_{y\in\mathcal{Y}}\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)}\|H_{\boldsymbol{x}}\phi(\boldsymbol{x}',y)\|.$$

The conclusion readily follows. □

Lemma A.3 collects useful asymptotic expansions about smoothed conditional moments.

LEMMA A.3. *Suppose that conditions $\mathcal{K}$ and $\mathcal{L}_g$ hold. Assume that $h_n\to 0$ as $n\to\infty$.*

(i) *Then, for any $b>0$,*

$$\mathbb{E}\left(K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right)-h_n^p\left(\int_{\mathbb{R}^p}K^b\right)g(\boldsymbol{x})=O(h_n^{p+1}).$$

(ii) *If in fact conditions $\mathcal{KS}$ and $\mathcal{D}_g$ hold then, for any $b>0$,*

$$\mathbb{E}\left(K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right)-h_n^p\left(\int_{\mathbb{R}^p}K^b\right)g(\boldsymbol{x})=O(h_n^{p+2}).$$

(iii) *Suppose moreover that condition $\mathcal{L}_m$ holds. Then, for any $a\in\{1,2\}$ and $b>0$,*

$$\mathbb{E}\left(Y^aK^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right)-h_n^p\left(\int_{\mathbb{R}^p}K^b\right)\mathbb{E}(Y^a|\boldsymbol{X}=\boldsymbol{x})g(\boldsymbol{x})=O(h_n^{p+1}).$$

(iv) *If in fact conditions $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_m$ hold then, for any $a \in \{1,2\}$ and $b > 0$,*

$$\mathbb{E}\left(Y^a K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) - h_n^p\left(\int_{\mathbb{R}^p} K^b\right)\mathbb{E}(Y^a|\boldsymbol{X}=\boldsymbol{x})g(\boldsymbol{x}) = O(h_n^{p+2}).$$

PROOF. Note, for $a = 0,1,2$, the identity

$$\mathbb{E}\left(Y^a K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) = h_n^p\int_{\|\boldsymbol{u}\|\le 1} K^b(\boldsymbol{u})\mathbb{E}(Y^a|\boldsymbol{X}=\boldsymbol{x}-h_n\boldsymbol{u})g(\boldsymbol{x}-h_n\boldsymbol{u})d\boldsymbol{u}$$

(under condition $\mathcal{L}_m$ when $a \in \{1,2\}$). Statements (i) and (iii) then immediately follow by Lipschitz continuity. To show (ii) and (iv), apply Lemma A.2(i) and use the identity $\int_{\|\boldsymbol{u}\|\le 1} K^b(\boldsymbol{u})\boldsymbol{u}\,d\boldsymbol{u} = \mathbf{0}_p$ due to the symmetry of K. □

Proposition A.1 below is a version, tailored to our needs, of the variance approximation result in Theorem 1 of [36], under a weaker condition on the mixing rate.

PROPOSITION A.1. *Suppose that $((\boldsymbol{X}_t, Y_t))_{t\ge 1}$ is stationary and $\alpha-$mixing, and that conditions $\mathcal{K}$, $\mathcal{L}_g$ and $\mathcal{B}_p$ hold. Assume that $h_n \to 0$ and $nh_n^p \to \infty$ as $n \to \infty$.*

(i) *If there exists $\eta > 1$ with $\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$, then*

$$\operatorname{Var}(\widehat{g}_n(\boldsymbol{x})) = O\left(\frac{1}{nh_n^p}\right).$$

If $((\boldsymbol{X}_t, Y_t))_{t\ge 1}$ is in fact $\rho-$mixing, condition $\mathcal{B}_p$ can be dropped and assumption $\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$ may be replaced by $\sum_{j=1}^{\infty} \rho(j) < \infty$.

(ii) *If moreover conditions $\mathcal{L}_m$ and $\mathcal{B}_m$ hold, and if there exists $\delta > 0$ with $\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \mathbb{E}(|Y|^{2+\delta}|\boldsymbol{X} = \boldsymbol{x}') < \infty$ for a certain $r > 0$ and $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2+\delta)$, then*

$$\operatorname{Var}(\widehat{m}_n(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x})) = O\left(\frac{1}{nh_n^p}\right).$$

If $((\boldsymbol{X}_t, Y_t))_{t\ge 1}$ is in fact $\rho-$mixing, conditions $\mathcal{B}_p$, $\mathcal{B}_m$ and $\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \mathbb{E}(|Y|^{2+\delta}|\boldsymbol{X} = \boldsymbol{x}') < \infty$ can be dropped and assumption $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ may be replaced by $\sum_{j=1}^{\infty} \rho(j) < \infty$.

It is worth noting that in the above proposition, the condition on the rate of convergence to 0 of the strong mixing coefficient is stronger than in the non-regression case. For instance, if the random average

$$\frac{1}{n}\sum_{t=1}^{n} Y_t \mathbb{1}_{\{X_t\in A\}}$$

is considered for a fixed A, with $\mathbb{E}(|Y|^{2+\delta}) < \infty$, then a central limit theorem holds as soon as $\sum_{j=1}^{\infty}[\alpha(j)]^{\delta/(2+\delta)} < \infty$. In our regression framework where $A = A_n$ has probability converging to 0, this criterion becomes $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for a certain $\eta > \delta/(2+\delta)$. That assumption on the strong mixing rate is standard, see for example the remark below condition A2 in [35]. Note also that assumption ($\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2+\delta)$) is stronger than assumption ($\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$ for some $\eta > 1$) since $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} = \sum_{j=1}^{\infty}[j^{\eta(2+\delta)/\delta}\alpha(j)]^{\delta/(2+\delta)}$.

PROOF. We start by showing (i). Obviously

$$\mathrm{Var}(\widehat{g}_n(\boldsymbol{x})) = \frac{1}{nh_n^{2p}} \times \mathrm{Var}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) + \frac{1}{nh_n^{2p}} \times 2\sum_{j=1}^{n-1} \frac{n-j}{n} \mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right).$$

Note that, by Lemma A.3(i), $\mathrm{Var}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) = O(h_n^p)$, and for n large enough,

$$\begin{aligned}
&\left|\mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \\
&\leq \mathbb{E}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right) K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) + \mathbb{E}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right)\right) \mathbb{E}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) \\
&\leq \left(\sup_{\mathbb{R}^p} K^2\right) \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, h_n), \boldsymbol{X}_{j+1} \in B(\boldsymbol{x}, h_n)) + (2g(\boldsymbol{x})h_n^p)^2
\end{aligned}$$

where conditions $\mathcal{K}$ and $\mathcal{L}_g$ were used. Using condition $\mathcal{B}_p$, we find, for any $j \geq 1$,

$$\left|\mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \leq h_n^p \varepsilon(h_n) \mathbb{1}_{\{j<t_0\}} + Ch_n^{2p} \mathbb{1}_{\{j \geq t_0\}}, \tag{1}$$

where $\varepsilon(h) \to 0$ as $h \to 0$ and C is a positive constant (not depending on j; throughout this proof the expressions of the function ε and of the constant C may change from line to line). Besides, the fact that K is bounded makes it possible to apply Ibragimov's inequality [34] to the $\alpha-$mixing sequence $(\boldsymbol{X}_t)$, yielding

$$\left|\mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \leq C\,\alpha(j). \tag{2}$$

Combine (1) and (2) to get, for any sequence $v_n < n-1$ tending to infinity,

$$\begin{aligned}
&\frac{1}{nh_n^{2p}} \times 2\sum_{j=1}^{n-1} \frac{n-j}{n} \mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) \\
&= O\left(\frac{1}{nh_n^p}\left[\varepsilon(h_n) + v_n h_n^p + h_n^{-p} \sum_{j=v_n+1}^{n-1} \alpha(j)\right]\right) \text{ (splitting the sum at } t_0 \text{ and } v_n) \\
&= O\left(\frac{1}{nh_n^p}\left[\varepsilon(h_n) + v_n h_n^p + (v_n h_n^p)^{-1} v_n^{1-\eta} \sum_{j=v_n+1}^{n-1} j^\eta \alpha(j)\right]\right) \\
&= O\left(\frac{1}{nh_n^p}\left[1 + v_n h_n^p + (v_n h_n^p)^{-1}\right]\right) \text{ (because } \eta > 1).
\end{aligned}$$

Choosing $v_n = h_n^{-p} \to \infty$ (which indeed satisfies $v_n < n-1$ for n large enough since $nh_n^p \to \infty$) entails

$$\frac{1}{nh_n^{2p}} \times 2\sum_{j=1}^{n-1} \frac{n-j}{n} \mathrm{Cov}\left(K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) = \mathrm{O}\left(\frac{1}{nh_n^p}\right)$$

and concludes the proof of (i).

We now prove (ii). The arguments are inspired by the proof of Theorem 2(b) in [36], with a couple of crucial modifications due to the fact that we do not assume that $(\boldsymbol{X}_1, \boldsymbol{X}_{j+1})$ has a p.d.f. with respect to the Lebesgue measure on $\mathbb{R}^{2p}$. Write

$$\begin{aligned}\operatorname{Var}(\widehat{m}_n(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x})) &= \frac{1}{nh_n^{2p}} \times \operatorname{Var}\left(Y\,K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) \\ &\quad + \frac{1}{nh_n^{2p}} \times 2\sum_{j=1}^{n-1} \frac{n-j}{n} \operatorname{Cov}\left(Y_1 K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1} K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right).\end{aligned}$$

Note that, by Lemma A.3(iii), $\operatorname{Var}\left(Y\,K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) = O(h_n^p)$, and also, for n large enough,

$$\begin{aligned}&\left|\operatorname{Cov}\left(Y_1 K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1} K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \\ &\le \mathbb{E}\left(|Y_1 Y_{j+1}| K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right) K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) + (|m(\boldsymbol{x})|g(\boldsymbol{x})+1)^2 h_n^{2p} \\ &\le \frac{1}{2}\mathbb{E}\left((Y_1^2+Y_{j+1}^2) K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right) K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) + (|m(\boldsymbol{x})|g(\boldsymbol{x})+1)^2 h_n^{2p} \\ &\le C\,\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x},h_n), \boldsymbol{X}_{j+1} \in B(\boldsymbol{x},h_n)) + (|m(\boldsymbol{x})|g(\boldsymbol{x})+1)^2 h_n^{2p}\end{aligned}$$

thanks to assumptions $\mathcal{K}$ and $\mathcal{B}_m$. It follows that, for any $j \ge 1$,

$$\left|\operatorname{Cov}\left(Y_1 K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1} K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \le h_n^p \varepsilon(h_n) \mathbb{1}_{\{j<t_0\}} + C h_n^{2p} \mathbb{1}_{\{j \ge t_0\}}. \tag{3}$$

Besides, a straightforward adaptation of the proof of (2.17) in [36] in dimension p (by replacing $C_h(u)$ therein by $h^{-p}K(u/h)$, and noting the typo one line before (2.17) therein, where the expression in full should read $M_3 \int_{-\infty}^{+\infty} |C_h(u-x)|^\delta f(u)\,du$) leads to

$$\left|\operatorname{Cov}\left(Y_1 K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1} K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right| \le C h_n^{2p/(2+\delta)} [\alpha(j)]^{\delta/(2+\delta)}. \tag{4}$$

Combining (3) and (4) and arguing as in the final stages of the proof of (i) with $v_n = h_n^{-p}$ yields

$$\begin{aligned}&\frac{1}{nh_n^{2p}} \times 2\sum_{j=1}^{n-1} \frac{n-j}{n} \operatorname{Cov}\left(Y_1 K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1} K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right) \\ &= O\left(\frac{1}{nh_n^p}\left[\varepsilon(h_n) + v_n h_n^p + h_n^{-p\delta/(2+\delta)} \sum_{j=v_n+1}^{n-1} [\alpha(j)]^{\delta/(2+\delta)}\right]\right) \\ &= O\left(\frac{1}{nh_n^p}\left[\varepsilon(h_n) + v_n h_n^p + (v_n h_n^p)^{-\delta/(2+\delta)} v_n^{\delta/(2+\delta)-\eta} \sum_{j=v_n+1}^{n-1} j^\eta [\alpha(j)]^{\delta/(2+\delta)}\right]\right) \\ &= O\left(\frac{1}{nh_n^p}\left[1 + v_n h_n^p + (v_n h_n^p)^{-\delta/(2+\delta)}\right]\right) = O\left(\frac{1}{nh_n^p}\right)\end{aligned}$$

and concludes the proof in this case.

For both (i) and (ii), the results under the $\rho-$mixing assumption follow from writing, for $a = 0, 1$,

$$\left|\mathrm{Cov}\left(Y_1^a\, K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right), Y_{j+1}^a\, K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{j+1}}{h_n}\right)\right)\right|$$
$$\leq \rho(j)\left\{\mathbb{E}\left(Y_1^{2a} K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right)\right) - \left(\mathbb{E}\left(Y_1^a K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_1}{h_n}\right)\right)\right)^2\right\}$$

and using Lemma A.3. □

Our next objective is to quantify the bias of our nonparametric estimators, under the regularity conditions provided in Sections 2 and 3. Because condition $\mathcal{D}_\omega$ is somewhat involved, we first give in Lemma A.4 below simple sufficient conditions ensuring that this regularity assumption is met. We shall extensively use this result for working out our examples in Section B.

LEMMA A.4. *Assume that, for some $r > 0$, one has, for any $\boldsymbol{x}' \in B(\boldsymbol{x}, r)$,*

$$\forall y > 0,\ \lim_{t\to\infty} \frac{\overline{F}(ty|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x}')} = y^{-1/\gamma(\boldsymbol{x}')}$$

for a certain (strictly) positive function γ. Set $L(y|\boldsymbol{x}') = y^{1/\gamma(\boldsymbol{x}')}\overline{F}(y|\boldsymbol{x}')$ for such $\boldsymbol{x}'$.

(i) *Suppose that there exists $y_0 \geq 0$ such that the functions $\nabla_{\boldsymbol{x}} \log L(y|\cdot)/\log(y)$, for $y \geq y_0$, are well-defined, continuous on $B(\boldsymbol{x}, r)$, and define an equicontinuous family at $\boldsymbol{x}$, namely:*

$$\forall y \geq y_0,\ \forall \varepsilon > 0,\ \exists \delta > 0,\ \boldsymbol{x}' \in B(\boldsymbol{x}, \delta) \Rightarrow \left\|\frac{\nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x}')}{\log(y)} - \frac{\nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x})}{\log(y)}\right\| \leq \varepsilon.$$

Then $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x})/\log(y) = 0$.

(ii) *Suppose that γ is twice continuously differentiable on $B(\boldsymbol{x}, r)$. If the partial Hessian matrix $H_{\boldsymbol{x}} \log L(y|\boldsymbol{x}')/\log(y)$ (or equivalently $H_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}')/\log(y)$) is well-defined and uniformly bounded in $\boldsymbol{x}' \in B(\boldsymbol{x}, r)$ and y large enough, then condition $\mathcal{D}_\omega$ holds with $\boldsymbol{\mu}(\boldsymbol{x}) = \lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y) = \nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x})$.*

PROOF. We start by proving (i). The function $L(\cdot|\boldsymbol{x}')$ is slowly varying, and as such, $\log L(y|\boldsymbol{x}')/\log(y) \to 0$ as $y \to \infty$ by Proposition 1.3.6(i) on p.16 of [2]. It follows that the function $\phi : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$ defined as $\phi(\boldsymbol{x}, y) = \log L(y|\boldsymbol{x})/\log(y)$ satisfies:

- $\phi(\boldsymbol{x}', y) \to 0$ as $y \to \infty$ for any $\boldsymbol{x}' \in B(\boldsymbol{x}, r)$,
- For any $y \geq y_0$, $\phi(\cdot, y)$ is continuously differentiable on $B(\boldsymbol{x}, r)$,
- The family of functions $\nabla_{\boldsymbol{x}}\phi(\cdot, y)$, for $y \geq y_0$, is equicontinuous at $\boldsymbol{x}$.

It is enough to prove then that $\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}, y) \to 0$ as $y \to \infty$.

Suppose not. Then there is $j \in \{1, \ldots, p\}$ such that $(\partial\phi/\partial x_j)(\boldsymbol{x}, y)$ does not converge to 0 as $y \to \infty$. Up to reordering the variables x_j, we may and will assume that $j = 1$, so that there exists $\varepsilon > 0$ and a sequence $y_n \to \infty$ such that $|(\partial\phi/\partial x_1)(\boldsymbol{x}, y_n)| > 2\varepsilon$ for any n. Equicontinuity of $\nabla_{\boldsymbol{x}}\phi(\cdot, y)$ obviously implies equicontinuity of each partial derivative, so we can choose $\delta \in (0, r)$ such that

$$\boldsymbol{x}' \in B(\boldsymbol{x}, \delta) \Rightarrow \left|\frac{\partial\phi}{\partial x_1}(\boldsymbol{x}', y_n) - \frac{\partial\phi}{\partial x_1}(\boldsymbol{x}, y_n)\right| \leq \varepsilon \text{ for any } n.$$

Let $\boldsymbol{e} = (1, 0, \ldots, 0)^\top \in \mathbb{R}^p$. Write

$$\phi(\boldsymbol{x} + \delta\boldsymbol{e}, y_n) - \phi(\boldsymbol{x}, y_n) = \delta \frac{\partial \phi}{\partial x_1}(\boldsymbol{x}, y_n) + \int_0^\delta \left(\frac{\partial \phi}{\partial x_1}(\boldsymbol{x} + t\boldsymbol{e}, y_n) - \frac{\partial \phi}{\partial x_1}(\boldsymbol{x}, y_n) \right) dt$$

and apply the reverse triangle inequality to obtain

$$|\phi(\boldsymbol{x} + \delta\boldsymbol{e}, y_n) - \phi(\boldsymbol{x}, y_n)| \geq \delta \left| \frac{\partial \phi}{\partial x_1}(\boldsymbol{x}, y_n) \right| - \delta\varepsilon \geq \delta\varepsilon > 0.$$

The left-hand side converges to 0 as $n \to \infty$, which is a contradiction. This completes the proof of (i).

To prove (ii), let $y_0 \geq 0$ be such that $\|H_{\boldsymbol{x}} \log L(y|\boldsymbol{x}')/\log(y)\| \leq c$, a finite constant, for $\boldsymbol{x}' \in B(\boldsymbol{x}, r)$ and $y \geq y_0$. Then, by the mean value theorem,

$$\forall y \geq y_0,\ \boldsymbol{x}' \in B(\boldsymbol{x}, r) \Rightarrow \left\| \frac{\nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x}')}{\log(y)} - \frac{\nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x})}{\log(y)} \right\| \leq c\|\boldsymbol{x}' - \boldsymbol{x}\|.$$

This equi-Lipschitz property at $\boldsymbol{x}$ guarantees in particular that the equicontinuity assumption in (i) is satisfied and therefore that

$$\frac{\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})}{\log(y)} = \frac{\nabla\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})} + \frac{\nabla_{\boldsymbol{x}} \log L(y|\boldsymbol{x})}{\log(y)} \to \frac{\nabla\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})} \text{ as } y \to \infty.$$

Besides, the bound

$$\limsup_{y\to\infty} \sup_{\substack{\boldsymbol{x}' \in B(\boldsymbol{x}, r/2) \\ \boldsymbol{x}' \neq \boldsymbol{x}}} \frac{1}{\|\boldsymbol{x}' - \boldsymbol{x}\|^2} \left| \frac{1}{\log(y)} \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} - (\boldsymbol{x}' - \boldsymbol{x})^\top \frac{\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})}{\log(y)} \right| < \infty$$

is an immediate consequence of Lemma A.2(ii). Conclude that condition $\mathcal{D}_\omega$ holds, which is the desired result. □

We will repeatedly use the following lemma in the evaluation of the bias of our nonparametric estimators under conditions $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$ and $\mathcal{D}_\omega$.

LEMMA A.5. *Assume that*

$$\forall y > 0,\ \lim_{t\to\infty} \frac{\overline{F}(ty|\boldsymbol{x})}{\overline{F}(t|\boldsymbol{x})} = y^{-1/\gamma(\boldsymbol{x})}$$

for some $\gamma(\boldsymbol{x}) > 0$*, and let* $a \in (0, 1/\gamma(\boldsymbol{x}))$*. Assume further that* $\omega_{h_n}(y_n|\boldsymbol{x}) \log(y_n) \to 0$ *for some sequences* $y_n \to \infty$ *and* $h_n \to 0$*. Then, for any* $s > 0$*,*

$$\int_{y_n}^\infty (z - y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1 + z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz < \infty$$

$$\text{and } \sup_{\|\boldsymbol{u}\|\leq 1} \int_{y_n}^\infty (z - y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \left| \log\left(\frac{\overline{F}(z|\boldsymbol{x} - h_n\boldsymbol{u})}{\overline{F}(z|\boldsymbol{x})} \right) \right|^s dz < \infty$$

for n *large enough. Moreover,*

$$\frac{\int_{y_n}^\infty (z - y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1 + z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz}{\int_{y_n}^\infty (z - y_n)^{a-1} \overline{F}(z|\boldsymbol{x})\, dz} = O(\log^s(y_n)).$$

PROOF. Pick an arbitrary $\delta > 0$. Proposition B.1.9.1 on p.366 of [19] and the convergences $\omega_{h_n}(y_n|\boldsymbol{x}) \to 0$ and $y_n \to \infty$ yield, for n large enough,

$$\forall z \geq y_n, \ \overline{F}(z|\boldsymbol{x}) z^{\omega_{h_n}(y_n|\boldsymbol{x})} \log^s(z) \leq z^{-1/\gamma(\boldsymbol{x})+\delta}.$$

Since $a < 1/\gamma(\boldsymbol{x})$, the finiteness of the integral $\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1+z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz$ for n large enough follows. Noting that

$$\forall z \geq y_n, \ \sup_{\|\boldsymbol{u}\| \leq 1} \frac{1}{\log^s(z)} \left| \log\left(\frac{\overline{F}(z|\boldsymbol{x} - h_n \boldsymbol{u})}{\overline{F}(z|\boldsymbol{x})} \right) \right|^s \leq \omega_{h_n}^s(y_n|\boldsymbol{x}) \to 0$$

the finiteness of $\sup_{\|\boldsymbol{u}\| \leq 1} \int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \left| \log(\overline{F}(z|\boldsymbol{x} - h_n \boldsymbol{u}) / \overline{F}(z|\boldsymbol{x})) \right|^s dz$ for n large enough follows as well. Write then

$$\begin{aligned}
&\frac{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1+z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz}{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x})\, dz} \\
&= \log^s(y_n) \frac{\int_1^{\infty} (t-1)^{a-1} \frac{\overline{F}(ty_n|\boldsymbol{x}) \log^s(ty_n)}{\overline{F}(y_n|\boldsymbol{x}) \log^s(y_n)} (1 + y_n^{\omega_{h_n}(y_n|\boldsymbol{x})} t^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dt}{\int_1^{\infty} (t-1)^{a-1} \frac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}\, dt} \\
&\leq 2 y_n^{\omega_{h_n}(y_n|\boldsymbol{x})} \log^s(y_n) \frac{\int_1^{\infty} (t-1)^{a-1} \frac{\overline{F}(ty_n|\boldsymbol{x}) \log^s(ty_n)}{\overline{F}(y_n|\boldsymbol{x}) \log^s(y_n)} t^{\omega_{h_n}(y_n|\boldsymbol{x})}\, dt}{\int_1^{\infty} (t-1)^{a-1} \frac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}\, dt}.
\end{aligned}$$

Use Potter bounds [see Proposition B.1.9.5 in 19, p.367] and the convergence $\omega_{h_n}(y_n|\boldsymbol{x}) \log(y_n) \to 0$ to get, for any $\delta > 0$ such that $a - 1/\gamma(\boldsymbol{x}) + \delta < 0$,

$$\frac{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1+z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz}{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x})\, dz} \leq 4 \log^s(y_n) \frac{\int_1^{\infty} (t-1)^{a-1} t^{-1/\gamma(\boldsymbol{x})+\delta}\, dt}{\int_1^{\infty} (t-1)^{a-1} t^{-1/\gamma(\boldsymbol{x})-\delta}\, dt}$$

when n is large enough. Both integrals in the ratio on the second line are finite, and therefore

$$\frac{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x}) \log^s(z)(1+z^{\omega_{h_n}(y_n|\boldsymbol{x})})\, dz}{\int_{y_n}^{\infty} (z-y_n)^{a-1} \overline{F}(z|\boldsymbol{x})\, dz} = O(\log^s(y_n)).$$

This completes the proof. □

We are now in position to state and prove our next key result about the bias of our nonparametric estimators. This lemma extends results of [27] (see Lemmas 2 and 5 therein). These were originally stated under a more restrictive condition than the standard conditional heavy tails condition. In particular, we give a precise quantification of the bias appearing as a result of our kernel smoothing procedures. Throughout the rest of this section we use the notation

$$\widehat{\psi}_n^{(a)}(y|\boldsymbol{x}) = \widehat{\varphi}_n^{(a)}(y|\boldsymbol{x}) \widehat{g}_n(\boldsymbol{x}) \text{ and } \psi^{(a)}(y|\boldsymbol{x}) = \varphi^{(a)}(y|\boldsymbol{x}) g(\boldsymbol{x}),$$

where

$$\widehat{\varphi}_n^{(a)}(y|\boldsymbol{x}) = \frac{1}{\widehat{g}_n(\boldsymbol{x})} \times \frac{1}{n h_n^p} \sum_{t=1}^{n} (Y_t - y)^a \mathbb{1}_{\{Y_t > y\}} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h_n} \right).$$

LEMMA A.6. *Let $\boldsymbol{x} \in \mathbb{R}^p$ be such that $g(\boldsymbol{x}) > 0$. Assume that*

$$\forall y > 0, \ \lim_{t \to \infty} \frac{\overline{F}(ty|\boldsymbol{x})}{\overline{F}(t|\boldsymbol{x})} = y^{-1/\gamma(\boldsymbol{x})}$$

for some $\gamma(\boldsymbol{x}) > 0$, and let $a \in [0, 1/\gamma(\boldsymbol{x}))$.

(i) *Then* $\psi^{(a)}(y|\boldsymbol{x}) = \dfrac{B(a+1, 1/\gamma(\boldsymbol{x})-a)}{\gamma(\boldsymbol{x})} g(\boldsymbol{x}) y^a \overline{F}(y|\boldsymbol{x})(1+o(1))$ *as* $y \to \infty$.

(ii) *Assume further that conditions* $\mathcal{K}$, $\mathcal{L}_g$ *and* $\mathcal{L}_\omega$ *hold. Let* $y_n \to \infty$, $h_n \to 0$ *be such that* $h_n \log(y_n) \to 0$. *Then one has, for any* $b > 0$,

$$\mathbb{E}\left((Y - y_n)^a \mathbb{1}_{\{Y > y_n\}} K^b \left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n} \right) \right) = h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) \left(\int_{\mathbb{R}^p} K^b \right) (1 + O(h_n \log(y_n))).$$

(iii) *If moreover conditions* $\mathcal{KS}$, $\mathcal{D}_g$ *and* $\mathcal{D}_\omega$ *hold then*

$$\mathbb{E}\left((Y - y_n)^a \mathbb{1}_{\{Y > y_n\}} K^b \left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n} \right) \right) = h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) \left(\int_{\mathbb{R}^p} K^b \right) \times \left(1 + \frac{h_n^2 \log^2(y_n)}{2} \int_{\mathbb{R}^p} \frac{K^b(\boldsymbol{u})}{\int_{\mathbb{R}^p} K^b} (\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2 \, d\boldsymbol{u} + o(h_n^2 \log^2(y_n)) \right)$$

where $\boldsymbol{\mu}(\boldsymbol{x}) = \lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}) / \log(y)$ *as defined in condition* $\mathcal{D}_\omega$.

In particular, for $b = 1$ and under the assumptions of Lemma A.6(ii),

$$h_n^p \mathbb{E}(\widehat{\psi}_n^{(a)}(y_n|\boldsymbol{x})) = h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) (1 + O(h_n \log(y_n))).$$

Under the assumptions of Lemma A.6(iii),

$$h_n^p \mathbb{E}(\widehat{\psi}_n^{(a)}(y_n|\boldsymbol{x})) = h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) \times \left(1 + \frac{h_n^2 \log^2(y_n)}{2} \int_{\mathbb{R}^p} K(\boldsymbol{u}) (\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2 \, d\boldsymbol{u} + o(h_n^2 \log^2(y_n)) \right).$$

PROOF. Statement (i) is proven exactly like Lemma 2 in [27], noting that the differentiability condition therein on the conditional survival function is unnecessary. To show statements (ii) and (iii), the fundamental identity is, using condition $\mathcal{K}$,

$$\frac{\mathbb{E}\left((Y - y_n)^a \mathbb{1}_{\{Y > y_n\}} K^b \left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n} \right) \right)}{h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) \int_{\mathbb{R}^p} K^b} - 1 = \int_{\|\boldsymbol{u}\| \le 1} \frac{K^b(\boldsymbol{u})}{\int_{\mathbb{R}^p} K^b} \left(\frac{\psi^{(a)}(y_n|\boldsymbol{x} - h_n \boldsymbol{u})}{\psi^{(a)}(y_n|\boldsymbol{x})} - 1 \right) d\boldsymbol{u}. \tag{5}$$

The identity $ab - 1 = (a-1) + (b-1) + (a-1)(b-1)$ yields

$$\frac{\psi^{(a)}(y_n|\boldsymbol{x} - h_n \boldsymbol{u})}{\psi^{(a)}(y_n|\boldsymbol{x})} - 1 = \left[\frac{g(\boldsymbol{x} - h_n \boldsymbol{u})}{g(\boldsymbol{x})} - 1 \right] + \left[\frac{\varphi^{(a)}(y_n|\boldsymbol{x} - h_n \boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})} - 1 \right] + \left[\frac{g(\boldsymbol{x} - h_n \boldsymbol{u})}{g(\boldsymbol{x})} - 1 \right] \times \left[\frac{\varphi^{(a)}(y_n|\boldsymbol{x} - h_n \boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})} - 1 \right]. \tag{6}$$

Obviously

$$\sup_{\|\boldsymbol{u}\| \le 1} \left| \frac{g(\boldsymbol{x} - h_n \boldsymbol{u})}{g(\boldsymbol{x})} - 1 \right| = O(h_n) \tag{7}$$

and following the proof of Lemma 3 in [27],

$$\sup_{\|\boldsymbol{u}\| \le 1} \left| \frac{\varphi^{(a)}(y_n|\boldsymbol{x} - h_n \boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})} - 1 \right| = O(\omega_{h_n}(y_n|\boldsymbol{x}) \log(y_n)). \tag{8}$$

Combining Equations (5)–(8) with the fact that $\omega_{h_n}(y_n|\boldsymbol{x}) = O(h_n)$ (by condition $\mathcal{L}_\omega$) immediately provides (ii).

To prove (iii), note that by condition $\mathcal{D}_g$ and Lemma A.2(i),

$$(9)\qquad \sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{g(\boldsymbol{x}-h_n\boldsymbol{u})}{g(\boldsymbol{x})}-1+h_n\boldsymbol{u}^\top\frac{\nabla g(\boldsymbol{x})}{g(\boldsymbol{x})}\right|=O(h_n^2).$$

We then first treat the case $a=0$, when $\varphi^{(0)}(\cdot|\cdot)=\overline{F}(\cdot|\cdot)$. Set, for any t, $t'>0$,

$$\Delta(t,t')=\frac{t}{t'}-1-\log(t/t')-\frac{\log^2(t/t')}{2}=\exp(\log(t/t'))-1-\log(t/t')-\frac{\log^2(t/t')}{2}.$$

A Taylor formula for the exponential function on the interval linking 0 to $\log(t/t')$ gives

$$|\Delta(t,t')|\le\frac{|\log^3(t/t')|}{6}\exp(|\log(t/t')|).$$

Applied to $t=\overline{F}(z|\boldsymbol{x}-h_n\boldsymbol{u})$ and $t'=\overline{F}(z|\boldsymbol{x})$, this inequality yields, for n large enough,

$$(10)\qquad \forall z\ge y_n,\ \sup_{\|\boldsymbol{u}\|\le 1}|\Delta(\overline{F}(z|\boldsymbol{x}-h_n\boldsymbol{u}),\overline{F}(z|\boldsymbol{x}))|\le\frac{\omega_{h_n}^3(y_n|\boldsymbol{x})}{6}\log^3(z)z^{\omega_{h_n}(y_n|\boldsymbol{x})}.$$

In particular, at $z=y_n$,

$$\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{\overline{F}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y_n|\boldsymbol{x})}-1-\log\left(\frac{\overline{F}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y_n|\boldsymbol{x})}\right)-\frac{1}{2}\log^2\left(\frac{\overline{F}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y_n|\boldsymbol{x})}\right)\right|$$

$$(11)\qquad =O(h_n^3\log^3(y_n))=o(h_n^2\log^2(y_n)).$$

By condition $\mathcal{D}_\omega$,

$$(12)\qquad \limsup_{y\to\infty}\sup_{\|\boldsymbol{u}\|\le 1}\frac{1}{\log(y)}\left|\log\left(\frac{\overline{F}(y|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y|\boldsymbol{x})}\right)+h_n\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x})\right|=O(h_n^2).$$

Since $\nabla_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x})/\log(y)$ is bounded for y large enough by condition $\mathcal{D}_\omega$, we deduce from this equation that

$$(13)\qquad \limsup_{y\to\infty}\sup_{\|\boldsymbol{u}\|\le 1}\frac{1}{\log^2(y)}\left|\log^2\left(\frac{\overline{F}(y|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y|\boldsymbol{x})}\right)-h_n^2(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x}))^2\right|=o(h_n^2).$$

Combine (11) with (12) and (13), to obtain

$$\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{\overline{F}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(y_n|\boldsymbol{x})}-1+h_n\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(y_n|\boldsymbol{x})-\frac{h_n^2}{2}(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(y_n|\boldsymbol{x}))^2\right|$$

$$(14)\qquad =O(h_n^2\log(y_n))+o(h_n^2\log^2(y_n))=o(h_n^2\log^2(y_n)).$$

Combining (5), (6), (9) and (14) and the identity $\int_{\|\boldsymbol{u}\|\le 1}K^b(\boldsymbol{u})\boldsymbol{u}\,d\boldsymbol{u}=\mathbf{0}_p$ results in

$$\mathbb{E}\left(\mathbb{1}_{\{Y>y_n\}}K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right)=h_n^p\psi^{(0)}(y_n|\boldsymbol{x})\left(\int_{\mathbb{R}^p}K^b\right)$$

$$\times\left(1+\frac{h_n^2}{2}\int_{\|\boldsymbol{u}\|\le 1}\frac{K^b(\boldsymbol{u})}{\int_{\mathbb{R}^p}K^b}(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(y_n|\boldsymbol{x}))^2\,d\boldsymbol{u}+o(h_n^2\log^2(y_n))\right).$$

The result immediately follows, in the case $a=0$, by definition of $\boldsymbol{\mu}(\boldsymbol{x})$ and our assumptions on K. When $a>0$, an integration by parts yields

$$\frac{\varphi^{(a)}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})}-1=\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\left(\frac{\overline{F}(z|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(z|\boldsymbol{x})}-1\right)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}.$$

It then follows from this identity combined with (10) and Lemma A.5 that, for n large enough,

$$\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{\varphi^{(a)}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})}-1-\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\log\left(\frac{\overline{F}(z|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(z|\boldsymbol{x})}\right)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right.$$
$$\left.-\frac{1}{2}\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\log^2\left(\frac{\overline{F}(z|\boldsymbol{x}-h_n\boldsymbol{u})}{\overline{F}(z|\boldsymbol{x})}\right)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right|$$
$$= O(\omega_{h_n}^3(y_n|\boldsymbol{x})\log^3(y_n)) = O(h_n^3\log^3(y_n)) = o(h_n^2\log^2(y_n)). \tag{15}$$

[Lemma A.5 ensures that the integrals within the supremum are well-defined.] By condition $\mathcal{D}_\omega$ and Lemma A.5,

$$\forall s>0,\ \int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\|\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x})\|^s\log(z)\,dz<\infty$$

for n large enough. Combine (12), (13) and (15) to get

$$\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{\varphi^{(a)}(y_n|\boldsymbol{x}-h_n\boldsymbol{u})}{\varphi^{(a)}(y_n|\boldsymbol{x})}-1+h_n\boldsymbol{u}^\top\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x})\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right.$$
$$\left.-\frac{h_n^2}{2}\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x}))^2\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right|=o(h_n^2\log^2(y_n)). \tag{16}$$

Combine now (5), (6), (9) and (16) with the identity $\int_{\|\boldsymbol{u}\|\le 1}K^b(\boldsymbol{u})\boldsymbol{u}\,d\boldsymbol{u}=\mathbf{0}_p$ to find

$$\mathbb{E}\left((Y-y_n)^a\mathbb{1}_{\{Y>y_n\}}K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right)=h_n^p\psi^{(a)}(y_n|\boldsymbol{x})\left(\int_{\mathbb{R}^p}K^b\right)$$
$$\times\left(1+\frac{h_n^2}{2}\int_{\|\boldsymbol{u}\|\le 1}\frac{K^b(\boldsymbol{u})}{\int_{\mathbb{R}^p}K^b}\left(\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x}))^2\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right)d\boldsymbol{u}+o(h_n^2\log^2(y_n))\right).$$

The final step is to write

$$\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{1}{\log^2(y_n)}\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x}))^2\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}-(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\right|$$
$$\le\frac{1}{\log^2(y_n)}\sup_{\|\boldsymbol{u}\|\le 1}\left|\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})[(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\log\overline{F}(z|\boldsymbol{x}))^2-(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\log^2(z)]\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right|$$
$$+\sup_{\|\boldsymbol{u}\|\le 1}(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\times\left|\frac{1}{\log^2(y_n)}\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\log^2(z)\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}-1\right|$$
$$\le\sup_{\substack{\|\boldsymbol{u}\|\le 1\\ z\ge y_n}}\left|\left(\boldsymbol{u}^\top\nabla_{\boldsymbol{x}}\frac{\log\overline{F}(z|\boldsymbol{x})}{\log(z)}\right)^2-(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\right|\times\frac{1}{\log^2(y_n)}\left|\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\log^2(z)\,dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})\,dz}\right|$$
$$+o(1)=o(1)$$

by condition $\mathcal{D}_\omega$, the Cauchy-Schwarz inequality, Lemma A.5 and Karamata's theorem [see Theorem B.1.5 in 19, p.363]. Hence the asymptotic expansion

$$\mathbb{E}\left((Y-y_n)^a \mathbb{1}_{\{Y>y_n\}} K^b\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\right) = h_n^p \psi^{(a)}(y_n|\boldsymbol{x}) \left(\int_{\mathbb{R}^p} K^b\right)$$
$$\times\left(1+\frac{h_n^2\log^2(y_n)}{2}\int_{\|\boldsymbol{u}\|\le 1}\frac{K^b(\boldsymbol{u})}{\int_{\mathbb{R}^p}K^b}(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\,d\boldsymbol{u}+o(h_n^2\log^2(y_n))\right)$$

as required. □

The following lemma is the key to the control of the empirical smoothed conditional tail probabilities and moments. It makes use of the fact that, if $z>1$ then, by definition of $\omega_h(z|\boldsymbol{x})$ and using the mean value theorem, one has

$$(17)\qquad \forall y\ge z,\ \sup_{\boldsymbol{x}'\in B(\boldsymbol{x},h)}\left|\frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})}-1\right| \le \omega_h(z|\boldsymbol{x})\log(y)\times y^{\omega_h(z|\boldsymbol{x})}.$$

LEMMA A.7. *Assume that conditions $\mathcal{M}$, $\mathcal{A}(l_n,r_n)$, $\mathcal{K}$, $\mathcal{L}_g$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ hold. Suppose that $y_n\to\infty$ and $h_n\to 0$ as $n\to\infty$ are such that $nh_n^p\overline{F}(y_n|\boldsymbol{x})\to\infty$, $h_n\log(y_n)\to 0$ and $r_nh_n^p\to 0$.*

(i) *Suppose that $y_{n,j}=c_j^{-\gamma(\boldsymbol{x})}y_n(1+o(1))$ for all $j\in\{1,\dots,J\}$ with $0<c_1<c_2<\dots<c_J\le 1$. Assume that there exists $\delta>0$ with $r_n^{1+\delta}/[nh_n^p\overline{F}(y_n|\boldsymbol{x})]^{\delta/2}\to 0$. Then*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(0)}(y_{n,j}|\boldsymbol{x})-\mathbb{E}(\widehat{\psi}_n^{(0)}(y_{n,j}|\boldsymbol{x}))}{\psi^{(0)}(y_{n,j}|\boldsymbol{x})}\right)_{1\le j\le J}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0}_J,\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\boldsymbol{M}\right),$$

where $\boldsymbol{M}$ is the symmetric matrix of size J having entries $M_{j,l}=c_l^{-1}$ (for $1\le j\le l\le J$).

(ii) *If moreover $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x})\log^2(y_n)\to 0$ then*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(0)}(y_{n,j}|\boldsymbol{x})}{\psi^{(0)}(y_{n,j}|\boldsymbol{x})}-1\right)_{1\le j\le J}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0}_J,\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\boldsymbol{M}\right).$$

If furthermore conditions $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_\omega$ hold then condition $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x})\log^2(y_n)\to 0$ may be replaced by the weaker bias assumption $\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\times h_n^2\log^2(y_n)\to c(\boldsymbol{x})\in[0,\infty)$, in which case

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(0)}(y_{n,j}|\boldsymbol{x})}{\psi^{(0)}(y_{n,j}|\boldsymbol{x})}-1\right)_{1\le j\le J}$$
$$\xrightarrow{d}\mathcal{N}\left(\left(\frac{c(\boldsymbol{x})}{2}\int_{\mathbb{R}^p}K(\boldsymbol{u})(\boldsymbol{u}^\top\boldsymbol{\mu}(\boldsymbol{x}))^2\,d\boldsymbol{u}\right)\boldsymbol{1}_J,\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\boldsymbol{M}\right)$$

where $\boldsymbol{1}_J$ is the column vector in $\mathbb{R}^J$ having all entries equal to 1.

In (i) and (ii), if $((\boldsymbol{X}_t,Y_t))_{t\ge 1}$ is in fact $\psi-$mixing with $\sum_{j=1}^\infty\psi(j)<\infty$, then all conditions on (l_n) and (r_n) (including condition $\mathcal{A}(l_n,r_n)$), conditions $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ can be dropped.

(iii) *Assume that there exists $\delta>0$ with $\gamma(\boldsymbol{x})\in(0,1/(2+\delta))$ and $r_n^{1+\delta}/[nh_n^p\overline{F}(y_n|\boldsymbol{x})]^{\delta/2}\to 0$. Let $z_n=\theta y_n(1+o(1))$ (for a certain $\theta>0$). Then*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x})-\mathbb{E}(\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x}))}{\psi^{(1)}(y_n|\boldsymbol{x})},\frac{\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x})-\mathbb{E}(\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x}))}{\psi^{(0)}(z_n|\boldsymbol{x})}\right)$$

$$\xrightarrow{d} \mathcal{N}\left((0,0), \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{V}(\boldsymbol{x})\right),$$

where $\boldsymbol{V}(\boldsymbol{x})$ *is a* 2×2 *symmetric matrix having entries*

$$\boldsymbol{V}_{1,1}(\boldsymbol{x}) = 2\frac{1-\gamma(\boldsymbol{x})}{1-2\gamma(\boldsymbol{x})},\ \boldsymbol{V}_{2,2}(\boldsymbol{x}) = \theta^{1/\gamma(\boldsymbol{x})} \text{ and } \boldsymbol{V}_{1,2}(\boldsymbol{x}) = \begin{cases} \dfrac{\theta+\gamma(\boldsymbol{x})-1}{\gamma(\boldsymbol{x})} & \text{if } \theta \geq 1, \\ \theta^{1/\gamma(\boldsymbol{x})} & \text{if } \theta < 1. \end{cases}$$

(iv) *If moreover* $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x})\log^2(y_n) \to 0$ *then*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x})}{\psi^{(1)}(y_n|\boldsymbol{x})}-1, \frac{\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x})}{\psi^{(0)}(z_n|\boldsymbol{x})}-1\right) \xrightarrow{d} \mathcal{N}\left((0,0), \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{V}(\boldsymbol{x})\right).$$

If furthermore conditions $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$ *and* $\mathcal{D}_\omega$ *hold then condition* $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x})\log^2(y_n) \to 0$ *may be replaced by the weaker bias assumption* $\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})} \times h_n^2\log^2(y_n) \to c(\boldsymbol{x}) \in [0,\infty)$*, in which case*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x})}{\psi^{(1)}(y_n|\boldsymbol{x})}-1, \frac{\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x})}{\psi^{(0)}(z_n|\boldsymbol{x})}-1\right)$$
$$\xrightarrow{d} \mathcal{N}\left(\left(\frac{c(\boldsymbol{x})}{2}\int_{\mathbb{R}^p} K(\boldsymbol{u})(\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2\, d\boldsymbol{u}\right) \times (1,1), \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{V}(\boldsymbol{x})\right).$$

In (iii) and (iv), if $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ *is in fact* $\psi-$*mixing with* $\sum_{j=1}^\infty \psi(j) < \infty$*, then all conditions on* (l_n) *and* (r_n) *(including condition* $\mathcal{A}(l_n, r_n)$*), conditions* $\mathcal{B}_p$ *and* $\mathcal{B}_\Omega$ *can be dropped. The condition on* $\gamma(\boldsymbol{x})$ *becomes* $0 < \gamma(\boldsymbol{x}) < 1/2$.

Note that condition $r_n h_n^p \to 0$ is a consequence of $r_n^{1+\delta}/[nh_n^p\overline{F}(y_n|\boldsymbol{x})]^{\delta/2} \to 0$ (for some $\delta > 0$) and $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \to 0$, which is why this condition will not appear in Lemma A.9 when the latter two assumptions hold. Note also that, in (i) and (ii), condition $r_n^{1+\delta}/[nh_n^p\overline{F}(y_n|\boldsymbol{x})]^{\delta/2} \to 0$ for some $\delta > 0$ is equivalent to assuming that $r_n/[nh_n^p\overline{F}(y_n|\boldsymbol{x})]^{1/2-\delta} \to 0$ for some $\delta > 0$. The latter can be viewed as a conditional version of the condition $r_n/[n\overline{F}(y_n)]^{1/2-\delta} \to 0$ used in [42].

PROOF. We prove (iii) and (iv), which are similar to, but more difficult than, (i) and (ii). If (iii) holds then (iv) is a direct consequence of Lemma A.6(ii) and (iii). We therefore concentrate on proving (iii). Pick $\beta_0, \beta_1 \in \mathbb{R}$. Using the Cramér-Wold device, it is enough to analyze the asymptotic behavior of

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left\{\beta_1\left(\frac{\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x}) - \mathbb{E}(\widehat{\psi}_n^{(1)}(y_n|\boldsymbol{x}))}{\psi^{(1)}(y_n|\boldsymbol{x})}\right) + \beta_0\left(\frac{\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x}) - \mathbb{E}(\widehat{\psi}_n^{(0)}(z_n|\boldsymbol{x}))}{\psi^{(0)}(z_n|\boldsymbol{x})}\right)\right\}$$
$$= \sum_{i=1}^n \sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} L_{i,n}$$

where the centered random variable $L_{i,n}$ is given as

$$L_{i,n} = \beta_1 \frac{(Y_i - y_n)\mathbb{1}_{\{Y_i > y_n\}}K((\boldsymbol{x}-\boldsymbol{X}_i)/h_n) - \mathbb{E}((Y-y_n)\mathbb{1}_{\{Y>y_n\}}K((\boldsymbol{x}-\boldsymbol{X})/h_n))}{\psi^{(1)}(y_n|\boldsymbol{x})}$$
$$+ \beta_0 \frac{\mathbb{1}_{\{Y_i > z_n\}}K((\boldsymbol{x}-\boldsymbol{X}_i)/h_n) - \mathbb{E}(\mathbb{1}_{\{Y>z_n\}}K((\boldsymbol{x}-\boldsymbol{X})/h_n))}{\psi^{(0)}(z_n|\boldsymbol{x})}.$$

The aim is to use a big-block/small-block argument to obtain the desired convergence. To do so we apply Lemma C.7(ii) in [18], which requires to analyze the asymptotic behavior of $\mathrm{Var}(L_{1,n})$ and $\mathrm{Cov}(L_{i,n}, L_{j,n})$. Combine Lemma A.6(i) and A.6(ii), the regular variation property of $\overline{F}(\cdot|\boldsymbol{x})$ and straightforward calculations to get

$$\mathrm{Var}(L_{1,n}) = \frac{h_n^p}{\overline{F}(y_n|\boldsymbol{x})} \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \left[\beta_1^2 \times \frac{2(1-\gamma(\boldsymbol{x}))}{1-2\gamma(\boldsymbol{x})} + \beta_0^2 \times \theta^{1/\gamma(\boldsymbol{x})} \right.$$
$$\left. + 2\beta_0\beta_1 \times \frac{\gamma(\boldsymbol{x})\theta^{1/\gamma(\boldsymbol{x})}(\max(1,\theta))^{1-1/\gamma(\boldsymbol{x})} + (1-\gamma(\boldsymbol{x}))(\max(1,\theta)-1)}{\gamma(\boldsymbol{x})}\right](1+o(1))$$

(18)
$$= \frac{h_n^p}{\overline{F}(y_n|\boldsymbol{x})} \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \left[\beta_1^2 \boldsymbol{V}_{1,1}(\boldsymbol{x}) + \beta_0^2 \boldsymbol{V}_{2,2}(\boldsymbol{x}) + 2\beta_0\beta_1 \boldsymbol{V}_{1,2}(\boldsymbol{x})\right](1+o(1)).$$

We turn to the calculation of $\mathrm{Cov}(L_{i,n}, L_{j,n}) = \beta_1^2 C_{i,j,n}^{(1)} + \beta_0^2 C_{i,j,n}^{(2)} + \beta_0\beta_1 (C_{i,j,n}^{(3)} + C_{i,j,n}^{(4)})$, with

$$C_{i,j,n}^{(1)} = \frac{\mathrm{Cov}\left((Y_i - y_n)\mathbb{1}_{\{Y_i > y_n\}} K((\boldsymbol{x} - \boldsymbol{X}_i)/h_n), (Y_j - y_n)\mathbb{1}_{\{Y_j > y_n\}} K((\boldsymbol{x} - \boldsymbol{X}_j)/h_n)\right)}{[\psi^{(1)}(y_n|\boldsymbol{x})]^2},$$
$$C_{i,j,n}^{(2)} = \frac{\mathrm{Cov}\left(\mathbb{1}_{\{Y_i > z_n\}} K((\boldsymbol{x} - \boldsymbol{X}_i)/h_n), \mathbb{1}_{\{Y_j > z_n\}} K((\boldsymbol{x} - \boldsymbol{X}_j)/h_n)\right)}{[\psi^{(0)}(z_n|\boldsymbol{x})]^2},$$
$$C_{i,j,n}^{(3)} = \frac{\mathrm{Cov}\left((Y_i - y_n)\mathbb{1}_{\{Y_i > y_n\}} K((\boldsymbol{x} - \boldsymbol{X}_i)/h_n), \mathbb{1}_{\{Y_j > z_n\}} K((\boldsymbol{x} - \boldsymbol{X}_j)/h_n)\right)}{\psi^{(1)}(y_n|\boldsymbol{x})\psi^{(0)}(z_n|\boldsymbol{x})},$$
$$\text{and } C_{i,j,n}^{(4)} = \frac{\mathrm{Cov}\left(\mathbb{1}_{\{Y_i > z_n\}} K((\boldsymbol{x} - \boldsymbol{X}_i)/h_n), (Y_j - y_n)\mathbb{1}_{\{Y_j > y_n\}} K((\boldsymbol{x} - \boldsymbol{X}_j)/h_n)\right)}{\psi^{(1)}(y_n|\boldsymbol{x})\psi^{(0)}(z_n|\boldsymbol{x})}.$$

We bound each of the $C_{i,j,n}^{(l)}$, for $l = 1,2,3,4$. We clearly have, by Lemma A.6(ii),

$$C_{i,j,n}^{(1)} = \frac{1}{[\psi^{(1)}(y_n|\boldsymbol{x})]^2} \mathbb{E}(\mathbb{E}((Y_i - y_n)_+ (Y_j - y_n)_+ | \boldsymbol{X}_i, \boldsymbol{X}_j) K((\boldsymbol{x} - \boldsymbol{X}_i)/h_n) K((\boldsymbol{x} - \boldsymbol{X}_j)/h_n))$$
$$+ h_n^{2p} R_n$$
$$\leq \frac{1}{[\psi^{(1)}(y_n|\boldsymbol{x})]^2} \sup_{\|\boldsymbol{u}_i\|, \|\boldsymbol{u}_j\| \leq 1} \mathbb{E}((Y_i - y_n)_+ (Y_j - y_n)_+ | \boldsymbol{X}_i = \boldsymbol{x} - \boldsymbol{u}_i h_n, \boldsymbol{X}_j = \boldsymbol{x} - \boldsymbol{u}_j h_n)$$
$$\times \left(\sup_{\mathbb{R}^p} K^2\right) \mathbb{P}(\boldsymbol{X}_i \in B(\boldsymbol{x}, h_n), \boldsymbol{X}_j \in B(\boldsymbol{x}, h_n)) + h_n^{2p} R_n$$

where R_n is bounded and independent of i and j. [We use this notation throughout without further mention, with a value of R_n that can change from line to line.] An integration by parts and a change of variables produce

$$\mathbb{E}((Y_i - y_n)_+ (Y_j - y_n)_+ | \boldsymbol{X}_i = \boldsymbol{x} - \boldsymbol{u}_i h_n, \boldsymbol{X}_j = \boldsymbol{x} - \boldsymbol{u}_j h_n)$$
$$= y_n^2 \int_1^\infty \int_1^\infty \mathbb{P}\left(Y_i > a y_n, Y_j > b y_n | \boldsymbol{X}_i = \boldsymbol{x} - \boldsymbol{u}_i h_n, \boldsymbol{X}_j = \boldsymbol{x} - \boldsymbol{u}_j h_n\right) da\, db.$$

Using assumption $\mathcal{B}_\Omega$, we find

$$\sup_{i \neq j} \sup_{\|\boldsymbol{u}_i\|, \|\boldsymbol{u}_j\| \leq 1} \int_1^\infty \int_1^\infty \mathbb{P}\left(Y_i > a y_n, Y_j > b y_n | \boldsymbol{X}_i = \boldsymbol{x} - \boldsymbol{u}_i h_n, \boldsymbol{X}_j = \boldsymbol{x} - \boldsymbol{u}_j h_n\right) da\, db$$

$$= O\left(\left[\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},h_n)}\int_1^\infty \sqrt{\overline{F}(ay_n|\boldsymbol{x}')}\,da\right]^2\right).$$

It then follows from Equation (17) that

$$\sup_{i\neq j}\sup_{\|\boldsymbol{u}_i\|,\|\boldsymbol{u}_j\|\leq 1}\int_1^\infty\int_1^\infty \mathbb{P}\left(Y_i>ay_n, Y_j>by_n|\boldsymbol{X}_i=\boldsymbol{x}-\boldsymbol{u}_ih_n, \boldsymbol{X}_j=\boldsymbol{x}-\boldsymbol{u}_jh_n\right)da\,db$$
$$= O\left(\overline{F}(y_n|\boldsymbol{x})\left[\int_1^\infty \sqrt{\frac{\overline{F}(ay_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}}\left(1+\sqrt{\omega_{h_n}(y_n|\boldsymbol{x})\log(ay_n)\times(ay_n)^{\omega_{h_n}(y_n|\boldsymbol{x})}}\right)da\right]^2\right).$$

Since $\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n) = O(h_n\log(y_n))\to 0$ by assumption $\mathcal{L}_\omega$, we find

$$\sup_{i\neq j}\sup_{\|\boldsymbol{u}_i\|,\|\boldsymbol{u}_j\|\leq 1}\int_1^\infty\int_1^\infty \mathbb{P}\left(Y_i>ay_n, Y_j>by_n|\boldsymbol{X}_i=\boldsymbol{x}-\boldsymbol{u}_ih_n, \boldsymbol{X}_j=\boldsymbol{x}-\boldsymbol{u}_jh_n\right)da\,db$$
$$= O\left(\overline{F}(y_n|\boldsymbol{x})\left[\int_1^\infty \sqrt{\frac{\overline{F}(ay_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}}a^\varepsilon\,da\right]^2\right)\quad\text{for any } \varepsilon>0.$$

Potter bounds [see Proposition B.1.9.5 p.367 in 19] and the fact that $\gamma(\boldsymbol{x})<1/2$ entail

$$\sup_{i\neq j}\sup_{\|\boldsymbol{u}_i\|,\|\boldsymbol{u}_j\|\leq 1}\int_1^\infty\int_1^\infty \mathbb{P}\left(Y_i>ay_n, Y_j>by_n|\boldsymbol{X}_i=\boldsymbol{x}-\boldsymbol{u}_ih_n, \boldsymbol{X}_j=\boldsymbol{x}-\boldsymbol{u}_jh_n\right)da\,db$$
$$= O(\overline{F}(y_n|\boldsymbol{x})).$$

Conclude, using condition $\mathcal{B}_p$ and again Lemma A.6(i), that there is a positive constant D (whose value may change from line to line) such that for n large enough and $i\neq j$,

$$|C_{i,j,n}^{(1)}| \leq D\frac{1}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{P}(\boldsymbol{X}_i\in B(\boldsymbol{x},h_n),\boldsymbol{X}_j\in B(\boldsymbol{x},h_n)) + h_n^{2p}R_n$$
$$\leq D\left(\frac{h_n^p\varepsilon_n}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|<t_0\}}+\frac{h_n^{2p}}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|\geq t_0\}}\right)$$

where $\varepsilon_n\to 0$ as $n\to\infty$. Then, combining the equality $z_n=\theta y_n(1+o(1))$, conditions $\mathcal{L}_\omega$, $\mathcal{B}_p$, $\mathcal{B}_\Omega$ and $h_n\log(y_n)\to 0$, Lemma A.6(i), Equation (17) and the regular variation property of $\overline{F}(\cdot|\boldsymbol{x})$, we find, for n large enough and $i\neq j$, that similarly as above,

$$|C_{i,j,n}^{(2)}| \leq D\left(\frac{h_n^p\varepsilon_n}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|<t_0\}}+\frac{h_n^{2p}}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|\geq t_0\}}\right).$$

The same arguments lead to identical bounds for the $C_{i,j,n}^{(3)}$ and $C_{i,j,n}^{(4)}$, and we may then conclude that, for n large enough and $i\neq j$,

$$|\operatorname{Cov}(L_{i,n},L_{j,n})| \leq D\left(\frac{h_n^p\varepsilon_n}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|<t_0\}}+\frac{h_n^{2p}}{\overline{F}(y_n|\boldsymbol{x})}\mathbb{1}_{\{|i-j|\geq t_0\}}\right).$$

As a consequence, for any sequence (u_n) tending to infinity and such that $u_n\leq n$,

$$\left|\sum_{1\leq i<j\leq u_n}\operatorname{Cov}(L_{i,n},L_{j,n})\right| = O\left(u_n\frac{h_n^p\varepsilon_n}{\overline{F}(y_n|\boldsymbol{x})}+u_n^2\frac{h_n^{2p}}{\overline{F}(y_n|\boldsymbol{x})}\right). \tag{19}$$

We may now develop our big-block/small-block argument (with l_n being the size of small blocks and $r_n - l_n$ being the size of big blocks) via Lemma C.7(ii) in [18], applied to the random sum

$$S_n = \sum_{i=1}^{n} \sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} L_{i,n}.$$

Using Equations (18) and (19), it is immediate that

$$\begin{aligned}
&\frac{n}{r_n} \operatorname{Var}\left(\sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} \sum_{j=1}^{l_n} L_{j,n} \right) \\
&= \frac{n}{r_n} \frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p} l_n \operatorname{Var}(L_{1,n}) + 2 \frac{n}{r_n} \frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p} \sum_{1 \le i < j \le l_n} \operatorname{Cov}(L_{i,n}, L_{j,n}) \\
&= O(l_n/r_n) + O(l_n \varepsilon_n / r_n) + O(l_n h_n^p \times l_n / r_n) = o(1)
\end{aligned}$$

because $l_n = o(r_n)$ and $r_n h_n^p \to 0$. Moreover, since $r_n = o(n)$ and $n - r_n \lfloor n/r_n \rfloor \le r_n$,

$$\begin{aligned}
&\operatorname{Var}\left(\sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} \sum_{j=1}^{n - r_n \lfloor n/r_n \rfloor} L_{j,n} \right) \\
&= \frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p} (n - r_n \lfloor n/r_n \rfloor) \operatorname{Var}(L_{1,n}) + 2 \frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p} \sum_{1 \le i < j \le n - r_n \lfloor n/r_n \rfloor} \operatorname{Cov}(L_{i,n}, L_{j,n}) \\
&= O(r_n/n) + O(r_n \varepsilon_n / n) + O(r_n h_n^p \times r_n / n) = o(1).
\end{aligned}$$

Then, using Equations (18) and (19) again,

$$\begin{aligned}
&\frac{n}{r_n} \operatorname{Var}\left(\sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} \sum_{j=1}^{r_n} L_{j,n} \right) \\
&= \frac{\overline{F}(y_n|\boldsymbol{x})}{h_n^p} \operatorname{Var}(L_{1,n}) + 2 \frac{\overline{F}(y_n|\boldsymbol{x})}{r_n h_n^p} \sum_{1 \le i < j \le r_n} \operatorname{Cov}(L_{i,n}, L_{j,n}) \\
&= \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \left[\beta_1^2 \boldsymbol{V}_{1,1}(\boldsymbol{x}) + \beta_0^2 \boldsymbol{V}_{2,2}(\boldsymbol{x}) + 2\beta_0\beta_1 \boldsymbol{V}_{1,2}(\boldsymbol{x}) \right] (1 + o(1)) + O(\varepsilon_n) + O(r_n h_n^p) \\
&\to \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix}^\top \boldsymbol{V}(\boldsymbol{x}) \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix}.
\end{aligned}$$

Finally, it is immediately found using the Hölder inequality and Lemma A.6(ii) that

$$\frac{n}{r_n} \mathbb{E}\left(\left(\sqrt{\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}} \right)^{2+\delta} \left| \sum_{j=1}^{r_n} L_{j,n} \right|^{2+\delta} \right) = O\left(r_n \left(\frac{r_n}{\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})}} \right)^{\delta} \right) = o(1).$$

It follows that the Lindeberg condition in Lemma C.7(ii) is satisfied. Applying Lemma C.7(ii) in [18] completes the proof.

The $\psi-$mixing case follows by combining the covariance inequality in Lemma 3.1 of [39] and Lemma A.6(ii) to get

$$\max_{1 \le l \le 4} \max_{i \ne j} |C_{i,j,n}^{(l)}| \le 2\psi(j-i) h_n^{2p}.$$

Use then the central limit theorem of [47] with $j_n = 1$ and $k_n = n$ (with the notation therein) to conclude. We omit the details. □

The next lemma is a technical result that allows to combine Proposition A.1 and Lemma A.7 under condition $\mathcal{H}_\delta$.

LEMMA A.8. *Assume that*

$$\forall y > 0, \ \lim_{t\to\infty} \frac{\overline{F}(ty|\boldsymbol{x})}{\overline{F}(t|\boldsymbol{x})} = y^{-1/\gamma(\boldsymbol{x})}$$

for some $0 < \gamma(\boldsymbol{x}) < 1/a$*, where* $a > 0$*. Assume further that* $\omega_{h_n}(y_n|\boldsymbol{x}) = o(1)$ *for some sequences* $y_n \to \infty$ *and* $h_n \to 0$*. Then there exists* $r > 0$ *such that*

$$\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \mathbb{E}(Y_+^a|\boldsymbol{X} = \boldsymbol{x}') < \infty.$$

PROOF. An integration by parts yields

$$\mathbb{E}(Y_+^a|\boldsymbol{X} = \boldsymbol{x}') - \mathbb{E}(Y_+^a|\boldsymbol{X} = \boldsymbol{x}) = \int_0^\infty at^{a-1}\left\{\frac{\overline{F}(t|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x})} - 1\right\}\overline{F}(t|\boldsymbol{x})\,dt.$$

The conditional moment $\mathbb{E}(Y_+^a|\boldsymbol{X} = \boldsymbol{x})$ is finite because $\gamma(\boldsymbol{x}) < 1/a$, see Exercise 1.16 p.35 in [19]. It therefore suffices to work on the right-hand side in the above identity. Since

$$\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \left|\int_0^y at^{a-1}\left\{\frac{\overline{F}(t|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x})} - 1\right\}\overline{F}(t|\boldsymbol{x})\,dt\right| \le 2y^a < \infty$$

for any $r, y > 0$, it is in fact sufficient to show that

$$\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \left|\int_y^\infty at^{a-1}\left\{\frac{\overline{F}(t|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x})} - 1\right\}\overline{F}(t|\boldsymbol{x})\,dt\right|$$

is finite *for some* $r, y > 0$. Now, according to Equation (17), for any $\boldsymbol{x}' \in B(\boldsymbol{x}, h_n)$,

$$\begin{aligned}&\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},h_n)} \left|\int_{y_n}^\infty at^{a-1}\left\{\frac{\overline{F}(t|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x})} - 1\right\}\overline{F}(t|\boldsymbol{x})\,dt\right| \\ &\le a\omega_{h_n}(y_n|\boldsymbol{x})\int_{y_n}^\infty t^{a+\omega_{h_n}(y_n|\boldsymbol{x})-1}|\log(t)|\overline{F}(t|\boldsymbol{x})\,dt.\end{aligned}$$

Let $\varepsilon > 0$ be so small that $1/\gamma(\boldsymbol{x}) > a+\varepsilon$, and n_0 be an integer so large that $\omega_{h_n}(y_n|\boldsymbol{x}) \le \varepsilon/2$ and $y_n \ge 1$ for $n \ge n_0$. On the interval $[1,\infty)$, $t^{-\varepsilon/2}\log t$ is bounded, so there is a constant $C > 0$ with

$$\begin{aligned}\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},h_n)} \left|\int_{y_n}^\infty at^{a-1}\left\{\frac{\overline{F}(t|\boldsymbol{x}')}{\overline{F}(t|\boldsymbol{x})} - 1\right\}\overline{F}(t|\boldsymbol{x})\,dt\right| &\le C\int_{y_n}^\infty (a+\varepsilon)t^{a+\varepsilon-1}\overline{F}(t|\boldsymbol{x})\,dt \\ &\le C\,\mathbb{E}(Y_+^{a+\varepsilon}|\boldsymbol{X} = \boldsymbol{x}) < \infty\end{aligned}$$

as soon as $n \ge n_0$. Take $y = y_{n_0}$ and $r = h_{n_0}$ to complete the proof. □

Lemma A.9 gives the required joint convergence result between the empirical smoothed estimators of $\overline{E}(\cdot|\boldsymbol{x})$ and $\overline{F}(\cdot|\boldsymbol{x})$ at intermediate levels that will guarantee the joint convergence of empirical smoothed quantiles and expectiles at such levels. Recall that $\widehat{\overline{F}}_n(\cdot|\boldsymbol{x}) = 1 - \widehat{F}_n(\cdot|\boldsymbol{x})$ and $\widehat{\overline{E}}_n(\cdot|\boldsymbol{x}) = 1 - \widehat{E}_n(\cdot|\boldsymbol{x})$.

LEMMA A.9. *Assume that conditions $\mathcal{M}$, $\mathcal{A}(l_n, r_n)$, $\mathcal{K}$, $\mathcal{L}_g$, $\mathcal{L}_\omega$, $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ hold. Suppose that $y_n \to \infty$ and $h_n \to 0$ as $n \to \infty$. Assume further that $nh_n^p \overline{F}(y_n|\boldsymbol{x}) \to \infty$.*

(i) *Suppose that $y_{n,j} = c_j^{-\gamma(\boldsymbol{x})} y_n(1 + o(1))$ for all $j \in \{1, \ldots, J\}$ with $0 < c_1 < c_2 < \ldots < c_J \leq 1$. Assume that $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \log^2(y_n) \to 0$, and that there exist $\delta > 0$ with $r_n^{1+\delta}/[nh_n^p \overline{F}(y_n|\boldsymbol{x})]^{\delta/2} \to 0$ and $\eta > 1$ with $\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$. Then*

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\widehat{\overline{F}}_n(y_{n,j}|\boldsymbol{x})}{\overline{F}(y_{n,j}|\boldsymbol{x})} - 1 \right)_{1 \leq j \leq J} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}_J, \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{M} \right)$$

with the notation of Lemma A.7(i). If furthermore conditions $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_\omega$ hold then condition $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \log^2(y_n) \to 0$ may be replaced by the weaker bias assumption $\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \times h_n^2 \log^2(y_n) \to c(\boldsymbol{x}) \in [0, \infty)$, in which case, provided $r_n h_n^p \to 0$,

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\widehat{\overline{F}}_n(y_{n,j}|\boldsymbol{x})}{\overline{F}(y_{n,j}|\boldsymbol{x})} - 1 \right)_{1 \leq j \leq J}$$
$$\xrightarrow{d} \mathcal{N}\left(\left(\frac{c(\boldsymbol{x})}{2} \int_{\mathbb{R}^p} K(\boldsymbol{u})(\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2 \, d\boldsymbol{u} \right) \boldsymbol{1}_J, \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{M} \right).$$

If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is moreover $\rho-$mixing, then condition $\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$ may be replaced by $\sum_{j=1}^{\infty} \rho(j) < \infty$.

If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is in fact also $\psi-$mixing in addition to $\rho-$mixing, with $\sum_{j=1}^{\infty} \psi(j) < \infty$ (instead of $\sum_{j=1}^{\infty} j^\eta \alpha(j) < \infty$ for some $\eta > 1$, or $\sum_{j=1}^{\infty} \rho(j) < \infty$), then all conditions on (l_n) and (r_n) (including condition $\mathcal{A}(l_n, r_n)$), conditions $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ can also be dropped.

(ii) *Assume that conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ hold. Suppose that $z_n = \theta y_n(1 + o(1))$ (for a certain $\theta > 0$), that $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \log^2(y_n) \to 0$, that $r_n^{1+\delta}/[nh_n^p \overline{F}(y_n|\boldsymbol{x})]^{\delta/2} \to 0$ as $n \to \infty$, and $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$. Then*

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\widehat{\overline{E}}_n(y_n|\boldsymbol{x})}{\overline{E}(y_n|\boldsymbol{x})} - 1, \frac{\widehat{\overline{F}}_n(z_n|\boldsymbol{x})}{\overline{F}(z_n|\boldsymbol{x})} - 1 \right) \xrightarrow{d} \mathcal{N}\left((0, 0), \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{V}(\boldsymbol{x}) \right)$$

with the notation of Lemma A.7(iii). If furthermore conditions $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$ and $\mathcal{D}_\omega$ hold then condition $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \log^2(y_n) \to 0$ may be replaced by the weaker bias assumption $\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \times h_n^2 \log^2(y_n) \to c(\boldsymbol{x}) \in [0, \infty)$, in which case, provided $r_n h_n^p \to 0$,

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\widehat{\overline{E}}_n(y_n|\boldsymbol{x})}{\overline{E}(y_n|\boldsymbol{x})} - 1, \frac{\widehat{\overline{F}}_n(z_n|\boldsymbol{x})}{\overline{F}(z_n|\boldsymbol{x})} - 1 \right)$$
$$\xrightarrow{d} \mathcal{N}\left(\left(\frac{c(\boldsymbol{x})}{2} \int_{\mathbb{R}^p} K(\boldsymbol{u})(\boldsymbol{u}^\top \boldsymbol{\mu}(\boldsymbol{x}))^2 \, d\boldsymbol{u} \right) \times (1, 1), \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})} \boldsymbol{V}(\boldsymbol{x}) \right).$$

If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is moreover $\rho-$mixing with $\sum_{j=1}^{\infty} \rho(j) < \infty$ (instead of $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$), then condition $\mathcal{B}_m$ can be dropped and condition $\mathcal{H}_\delta$ can be replaced by $\gamma(\boldsymbol{x}) < 1/(2 + \delta)$.

If $((\boldsymbol{X}_t, Y_t))_{t \geq 1}$ is in fact also $\psi-$mixing in addition to $\rho-$mixing, with $\sum_{j=1}^{\infty} \psi(j) < \infty$ (instead of $\sum_{j=1}^{\infty} j^\eta [\alpha(j)]^{\delta/(2+\delta)} < \infty$ for some $\eta > \delta/(2 + \delta)$, or $\sum_{j=1}^{\infty} \rho(j) < \infty$), then all conditions on (l_n) and (r_n) (including condition $\mathcal{A}(l_n, r_n)$), conditions $\mathcal{B}_p$ and $\mathcal{B}_\Omega$ can also be dropped. Condition $\mathcal{H}_\delta$ can be replaced by $0 < \gamma(\boldsymbol{x}) < 1/2$.

PROOF. Again we prove (ii), which is similar to, but more difficult than, statement (i). Follow the proof of Lemma 7 in [27] but apply Proposition A.1 and Lemma A.7 rather than the Lemmas 4 and 6 of [27]. Note that Proposition A.1 applies because of Lemma A.8, which combined with condition $\mathcal{H}_\delta$ yields $\sup_{\boldsymbol{x}'\in B(\boldsymbol{x},r)} \mathbb{E}(|Y|^{2+\delta}|\boldsymbol{X}=\boldsymbol{x}') < \infty$ for some $r > 0$. □

The following result is a refinement of Lemma 8 in [27].

LEMMA A.10. *Assume that* $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ *holds. Suppose also that* $\gamma(\boldsymbol{x}) < 1$, $\rho(\boldsymbol{x}) < 0$ *and* $\mathbb{E}(Y_- \mid \boldsymbol{X}=\boldsymbol{x}) < \infty$. *Let* $\tau_n, \tau_n' \to 1$ *be such that* $(1-\tau_n')/(1-\tau_n) \to 0$. *Then:*

$$\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\gamma(\boldsymbol{x})} \frac{e(\tau_n|\boldsymbol{x})}{e(\tau_n'|\boldsymbol{x})} = 1 + \frac{\gamma(\boldsymbol{x})(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})}}{q(\tau_n|\boldsymbol{x})}(\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})+o(1)) + \frac{(1-\gamma(\boldsymbol{x}))(1/\gamma(\boldsymbol{x})-1)^{-\rho(\boldsymbol{x})}}{\rho(\boldsymbol{x})(1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x}))} A((1-\tau_n)^{-1}|\boldsymbol{x})(1+o(1)).$$

PROOF. Combine Proposition 1(i) in [17] with

$$\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\gamma(\boldsymbol{x})} \frac{q(\tau_n|\boldsymbol{x})}{q(\tau_n'|\boldsymbol{x})} = 1 + \frac{1}{\rho(\boldsymbol{x})} A((1-\tau_n)^{-1}|\boldsymbol{x})(1+o(1))$$

(from *e.g.* p.139 in [19]). We omit the straightforward calculation. □

A.5. Proofs of the main results. **Proof of Theorems 2.1 and 3.1.** The proof of Theorem 2.1 goes in exactly the same way as that of Theorem 3.1, so we only prove the latter, which is more difficult. Set $\sigma_n = 1/\sqrt{nh_n^p(1-\tau_n)}$, pick $\boldsymbol{z}=(z_1,z_2)$ and define

$$\Phi_n(\boldsymbol{z}) = \mathbb{P}\left(\left\{\sigma_n^{-1}\left(\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)\le z_1\right\}\cap\left\{\sigma_n^{-1}\left(\frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\right)\le z_2\right\}\right)$$

As in the proof of Theorem 1 in [27], we find, with $y_n = e(\tau_n|\boldsymbol{x})$, $y_n' = e(\tau_n|\boldsymbol{x})(1+z_1\sigma_n)$ and $z_n' = q(\beta_n|\boldsymbol{x})(1+z_2\sigma_n)$,

$$\Phi_n(\boldsymbol{z}) = \mathbb{P}\left(\left\{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\overline{E}}_n(y_n'|\boldsymbol{x})}{\overline{E}(y_n'|\boldsymbol{x})}-1\right)\le\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{E}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{E}(y_n'|\boldsymbol{x})}-1\right)\right\}\right.$$
$$\left.\cap\left\{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\overline{F}}_n(z_n'|\boldsymbol{x})}{\overline{F}(z_n'|\boldsymbol{x})}-1\right)\le\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{F}(q(\beta_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{F}(z_n'|\boldsymbol{x})}-1\right)\right\}\right).$$

Combine the local uniformity of condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ (see Lemma 5 in [17]) with assumption $A((1-\tau_n)^{-1}|\boldsymbol{x}) = O(\sigma_n)$, the asymptotic proportionality between $1-\beta_n$ and $1-\tau_n$, and the regular variation property of $A(\cdot|\boldsymbol{x})$ to find

$$\overline{F}(q(\beta_n|\boldsymbol{x})(1+z_2\sigma_n)|\boldsymbol{x}) = \overline{F}(q(\beta_n|\boldsymbol{x})|\boldsymbol{x})\left(1-\frac{z_2\sigma_n}{\gamma(\boldsymbol{x})}(1+o(1))\right).$$

Recall Lemma A.3(iv) in [46] which, for $p=2$ and applied to the conditional distribution of Y given $\boldsymbol{X}=\boldsymbol{x}$, reads

$$\frac{\overline{E}(u_n(1+\varepsilon_n)|\boldsymbol{x})}{\overline{E}(u_n|\boldsymbol{x})} = 1 - \frac{\varepsilon_n}{\gamma(\boldsymbol{x})}(1+o(1))$$

as soon as $u_n \to \infty$ and $\varepsilon_n \to 0$ are such that $A(1/\overline{F}(u_n|\boldsymbol{x})|\boldsymbol{x}) = O(\varepsilon_n)$. This equation provides

$$\overline{E}(e(\tau_n|\boldsymbol{x})(1+z_1\sigma_n)|\boldsymbol{x}) = \overline{E}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x}) \left(1 - \frac{z_1\sigma_n}{\gamma(\boldsymbol{x})}(1+o(1))\right).$$

Here the asymptotic proportionality between $e(\tau_n|\boldsymbol{x})$ and $q(\tau_n|\boldsymbol{x})$ was used, that is,

$$\frac{e(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})} \to \left(\frac{1}{\gamma(\boldsymbol{x})} - 1\right)^{-\gamma(\boldsymbol{x})} \quad \text{as } n \to \infty$$

[17, Proposition 1(i)], together with the regular variation property of $A(\cdot|\boldsymbol{x})$ and the choices $u_n = e(\tau_n|\boldsymbol{x})$ and $\varepsilon_n = z_1\sigma_n$. Therefore:

$$\frac{\overline{E}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{E}(e(\tau_n|\boldsymbol{x})(1+z_1\sigma_n)|\boldsymbol{x})} - 1 = \frac{z_1}{\gamma(\boldsymbol{x})}\sigma_n(1+o(1)),$$
$$\text{and } \frac{\overline{F}(q(\beta_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{F}(q(\beta_n|\boldsymbol{x})(1+z_2\sigma_n)|\boldsymbol{x})} - 1 = \frac{z_2}{\gamma(\boldsymbol{x})}\sigma_n(1+o(1)).$$

Continuity of the mapping

$$y \mapsto E(y|\boldsymbol{x}) = 1 - \frac{\mathbb{E}[(Y-y)\mathbb{1}_{\{Y>y\}}|\boldsymbol{X}=\boldsymbol{x}]}{\mathbb{E}[|Y-y||\boldsymbol{X}=\boldsymbol{x}]}$$

is immediate, by the dominated convergence theorem, so that $\overline{E}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x}) = 1 - \tau_n$. Lemma A.3(iii) in [46] applied to the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$ then gives

$$\frac{\overline{F}(y_n|\boldsymbol{x})}{1-\tau_n} = \frac{\overline{F}(y_n|\boldsymbol{x})}{\overline{E}(y_n|\boldsymbol{x})} \to \frac{1}{\gamma(\boldsymbol{x})} - 1 \quad \text{as } n \to \infty \tag{20}$$

and therefore

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{E}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{E}(y_n'|\boldsymbol{x})} - 1\right) = \frac{z_1}{\gamma(\boldsymbol{x})}\sqrt{\frac{1-\gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}}(1+o(1))$$
$$\text{and } \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{F}(q(\beta_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{F}(z_n'|\boldsymbol{x})} - 1\right) = \frac{z_2}{\gamma(\boldsymbol{x})}\sqrt{\frac{1-\gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}}(1+o(1)).$$

Conclude as in the end of the proof of Theorem 1 in [27], by applying Lemma A.9 to handle the current dependent data context. [Under conditions $\mathcal{KS}$, $\mathcal{D}_g$, $\mathcal{D}_m$ and $\mathcal{D}_\omega$, the bias component is obtained by noting that $\overline{F}(y_n|\boldsymbol{x})/(1-\tau_n) \to 1/\gamma(\boldsymbol{x}) - 1$, and therefore $c(\boldsymbol{x}) = \Delta\sqrt{1/\gamma(\boldsymbol{x}) - 1}$ with the notation of Lemma A.9. In the proof of Theorem 2.1, one should set instead $y_n = q(\tau_n|\boldsymbol{x})$ and therefore $\overline{F}(y_n|\boldsymbol{x})/(1-\tau_n) \to 1$, resulting in $c(\boldsymbol{x}) = \Delta$.] □

Proof of Theorems 2.2 and 3.3. The key is to write

$$\log\left(\frac{\widehat{q}^W_{n,\tau_n}(\tau_n'|\boldsymbol{x})}{q(\tau_n'|\boldsymbol{x})}\right) = \log\left(\frac{1-\tau_n}{1-\tau_n'}\right)(\widehat{\gamma}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) + \log\left(\frac{\overline{q}_n(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})}\right)$$
$$+ \log\left(\left[\frac{1-\tau_n}{1-\tau_n'}\right]^{\gamma(\boldsymbol{x})}\frac{q(\tau_n|\boldsymbol{x})}{q(\tau_n'|\boldsymbol{x})}\right)$$
$$\text{and } \log\left(\frac{\widehat{e}^W_{n,\tau_n}(\tau_n'|\boldsymbol{x})}{e(\tau_n'|\boldsymbol{x})}\right) = \log\left(\frac{1-\tau_n}{1-\tau_n'}\right)(\widehat{\gamma}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) + \log\left(\frac{\overline{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}\right)$$

$$+ \log\left(\left[\frac{1-\tau_n}{1-\tau_n'} \right]^{\gamma(\boldsymbol{x})} \frac{e(\tau_n|\boldsymbol{x})}{e(\tau_n'|\boldsymbol{x})} \right).$$

The conclusion now follows from our assumptions on $\widehat{\gamma}(\boldsymbol{x})$, $\overline{q}_n(\tau_n|\boldsymbol{x})$ (for Theorem 2.2) or $\overline{e}_n(\tau_n|\boldsymbol{x})$ (for Theorem 3.3), and the identity

$$\left(\frac{1-\tau_n}{1-\tau_n'} \right)^{\gamma(\boldsymbol{x})} \frac{q(\tau_n|\boldsymbol{x})}{q(\tau_n'|\boldsymbol{x})} = 1 + \frac{1}{\rho(\boldsymbol{x})} A\left((1-\tau_n)^{-1}|\boldsymbol{x} \right)(1+o(1))$$

(see p.139 in [19], for Theorem 2.2) or Lemma A.10 (for Theorem 3.3), together with a straightforward application of the delta-method and (for Theorem 3.3) the asymptotic proportionality relationship $e(\tau|\boldsymbol{x})/q(\tau|\boldsymbol{x}) \to (1/\gamma(\boldsymbol{x})-1)^{-\gamma(\boldsymbol{x})}$ (as $\tau \uparrow 1$) linking tail conditional expectiles and quantiles. □

Proof of Theorem 2.3. For $j \in \{1,\ldots,J\}$, set $c_j = 1/(J-j+1)$ (in particular $c_J = 1$ and $\log(J!) = -\sum_{j=1}^{J-1} \log(c_j)$). Write

$$\begin{aligned}
&\sqrt{nh_n^p(1-\tau_n)}\left(\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x}) - \gamma(\boldsymbol{x}) \right) \\
&= \sqrt{nh_n^p(1-\tau_n)} \left(\frac{1}{\log(J!)} \sum_{j=1}^{J-1} \log\left(\frac{\widehat{q}_n(1-c_j(1-\tau_n)|\boldsymbol{x})}{q(1-c_j(1-\tau_n)|\boldsymbol{x})} \right) \right. \\
&\qquad\qquad \left. - \frac{J-1}{\log(J!)} \log\left(\frac{\widehat{q}_n(1-c_J(1-\tau_n)|\boldsymbol{x})}{q(1-c_J(1-\tau_n)|\boldsymbol{x})} \right) \right) \\
&+ \sqrt{nh_n^p(1-\tau_n)} \frac{1}{\log(J!)} \sum_{j=1}^{J-1} \left(\log\left(\frac{q(1-c_j(1-\tau_n)|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})} \right) + \gamma(\boldsymbol{x}) \log(c_j) \right) \\
&=: A_n + B_n.
\end{aligned}$$

Theorem 2.3.9 in [19] yields, for any $u > 0$,

$$\log\left(\frac{q(1-(1-\tau_n)/u|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})} \right) = \gamma(\boldsymbol{x}) \log(u) + A((1-\tau_n)^{-1}|\boldsymbol{x}) \left(\frac{u^{\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})} + o(1) \right).$$

Condition $\sqrt{nh_n^p(1-\tau_n)} A((1-\tau_n)^{-1}|\boldsymbol{x}) \to \lambda(\boldsymbol{x})$ then entails

$$\lim_{n\to\infty} B_n = \frac{1}{\log(J!)} \left(\sum_{j=1}^{J-1} \frac{c_j^{-\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})} \right) \lambda(\boldsymbol{x}) = \frac{1}{\log(J!)} \left(\sum_{j=2}^{J} \frac{j^{\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})} \right) \lambda(\boldsymbol{x}).$$

Besides, combining Theorem 2.1 and a Taylor expansion of $u \mapsto \log(1+u)$ in a neighborhood of 0 gives

$$A_n = \sqrt{nh_n^p(1-\tau_n)} \frac{1}{\log(J!)} (1,\ldots,1,-(J-1)) \begin{pmatrix} \dfrac{\widehat{q}_n(1-c_1(1-\tau_n)|\boldsymbol{x})}{q(1-c_1(1-\tau_n)|\boldsymbol{x})} - 1 \\ \vdots \\ \dfrac{\widehat{q}_n(1-c_{J-1}(1-\tau_n)|\boldsymbol{x})}{q(1-c_{J-1}(1-\tau_n)|\boldsymbol{x})} - 1 \\ \dfrac{\widehat{q}_n(1-c_J(1-\tau_n)|\boldsymbol{x})}{q(1-c_J(1-\tau_n)|\boldsymbol{x})} - 1 \end{pmatrix} + o_{\mathbb{P}}(1).$$

Combining Theorem 2.1 and the elementary identity $\sum_{1\le i,j\le p}\max(i,j)=p(p+1)(4p-1)/6$ leads to the asymptotic normality of $\sqrt{nh_n^p(1-\tau_n)}(\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})-\gamma(\boldsymbol{x}))$, with the announced asymptotic mean and variance. It is then also a consequence of Theorem 2.1 that

$$\sqrt{nh_n^p(1-\tau_n)}\left(\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})-\gamma(\boldsymbol{x}),\frac{\widehat{q}_n(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})}-1\right)$$

is in fact jointly asymptotically normal, with the announced asymptotic mean vector, depending on the set of regularity assumptions made. The diagonal of the asymptotic covariance matrix has already been identified; according to the above representation of A_n, the off-diagonal element is nothing but the last element of the row vector $(1,\ldots,1,-(J-1))\boldsymbol{M}$, where $\boldsymbol{M}$ has entries $M_{j,l}(\boldsymbol{x})=1/\max(c_j,c_l)=1/c_{\max(j,l)}$. This is clearly 0, which completes the proof. □

Proof of Theorem 3.2. The key to the proof is the joint convergence in Theorem 2.3. Write

$$\log\frac{\breve{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}=\log((1/\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})-1)^{-\widehat{\gamma}_{\tau_n}^{(J)}(\boldsymbol{x})})-\log((1/\gamma(\boldsymbol{x})-1)^{-\gamma(\boldsymbol{x})})+\log\frac{\widehat{q}_n(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})}$$
$$-\log\left((1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})}\frac{e(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})}\right).$$

Note that the derivative of the function $z\mapsto\log((1/z-1)^{-z})=-z\log(1/z-1)$ on $(0,1)$ is $z\mapsto(1-z)^{-1}-\log(1/z-1)$. The result then follows from a straightforward application of Theorem 2.3 (for the joint convergence of the two random terms) in conjunction with Proposition 1(i) in [17] (for the convergence of the bias term) and the delta-method.

Proof of Theorem 3.4. Set $\kappa=1/\gamma(\boldsymbol{x})-1$ and $\sigma_n=1/\sqrt{nh_n^p(1-\tau_n)}$ and focus on the event

$$A_n(z)=\left\{\sigma_n^{-1}\left(\frac{\widehat{\overline{F}}_n(\widehat{e}_n(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{1-\tau_n}-\kappa\right)\le z\right\}=\left\{\widehat{\overline{F}}_n(\widehat{e}_n(\tau_n|\boldsymbol{x})|\boldsymbol{x})\le(1-\tau_n)(\kappa+z\sigma_n)\right\}.$$

Equivalently $A_n(z)=\{\widehat{e}_n(\tau_n|\boldsymbol{x})\ge\widehat{q}_n(\beta_n|\boldsymbol{x})\}$, where $\beta_n=1-(1-\tau_n)(\kappa+z\sigma_n)$, and so

$$A_n(z)=\left\{\sigma_n^{-1}\left(\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)\ge\sigma_n^{-1}\left(\frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\right)\frac{q(\beta_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}+\sigma_n^{-1}\left(\frac{q(\beta_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)\right\}.$$

According to Theorem 2.3.9 in [19], condition $\mathcal{C}_2(\gamma(\boldsymbol{x}),\rho(\boldsymbol{x}),A(\cdot|\boldsymbol{x}))$ provides:

$$\frac{q(\beta_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})}=(\kappa+z\sigma_n)^{-\gamma(\boldsymbol{x})}\left(1+\frac{(\kappa+z\sigma_n)^{-\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})}A((1-\tau_n)^{-1}|\boldsymbol{x})(1+o(1))\right)$$
$$=\kappa^{-\gamma(\boldsymbol{x})}\left(1-\gamma(\boldsymbol{x})\frac{z\sigma_n}{\kappa}(1+o(1))\right)\left(1+\frac{\kappa^{-\rho(\boldsymbol{x})}-1}{\rho(\boldsymbol{x})}A((1-\tau_n)^{-1}|\boldsymbol{x})(1+o(1))\right).$$

Apply Proposition 1 in [17] to get

$$\frac{q(\beta_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}=\left(1-\frac{\gamma^2(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}z\sigma_n(1+o(1))\right)$$
$$\times\left(1-\frac{\gamma(\boldsymbol{x})(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})}}{q(\tau_n|\boldsymbol{x})}(\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})+o(1))-\frac{(1/\gamma(\boldsymbol{x})-1)^{-\rho(\boldsymbol{x})}}{1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x})}A((1-\tau_n)^{-1}|\boldsymbol{x})(1+o(1))\right).$$

Linearizing yields

$$\sigma_n^{-1}\left(\frac{q(\beta_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)$$
$$\to -\frac{\gamma^2(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}z-\frac{(1/\gamma(\boldsymbol{x})-1)^{-\rho(\boldsymbol{x})}}{1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x})}\lambda_1(\boldsymbol{x})-\gamma(\boldsymbol{x})(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})}\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})\lambda_2(\boldsymbol{x})$$

as $n\to\infty$. Using Theorem 3.1, we can therefore rewrite $A_n(z)$ as

$$\left\{\frac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\left[-\sigma_n^{-1}\left(\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)+\sigma_n^{-1}\left(\frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\right)\right]\right.$$
$$\left.-\frac{(1/\gamma(\boldsymbol{x})-1)^{1-\rho(\boldsymbol{x})}}{\gamma(\boldsymbol{x})(1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x}))}\lambda_1(\boldsymbol{x})-(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})+1}\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})\lambda_2(\boldsymbol{x})+o_{\mathbb{P}}(1)\le z\right\}.$$

It follows that the asymptotic distribution of

$$\sigma_n^{-1}\left(\frac{\widehat{\overline{F}}_n(\widehat{e}_n(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{1-\tau_n}-\kappa,\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)$$

is that of

$$\begin{pmatrix}-\frac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})} & \frac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\\ 1 & 0\end{pmatrix}\sigma_n^{-1}\begin{pmatrix}\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\\ \frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\end{pmatrix}$$
$$-\begin{pmatrix}\frac{(1/\gamma(\boldsymbol{x})-1)^{-\rho(\boldsymbol{x})}}{\gamma(\boldsymbol{x})(1-\gamma(\boldsymbol{x})-\rho(\boldsymbol{x}))}\lambda_1(\boldsymbol{x})+(1/\gamma(\boldsymbol{x})-1)^{\gamma(\boldsymbol{x})+1}\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})\lambda_2(\boldsymbol{x})\\ 0\end{pmatrix}.$$

The conclusion follows from an application of Theorem 3.1 with $\kappa=1/\gamma(\boldsymbol{x})-1$ and a use of the delta-method for the function $u\mapsto 1/(1+u)$. □

APPENDIX B: DISCUSSION OF THE EXAMPLES

Here we provide an extended discussion of the examples in Section 4. We treat each example in turn; to enrich the discussion, we do not structure this section to include successive formal proofs of each proposition, but we explain in detail how to prove the validity of each of our conditions under the stated assumptions. We also provide alternative ways of checking these conditions, as well as more general conditions, when relevant. We start by the general case when the data are $m-$dependent (possibly i.i.d.), and we then give a full treatment for the location-scale model, the nonlinear regression model, and autoregressive models.

The $m-$dependent case (including the case of i.i.d. data). Suppose that $((\boldsymbol{X}_t,Y_t))_{t\ge1}$ is an $m-$dependent sequence. Then condition $\mathcal{M}$ reduces to second-order regular variation of the conditional survival function; numerous examples of commonly used distributions that satisfy this assumption can be found in [1]. Assuming the validity of this second-order regular variation condition, we check the other conditions as follows.

Checking conditions $\mathcal{B}_p$, $\mathcal{B}_m$ and $\mathcal{B}_\Omega$ is unnecessary because an $m-$dependent process is in particular $\psi-$mixing, and hence $\rho-$mixing. Condition $\mathcal{M}$ then reduces to second-order regular variation of the conditional survival function; numerous examples of commonly used distributions that satisfy this assumption (including the Generalized Pareto, Burr, Student and Fréchet distributions) can be found in [1].

Condition $\mathcal{L}_g$ is of course satisfied as soon as the p.d.f. g of $\boldsymbol{X}$ satisfies $g(\boldsymbol{x}) > 0$ and is continuously differentiable in a neighborhood V of $\boldsymbol{x}$, since it is then Lipschitz continuous on any closed subset of V by the mean value theorem. Differentiability of g at $\boldsymbol{x}$ is in fact sufficient for condition $\mathcal{L}_g$ to hold, since then $\|g(\boldsymbol{x}') - g(\boldsymbol{x}) - (\boldsymbol{x}' - \boldsymbol{x})^\top \nabla g(\boldsymbol{x})\| \le \|\boldsymbol{x}' - \boldsymbol{x}\|$ for $\boldsymbol{x}'$ sufficiently close to $\boldsymbol{x}$, by the very definition of the differentiability property. Similarly, condition $\mathcal{D}_g$ is satisfied as soon as ∇g is continuously differentiable on V, although continuity of ∇g on V and its differentiability at $\boldsymbol{x}$ are sufficient.

Condition $\mathcal{L}_m$ is verified as soon as Y given $\boldsymbol{X} = \boldsymbol{x}$ has a conditional p.d.f. $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ that is sufficiently smooth with respect to $\boldsymbol{x}$ uniformly in y. Indeed, since for any $a > 0$,

$$|\mathbb{E}(Y^a|\boldsymbol{X} = \boldsymbol{x}) - \mathbb{E}(Y^a|\boldsymbol{X} = \boldsymbol{x}')| \le \int_{\mathbb{R}} |y|^a |f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')|\, dy,$$

it turns out that condition $\mathcal{L}_m$ holds if there is a neighborhood V of $\boldsymbol{x}$ such that

$$\forall \boldsymbol{x}' \in V,\ |f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')| \le \phi_{\boldsymbol{x}}(y)\|\boldsymbol{x} - \boldsymbol{x}'\|, \text{ where } \phi_{\boldsymbol{x}} \text{ is measurable,}$$

$$\text{bounded in a neighborhood of 0, and such that } \int_{\mathbb{R}} y^2 (f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) + \phi_{\boldsymbol{x}}(y))\, dy < \infty.$$

This is nothing but a Lipschitz assumption on $f_{Y|\boldsymbol{X}}(y|\cdot)$, with the Lipschitz coefficient being a sufficiently integrable function of y; replacing y^2 by $|y|^{2+\delta}$ ensures that the moment restriction in condition $\mathcal{H}_\delta$ holds also. [Alternatively, this moment condition holds if Y has a finite *unconditional* moment of order $(2+\delta)$ and the conditional density of $\boldsymbol{X}$ given $Y = y$ is suitably uniformly bounded, see Lemma A.1(i).] The above Lipschitz assumption can be checked directly on the joint p.d.f. f of $(\boldsymbol{X}, Y)$: if it is such that $f(\cdot, y)$ is continuously differentiable on V for any y, then for any $\boldsymbol{x}' \in V$,

$$\begin{aligned}
|f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')| &= \left| \frac{f(\boldsymbol{x}, y)}{g(\boldsymbol{x})} - \frac{f(\boldsymbol{x}', y)}{g(\boldsymbol{x}')} \right| \\
&\le \left| \frac{1}{g(\boldsymbol{x}')} - \frac{1}{g(\boldsymbol{x})} \right| \sup_{\boldsymbol{x}' \in V} f(\boldsymbol{x}', y) + \frac{1}{g(\boldsymbol{x})} \left(\sup_{\boldsymbol{x}' \in V} \|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}', y)\| \right) \|\boldsymbol{x}' - \boldsymbol{x}\|
\end{aligned}$$

by the mean value theorem. Then, if g is Lipschitz continuous at $\boldsymbol{x}$, a possible construction of $\phi_{\boldsymbol{x}}$ is (up to a multiplicative term depending on $\boldsymbol{x}$ only)

$$\phi_{\boldsymbol{x}}(y) = \sup_{\boldsymbol{x}' \in V} f(\boldsymbol{x}', y) + \sup_{\boldsymbol{x}' \in V} \|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}', y)\|$$

for a sufficiently small neighborhood V of $\boldsymbol{x}$. To check condition $\mathcal{D}_m$, one possibility is to assume that a similar Lipschitz condition also holds on the partial gradient of $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$:

$$\forall \boldsymbol{x}' \in V,\ |f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')| + \|\nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')\| \le \phi_{\boldsymbol{x}}(y)\|\boldsymbol{x} - \boldsymbol{x}'\|$$

where, in addition to the above regularity and integrability requirements on $\phi_{\boldsymbol{x}}(y)$, it is assumed that $\nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\cdot)$ is continuous on V for any y, and $\nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ is bounded in a neighborhood of 0 and is such that $\int_{\mathbb{R}} y^2 \|\nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\|\, dy < \infty$. In that case, $\int_{\mathbb{R}} y \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy$ and $\int_{\mathbb{R}} y^2 \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy$ exist and, writing, for $a = 1, 2$,

$$\left| \mathbb{E}(Y^a|\boldsymbol{X} = \boldsymbol{x}') - \mathbb{E}(Y^a|\boldsymbol{X} = \boldsymbol{x}) - (\boldsymbol{x}' - \boldsymbol{x})^\top \int_{\mathbb{R}} y^a \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy \right|$$

$$
\begin{aligned}
&\le \int_{\mathbb{R}} |y|^a |f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}') - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - (\boldsymbol{x}'-\boldsymbol{x})^\top \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})|\, dy \\
&\le \int_{\mathbb{R}} |y|^a \left(\int_0^1 (\boldsymbol{x}'-\boldsymbol{x})^\top \|\nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}+t(\boldsymbol{x}'-\boldsymbol{x})) - \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\|\, dt \right) dy \\
&\le \frac{\|\boldsymbol{x}'-\boldsymbol{x}\|^2}{2} \int_{\mathbb{R}} |y|^a \phi_{\boldsymbol{x}}(y)\, dy,
\end{aligned}
$$

it is clear that the gradient of $\mathbb{E}(Y^a|\boldsymbol{X}=\cdot)$ at $\boldsymbol{x}$ is $\int_{\mathbb{R}} y^a \nabla_{\boldsymbol{x}} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy$, and that this gradient is Lipschitz continuous at $\boldsymbol{x}$, as required by condition $\mathcal{D}_m$. A similar discussion as above, at the level of the joint p.d.f. f, can be made to show that a possible construction of $\phi_{\boldsymbol{x}}$ is (up to a multiplicative term depending on $\boldsymbol{x}$ only)

$$\phi_{\boldsymbol{x}}(y) = \sup_{\boldsymbol{x}'\in V} f(\boldsymbol{x}',y) + \sup_{\boldsymbol{x}'\in V} \|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}',y)\| + \sup_{\boldsymbol{x}'\in V} \|H_{\boldsymbol{x}} f(\boldsymbol{x}',y)\|.$$

Finally, a general criterion for the control of the oscillation of the log-conditional survival function can be provided by assuming that the heavy tail assumption holds in a neighborhood V of $\boldsymbol{x}$ and by introducing the corresponding Karamata representation of the (hence regularly varying) conditional survival function, that is

$$
\begin{aligned}
&\forall \boldsymbol{x}' \in V,\ \forall y \ge y_0,\ \overline{F}(y|\boldsymbol{x}') = y^{-1/\gamma(\boldsymbol{x}')} L(y|\boldsymbol{x}'), \\
&\qquad \text{with } L(y|\boldsymbol{x}') = \exp\left(\eta(y|\boldsymbol{x}') + \int_{y_0}^{y} \frac{\epsilon(u|\boldsymbol{x}')}{u} du \right),
\end{aligned}
$$

for some $y_0 > 0$ [see Theorem 1.3.1 p.12 in 2], where $\gamma > 0$ and $\eta(\cdot|\boldsymbol{x}')$ and $\epsilon(\cdot|\boldsymbol{x}')$ are measurable functions converging, respectively, to a constant and 0 at infinity. With this notation,

$$
\begin{aligned}
&\frac{1}{\log(y)} \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} \\
&\qquad = -\left(\frac{1}{\gamma(\boldsymbol{x}')} - \frac{1}{\gamma(\boldsymbol{x})} \right) + \frac{\eta(y|\boldsymbol{x}') - \eta(y|\boldsymbol{x})}{\log(y)} + \frac{1}{\log(y)} \int_{y_0}^{y} \frac{\epsilon(u|\boldsymbol{x}') - \epsilon(u|\boldsymbol{x})}{u} du.
\end{aligned}
$$

It is then clear that if there is $c > 0$ such that

$$\forall \boldsymbol{x}' \in V,\ \left| \frac{1}{\gamma(\boldsymbol{x})} - \frac{1}{\gamma(\boldsymbol{x}')} \right| + \sup_{y\ge y_0} \left| \frac{\eta(y|\boldsymbol{x}) - \eta(y|\boldsymbol{x}')}{\log(y)} \right| + \sup_{y \ge y_0} \left| \epsilon(y|\boldsymbol{x}) - \epsilon(y|\boldsymbol{x}') \right| \le c\, \|\boldsymbol{x} - \boldsymbol{x}'\|$$

then condition $\mathcal{L}_\omega$ is satisfied. This will in particular be the case provided V is a (without loss of generality) compact neighborhood of $\boldsymbol{x}$ such that γ is continuously differentiable on V, and (by the mean value theorem) $\eta(y|\cdot)/\log(y)$ and $\epsilon(y|\cdot)$ are continuously differentiable on V for all $y \ge y_0$, with bounded gradients on $V \times [y_0, \infty)$. To check condition $\mathcal{D}_\omega$, note that

$$\log L(y|\boldsymbol{x}') = \eta(y|\boldsymbol{x}') + \int_{y_0}^{y} \frac{\epsilon(u|\boldsymbol{x}')}{u} du.$$

If $\eta(y|\cdot)/\log(y)$ and $\epsilon(y|\cdot)$ are twice continuously differentiable on V for any $y \ge y_0$ with bounded Hessian matrices on the cylinder $V \times [y_0, \infty)$, then (by differentiating under the integral) it is straightforward to prove that the partial Hessian matrix $H_{\boldsymbol{x}} \log L(y|\boldsymbol{x}')/\log(y)$ is well-defined and bounded on $V \times [y_0, \infty)$. If then γ is twice continuously differentiable on V, Lemma A.4(ii) entails that condition $\mathcal{D}_\omega$ is satisfied, with $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y) = \nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x})$. We summarize this discussion in the following synthetic result.

PROPOSITION B.1 (The i.i.d. or $m-$dependent case). *Assume that $(\boldsymbol{X},Y)$ has a joint p.d.f. f. Let $\mathcal{X}$ denote the support of $\boldsymbol{X}$, assumed to have nonempty interior, let $\boldsymbol{x}$ belong to the interior of $\mathcal{X}$ and $V\subset\mathcal{X}$ be a neighborhood of $\boldsymbol{x}$ such that the p.d.f. g of $\boldsymbol{X}$ is continuously differentiable on V and $g(\boldsymbol{x})>0$. Then condition $\mathcal{L}_g$ holds; if g is also twice continuously differentiable on V, then condition $\mathcal{D}_g$ holds as well.*

If moreover Y given $\boldsymbol{X}$ has a conditional p.d.f. $f_{Y|\boldsymbol{X}}$, such that

$$\forall \boldsymbol{x}'\in V,\ |f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})-f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')|\le\phi_{\boldsymbol{x}}(y)\|\boldsymbol{x}-\boldsymbol{x}'\|$$

where $\phi_{\boldsymbol{x}}$ is measurable, bounded in a neighborhood of 0 and such that $\int_{\mathbb{R}}|y|^{2+\delta}(f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})+\phi_{\boldsymbol{x}}(y))\,dy<\infty$, then conditions $\mathcal{H}_\delta$ and $\mathcal{L}_m$ hold (when $\gamma(\boldsymbol{x})<1/(2+\delta)$). If in addition $f_{Y|\boldsymbol{X}}(y|\cdot)$ is continuously differentiable on V for any y, the partial gradient function $\nabla_{\boldsymbol{x}}f_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ is bounded in a neighborhood of 0 and satisfies also

$$\forall \boldsymbol{x}'\in V,\ \|\nabla_{\boldsymbol{x}}f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})-\nabla_{\boldsymbol{x}}f_{Y|\boldsymbol{X}}(y|\boldsymbol{x}')\|\le\phi_{\boldsymbol{x}}(y)\|\boldsymbol{x}-\boldsymbol{x}'\|$$

with $\int_{\mathbb{R}}|y|^{2+\delta}\|\nabla_{\boldsymbol{x}}f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\|\,dy<\infty$, then condition $\mathcal{D}_m$ holds as well.

Assume finally that for any $\boldsymbol{x}'$ in a neighborhood V of $\boldsymbol{x}$, $\overline{F}(\cdot|\boldsymbol{x}')$ is heavy-tailed, and recall the Karamata representation of heavy-tailed conditional survival functions,

$$\forall y\ge y_0,\ \overline{F}(y|\boldsymbol{x}')=y^{-1/\gamma(\boldsymbol{x}')}\exp\left(\eta(y|\boldsymbol{x}')+\int_{y_0}^{y}\frac{\epsilon(u|\boldsymbol{x}')}{u}du\right)$$

for some $y_0>0$ [see Theorem 1.3.1 on p.12 of 2], where $\gamma>0$ and $\eta(\cdot|\boldsymbol{x}')$ and $\epsilon(\cdot|\boldsymbol{x}')$ are measurable functions converging, respectively, to a constant and 0 at infinity. If γ is continuously differentiable on V, and $\eta(y|\cdot)/\log(y)$ and $\epsilon(y|\cdot)$ are continuously differentiable on V for any $y\ge y_0$ with bounded gradients on the cylinder $V\times[y_0,\infty)$, then condition $\mathcal{L}_\omega$ holds. If moreover γ is twice continuously differentiable on V, and $\eta(y|\cdot)/\log(y)$ and $\epsilon(y|\cdot)$ are twice continuously differentiable on V for any $y\ge y_0$ with bounded Hessian matrices on the cylinder $V\times[y_0,\infty)$, then condition $\mathcal{D}_\omega$ holds, with $\lim_{y\to\infty}\nabla_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x})/\log(y)=\nabla\gamma(\boldsymbol{x})/\gamma^2(\boldsymbol{x})$.

Location-scale model with possible temporal misspecification. In this case, strong mixing of the sequence $((\boldsymbol{X}_t,Y_t))_{t\ge1}$ is a consequence of the strong mixing of $((\boldsymbol{X}_t,\varepsilon_t))_{t\ge1}$. Indeed, since

$$\sigma(\{(\boldsymbol{X}_j,\varepsilon_j),a\le j\le b\})=\sigma(\sigma(\{\boldsymbol{X}_j,a\le j\le b\})\cup\sigma(\{\varepsilon_j,a\le j\le b\}))$$

with the two $\sigma-$algebras on the right-hand side being independent by assumption, Lemma 8 in [5] applied to the $\sigma-$algebras $\mathcal{A}_1=\sigma(\{\boldsymbol{X}_j,1\le j\le k\})$, $\mathcal{B}_1=\sigma(\{\boldsymbol{X}_j,j\ge k+n\})$, $\mathcal{A}_2=\sigma(\{\varepsilon_j,1\le j\le k\})$ and $\mathcal{B}_2=\sigma(\{\varepsilon_j,j\ge k+n\})$ yields $\alpha(n)\le\alpha_{\boldsymbol{X}}(n)+\alpha_\varepsilon(n)$, where $\alpha_{\boldsymbol{X}}(n)$ and $\alpha_\varepsilon(n)$ are the strong mixing coefficients of $(\boldsymbol{X}_t)_{t\ge1}$ and $(\varepsilon_t)_{t\ge1}$. In other words, strong mixing of $((\boldsymbol{X}_t,Y_t))_{t\ge1}$ holds as soon as $(\boldsymbol{X}_t)_{t\ge1}$ and $(\varepsilon_t)_{t\ge1}$ themselves are strongly mixing. In particular, if $(\varepsilon_t)_{t\ge1}$ is an i.i.d. sequence (meaning that the location-scale regression model is correctly specified), strong mixing of $((\boldsymbol{X}_t,Y_t))_{t\ge1}$ reduces to strong mixing of $(\boldsymbol{X}_t)_{t\ge1}$ alone.

In addition, denoting by ε a random variable having the same marginal distribution as the ε_t, we clearly have

$$Y|\boldsymbol{X}=\boldsymbol{x}\stackrel{d}{=}m(\boldsymbol{x})+\sigma(\boldsymbol{x})\varepsilon$$

and $(Y_1,Y_{t+1})|\{\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{X}_{t+1}=\boldsymbol{x}_{t+1}\}\stackrel{d}{=}(m(\boldsymbol{x}_1)+\sigma(\boldsymbol{x}_1)\varepsilon_1,m(\boldsymbol{x}_{t+1})+\sigma(\boldsymbol{x}_{t+1})\varepsilon_{t+1})$.

This entails in particular that the conditional second-order heavy tails assumption should essentially reduce to regular variation properties of the distribution of ε with tail index $\gamma > 0$. However, the presence of the location and scale components $m(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ are known to have a substantial influence on these regular variation properties (see p.83 in [19]), so we provide a full discussion next. Assume that

$$\forall z > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}_\varepsilon(t))}\left(\frac{\overline{F}_\varepsilon(tz)}{\overline{F}_\varepsilon(t)} - z^{-1/\gamma}\right) = \begin{cases} z^{-1/\gamma}\dfrac{z^{\rho/\gamma}-1}{\gamma\rho} & \text{if } \rho < 0, \\ z^{-1/\gamma}\dfrac{\log(z)}{\gamma^2} & \text{if } \rho = 0 \end{cases}$$

where $\rho \le 0$ and A is a measurable function having constant sign, or equivalently,

$$\forall z > 0,\ \lim_{t\to\infty} \frac{1}{A(t)}\left(\frac{U_\varepsilon(tz)}{U_\varepsilon(t)} - z^{\gamma}\right) = \begin{cases} z^{\gamma}\dfrac{z^{\rho}-1}{\rho} & \text{if } \rho < 0, \\ z^{\gamma}\log(z) & \text{if } \rho = 0 \end{cases}$$

where $U_\varepsilon(t) = q_\varepsilon(1-1/t) = \inf\{z \in \mathbb{R} \,|\, F_\varepsilon(z) \ge 1-1/t\}$ is the tail quantile function of ε (see Theorem 2.3.9 on p.48 of [19] for the equivalence between these two convergences). The function $|A|$ is regularly varying with index ρ. Then, for any $m \in \mathbb{R}$ and $\sigma > 0$, a straightforward calculation shows that the tail quantile function $m + \sigma U_\varepsilon$ of $m + \sigma\varepsilon$ satisfies

$$\forall z > 0,\ \frac{m+\sigma U_\varepsilon(tz)}{m+\sigma U_\varepsilon(t)} - z^{\gamma} = \left(1 + \frac{m}{\sigma U_\varepsilon(t)}\right)^{-1}\left[\left(\frac{U_\varepsilon(tz)}{U_\varepsilon(t)} - z^{\gamma}\right) - \frac{m\gamma}{\sigma U_\varepsilon(t)} \times z^{\gamma}\frac{z^{-\gamma}-1}{-\gamma}\right].$$

As a consequence of this calculation, if $\gamma \ne -\rho$, then assumption $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ holds for the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$, with $\gamma(\boldsymbol{x}) = \gamma$ for any $\boldsymbol{x}$, when:

- $m(\boldsymbol{x}) = 0$ or $\gamma > -\rho$, in which case $\rho(\boldsymbol{x}) = \rho$ and $A(\cdot|\boldsymbol{x}) = A$,
- $m(\boldsymbol{x}) \ne 0$ and $\gamma < -\rho$, in which case $\rho(\boldsymbol{x}) = -\gamma$ and $A(\cdot|\boldsymbol{x}) = m(\boldsymbol{x})\gamma/(\sigma(\boldsymbol{x})U_\varepsilon)$.

[It follows that in the pure scale model where $m(\boldsymbol{x}) = 0$, no restriction on γ and ρ is necessary.] When a location component is present, this restriction on γ and ρ can be lifted if the tail quantile function of ε satisfies an asymptotic expansion of the form

$$(21) \qquad U_\varepsilon(z) = Cz^{\gamma}(1 + Dz^{\rho} + D'z^{\rho+\rho'}(1+o(1))) \text{ as } z \to \infty$$

where $C > 0$, $D, D' \ne 0$ and $\rho, \rho' < 0$. Indeed, note first that any tail quantile function satisfying the weaker expansion

$$(22) \qquad U_\varepsilon(z) = Cz^{\gamma}(1 + Dz^{\rho}(1+o(1))) \text{ as } z \to \infty$$

will also satisfy the second-order regular variation condition with $A(t) = D\rho t^{\rho}$. If (21) holds, then

$$m + \sigma U_\varepsilon(z) = \sigma C z^{\gamma}\left(1 + \frac{m}{\sigma C}z^{-\gamma} + Dz^{\rho} + D'z^{\rho+\rho'}(1+o(1))\right) \text{ as } z \to \infty.$$

In that situation, even if $\gamma = -\rho$, assumption $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ always holds for the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$, with

- $\rho(\boldsymbol{x}) = \rho = -\gamma$ and $A(t|\boldsymbol{x}) = (D + m/(\sigma C))\rho t^{\rho}$ if $m + \sigma CD \ne 0$,
- $\rho(\boldsymbol{x}) = \rho + \rho'$ and $A(t|\boldsymbol{x}) = D'(\rho+\rho')t^{\rho+\rho'}$ otherwise.

Assumption (21), which puts the distribution of ε into a subset of the Hall class [named after 33], is reasonable and satisfied by most of the classes of distributions used in the modeling of heavy tails, such as:

- The Fréchet distribution with survival function $\overline{F}(y) = 1 - \exp(-y^{-1/\gamma})$ ($y > 0$),
- The Burr distribution with survival function $\overline{F}(y) = (1 + y^{-\rho/\gamma})^{1/\rho}$ ($y > 0$).

A notable exception is the Generalized Pareto distribution, for which it may happen that $m + \sigma U_\varepsilon$ is the tail quantile function of the pure Pareto distribution, that is, $m + \sigma U_\varepsilon(z) = z^\gamma$ ($z > 1$). More generally, any example where $m + \sigma U_\varepsilon$ is asymptotically equivalent to a multiple of z^γ at a rate faster than any polynomial, that is, $m + \sigma U_\varepsilon(z) = cz^\gamma(1 + o(z^{-\kappa}))$ for any $\kappa > 0$ (as $z \to \infty$). This represents the ideal case when extrapolation bias is absent, and our results continue to hold with the convention $\rho(\boldsymbol{x}) = -\infty$ and $A(\cdot|\boldsymbol{x}) \equiv 0$. It may happen though that (21) is not straightforward to check directly because the quantile function is not easy to compute and only the p.d.f. f_ε of ε has a simple form. In this case, a workaround is to assume directly that f_ε, rather than U_ε, satisfies such an asymptotic expansion, that is,

$$f_\varepsilon(z) = c_0 z^{-1/\gamma - 1}(1 + d_0 z^{-a} + d_0' z^{-a-b}(1 + o(1))) \text{ as } z \to \infty \tag{23}$$

where $c_0 > 0$, $d_0, d_0' \neq 0$ and $a, b > 0$. Writing $\overline{F}_\varepsilon(z) = \int_z^\infty f_\varepsilon(t)\,dt$ immediately entails that $\overline{F}_\varepsilon$ itself then satisfies an analogous expansion:

$$\overline{F}_\varepsilon(z) = cz^{-1/\gamma}(1 + dz^{-a} + d' z^{-a-b}(1 + o(1))) \text{ as } z \to \infty \tag{24}$$

where $c = c_0\gamma > 0$, $d = d_0/(1 + a\gamma) \neq 0$ and $d' = d_0'/(1 + (a + b)\gamma) \neq 0$. The identity $\overline{F}_\varepsilon(U_\varepsilon(z)) = 1/z$, valid for any $z > 1$ because of the continuity of F_ε, now provides

$$U_\varepsilon(z) = c^\gamma z^\gamma (1 + d(U_\varepsilon(z))^{-a} + d'(U_\varepsilon(z))^{-a-b}(1 + o(1)))^\gamma \text{ as } z \to \infty.$$

We then successively obtain $U_\varepsilon(z) = c^\gamma z^\gamma(1 + o(1))$, $U_\varepsilon(z) = c^\gamma z^\gamma(1 + \gamma d c^{-a\gamma} z^{-a\gamma}(1 + o(1)))$ and finally

$$U_\varepsilon(z) = c^\gamma z^\gamma \Big(1 + \gamma d c^{-a\gamma} z^{-a\gamma} - \frac{1}{2}\gamma d^2 c^{-2a\gamma}(1 + (2a - 1)\gamma) z^{-2a\gamma} \\ + \gamma d' c^{-(a+b)\gamma} z^{-(a+b)\gamma} + o(z^{-2a\gamma}) + o(z^{-(a+b)\gamma})\Big)$$

as $z \to \infty$. When $a \neq 1$, (22) holds with $\rho = -a\gamma \neq -\gamma$, so no further discussion is necessary. When $a = 1$ and $b \neq 1$, this obviously yields an asymptotic expansion of the form (21). Otherwise, when $a = b = 1$, the last two terms decay at the same rate, and this will be an asymptotic expansion of the form (21), meaning that once again $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ will be satisfied, as long as $2d' \neq d^2(1+\gamma)$, or equivalently $2d_0'(1+\gamma) \neq d_0^2(1+2\gamma)$. For instance:

- Consider the Student distribution with $\nu > 0$ degrees of freedom, with p.d.f. f_ε given by

$$f_\varepsilon(z) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left(1 + \frac{z^2}{\nu}\right)^{-(\nu+1)/2}$$

It is readily found that Assumption (23) holds with $\gamma = 1/\nu$, $a = b = 2$, $d_0 = -\nu(\nu+1)/2$ and $d_0' = \nu^2(\nu+1)(\nu+3)/8$. Since $a = 2$, the related tail quantile function U_ε is second-order regularly varying with second-order parameter $\rho = -2\gamma \neq -\gamma$, so condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ will be satisfied in the location-scale regression model.
- Consider the Fisher distribution with (ν_1, ν_2) degrees of freedom, defined as

$$f_\varepsilon(z) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} z^{\nu_1/2 - 1}\left(1 + \frac{\nu_1}{\nu_2} z\right)^{-(\nu_1+\nu_2)/2}$$

(where $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}\,du$ denotes the Beta function at $x, y > 0$). Here a straightforward calculation gives $\gamma = 2/\nu_2$, $a = b = 1$, $d_0 = -\nu_2(\nu_1 + \nu_2)/(2\nu_1)$ and $d_0' =$

$\nu_2^2(\nu_1+\nu_2)(\nu_1+\nu_2+2)/(8\nu_1^2)$, so that the Fisher distribution also satisfies Assumption (24) with $2d_0'(1+\gamma) \neq d_0^2(1+2\gamma)$ if and only if $\nu_1 \neq 2$. Condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ will then be satisfied in the location-scale regression model when $\nu_1 \neq 2$. [If on the contrary $\nu_1 = 2$, the Fisher distribution is in fact a Generalized Pareto distribution, whose relevance to our results has already been discussed above.]

As a conclusion, condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ holds provided the p.d.f. f_ε satisfies an asymptotic expansion of the form (23), with either $a \neq 1$, $a = 1 \neq b$, or $a = b = 1$ and $2d_0'(1+\gamma) \neq d_0^2(1+2\gamma)$.

In any of the regular variation models discussed above, since $\gamma(\boldsymbol{x}) = \gamma$ for any $\boldsymbol{x}$, it follows that condition $\mathcal{H}_\delta$ holds provided $\gamma < 1/(2+\delta)$, $\mathbb{E}(\varepsilon_-^{2+\delta}) < \infty$ and the functions m and σ are bounded in a neighborhood of $\boldsymbol{x}$.

Condition $\mathcal{L}_g$ clearly holds if the p.d.f. of $\boldsymbol{X}$ is such that $g(\boldsymbol{x}) > 0$ and g is continuously differentiable in a neighborhood of $\boldsymbol{x}$, and condition $\mathcal{L}_m$ reduces to assuming that ε has a finite second moment and the location and scale components m and σ are themselves Lipschitz continuous at $\boldsymbol{x}$, which will happen if m and σ are in fact continuously differentiable in a neighborhood of $\boldsymbol{x}$ (again, differentiability at $\boldsymbol{x}$ is sufficient). Under the latter conditions, condition $\mathcal{B}_m$ is automatically satisfied; condition $\mathcal{B}_p$ is automatically true under condition $\mathcal{L}_g$ if the random pairs $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ have absolutely continuous distributions and $(\boldsymbol{X}_t)$ is actually $\beta-$mixing (or absolutely regular), because then the p.d.f.s g_t of the pairs $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ are uniformly bounded in t, see Remark 1 in [11]. It follows that conditions $\mathcal{L}_g$, $\mathcal{L}_m$, $\mathcal{B}_p$ and $\mathcal{B}_m$ hold under the assumptions provided in the first two items of Proposition 4.1, in addition to $\gamma < 1/(2+\delta)$ and $\mathbb{E}(\varepsilon_-^{2+\delta}) < \infty$ for $\mathcal{L}_m$ and $\mathcal{B}_m$. Conditions $\mathcal{D}_g$ and $\mathcal{D}_m$ then clearly hold if g, m and σ have Lipschitz continuous gradients at $\boldsymbol{x}$; this is in particular true if these three functions are twice continuously differentiable in a neighborhood of $\boldsymbol{x}$.

Finally, $\overline{F}(y|\boldsymbol{x}) = \overline{F}_\varepsilon((y - m(\boldsymbol{x}))/\sigma(\boldsymbol{x}))$ in this model, and thus

$$\frac{1}{\log(y)}\left|\log\frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})}\right| = \frac{1}{\log(y)}\left|\log\frac{\overline{F}_\varepsilon((y-m(\boldsymbol{x}'))/\sigma(\boldsymbol{x}'))}{\overline{F}_\varepsilon((y-m(\boldsymbol{x}))/\sigma(\boldsymbol{x}))}\right|.$$

If ε has a p.d.f. f_ε which is continuous in a neighborhood of infinity then, for y large enough, by the mean value theorem:

$$\left|\log\frac{\overline{F}_\varepsilon((y-m(\boldsymbol{x}'))/\sigma(\boldsymbol{x}'))}{\overline{F}_\varepsilon((y-m(\boldsymbol{x}))/\sigma(\boldsymbol{x}))}\right| \le \frac{\left|\frac{y-m(\boldsymbol{x})}{\sigma(\boldsymbol{x})} - \frac{y-m(\boldsymbol{x}')}{\sigma(\boldsymbol{x}')}\right|}{\min\left(\frac{y-m(\boldsymbol{x})}{\sigma(\boldsymbol{x})}, \frac{y-m(\boldsymbol{x}')}{\sigma(\boldsymbol{x}')}\right)} \sup_{z\in I_{\boldsymbol{x},\boldsymbol{x}'}(y)} |z(\log\overline{F}_\varepsilon)'(z)|$$

$$\text{with } I_{\boldsymbol{x},\boldsymbol{x}'}(y) = \left[\frac{y-m(\boldsymbol{x})}{\sigma(\boldsymbol{x})}, \frac{y-m(\boldsymbol{x}')}{\sigma(\boldsymbol{x}')}\right].$$

It follows that condition $\mathcal{L}_\omega$ holds (and even $\omega_{h_n}(y_n|\boldsymbol{x}) = o(h_n)$ for any $y_n \to \infty$) provided m and σ are Lipschitz continuous at $\boldsymbol{x}$ and the distribution of ε satisfies the classical first-order von Mises condition $-z(\log\overline{F}_\varepsilon)'(z) = zf_\varepsilon(z)/\overline{F}_\varepsilon(z) \to 1/\gamma$ as $z \to \infty$ [see condition (1.1.34) p.17 in 19]. This first-order von Mises condition is in particular a consequence of Assumption (23). To check condition $\mathcal{D}_\omega$, recall that the chain rule for a function $\phi : \boldsymbol{x} \mapsto f(u(\boldsymbol{x}))$, where $f : \mathbb{R} \to \mathbb{R}$ and $u : \mathbb{R}^p \to \mathbb{R}$ are twice differentiable, yields

$$H\phi(\boldsymbol{x}) = f''(u(\boldsymbol{x}))\nabla u(\boldsymbol{x})[\nabla u(\boldsymbol{x})]^\top + f'(u(\boldsymbol{x}))Hu(\boldsymbol{x}).$$

If m and σ are twice continuously differentiable in a compact (without loss of generality) neighborhood V of $\boldsymbol{x}$, the gradient and Hessian matrix of $\boldsymbol{x} \mapsto (y - m(\boldsymbol{x}))/\sigma(\boldsymbol{x})$ are respectively

$$\frac{y-m(\boldsymbol{x})}{\sigma(\boldsymbol{x})} \times \sigma(\boldsymbol{x})\nabla(1/\sigma)(\boldsymbol{x}) - \frac{\nabla m(\boldsymbol{x})}{\sigma(\boldsymbol{x})}$$

and

$$\frac{y-m(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\times\sigma(\boldsymbol{x})H(1/\sigma)(\boldsymbol{x})-\frac{1}{\sigma(\boldsymbol{x})}Hm(\boldsymbol{x})-\nabla m(\boldsymbol{x})[\nabla(1/\sigma)(\boldsymbol{x})]^\top-\nabla(1/\sigma)(\boldsymbol{x})[\nabla m(\boldsymbol{x})]^\top.$$

It follows that if the p.d.f. f_ε itself is continuously differentiable in a neighborhood of infinity then, under the second-order von Mises condition $-z(\log f_\varepsilon)'(z)=-zf_\varepsilon'(z)/f_\varepsilon(z)\to 1/\gamma+1$ as $z\to\infty$ (in addition to its first-order version), the partial Hessian matrix $H_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x}')/\log(y)$ converges to the zero matrix as $y\to\infty$ uniformly in $\boldsymbol{x}'\in V$. Then, by Lemma A.4(ii), condition $\mathcal{D}_\omega$ is satisfied, with $\lim_{y\to\infty}\nabla_{\boldsymbol{x}}\log\overline{F}(y|\boldsymbol{x})/\log(y)=0$ (since γ is a constant function of $\boldsymbol{x}$ here).

Finally,

$$\begin{aligned}&\frac{\mathbb{P}(Y_1>y,Y_{t+1}>y'|\boldsymbol{X}_1=\boldsymbol{x}',\boldsymbol{X}_{t+1}=\boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}}\\&=\frac{\mathbb{P}(\varepsilon_1>(y-m(\boldsymbol{x}'))/\sigma(\boldsymbol{x}'),\varepsilon_{t+1}>(y'-m(\boldsymbol{x}''))/\sigma(\boldsymbol{x}''))}{\sqrt{\overline{F}_\varepsilon((y-m(\boldsymbol{x}'))/\sigma(\boldsymbol{x}'))\overline{F}_\varepsilon((y'-m(\boldsymbol{x}''))/\sigma(\boldsymbol{x}''))}}\le 1\end{aligned}$$

by the Cauchy-Schwarz inequality, so that condition $\mathcal{B}_\Omega$ is satisfied under no further condition whatsoever.

Nonlinear regression model. As in the example of the location-scale model, strong mixing of the sequence $((\boldsymbol{X}_t,Y_t))_{t\ge1}$ is a consequence of the strong mixing of $((\boldsymbol{X}_t,U_t))_{t\ge1}$, which itself follows from strong mixing of $(\boldsymbol{X}_t)_{t\ge1}$ and $(U_t)_{t\ge1}$. Besides, the conditional distributions in this model are

$$Y|\boldsymbol{X}=\boldsymbol{x}\overset{d}{=}q(U,\boldsymbol{\theta}(\boldsymbol{x}))$$

$$\text{and } (Y_1,Y_{t+1})|\{\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{X}_{t+1}=\boldsymbol{x}_{t+1}\}\overset{d}{=}(q(U_1,\boldsymbol{\theta}(\boldsymbol{x}_1)),q(U_{t+1},\boldsymbol{\theta}(\boldsymbol{x}_{t+1}))).$$

From the first equality above, the conditional distribution function of Y given $\boldsymbol{X}=\boldsymbol{x}$ is exactly $F(\cdot,\boldsymbol{\theta}(\boldsymbol{x}))$. The conditional second-order regular variation assumption is then a direct consequence of the second-order regular variation property of $\overline{F}(\cdot,\boldsymbol{\theta})$.

The validity of conditions $\mathcal{L}_g$, $\mathcal{B}_p$ and $\mathcal{D}_g$ follows from the assumptions on the distributions of the pairs $(\boldsymbol{X}_1,\boldsymbol{X}_{t+1})$ and their first marginal, see the example of the location scale-regression model for a complete discussion.

The validity of conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ depends on the particular form of the model and how smooth the function $\boldsymbol{\theta}(\cdot)$ is. We focus throughout this example on classical heavy-tailed distributions, such as those of Table 2.1 p.59 in [1], which are concentrated on $(0,\infty)$. For such models, the left tail moment requirement is trivially true, so checking condition $\mathcal{H}_\delta$ is done by calculating the value of γ as a function of $\boldsymbol{\theta}$ and ensuring that $\gamma(\boldsymbol{x})=\gamma(\boldsymbol{\theta}(\boldsymbol{x}))<1/(2+\delta)$. [Otherwise, one requires extra assumptions about the left tail of $\overline{F}(\cdot,\boldsymbol{\theta})$, such as symmetry of the tails, or the fact that the left tail is dominated by the right tail.] We turn to checking conditions $\mathcal{L}_m$ and $\mathcal{B}_m$. Let $m_1(\boldsymbol{\theta})$ and $m_2(\boldsymbol{\theta})$ denote the expectation and second moment of the parametric model $F(\cdot,\boldsymbol{\theta})$. Then conditions $\mathcal{L}_m$ and $\mathcal{B}_m$ will be satisfied provided $\boldsymbol{x}\mapsto m_1(\boldsymbol{\theta}(\boldsymbol{x}))$ and $\boldsymbol{x}\mapsto m_2(\boldsymbol{\theta}(\boldsymbol{x}))$ exist, are finite, and are Lipschitz continuous at $\boldsymbol{x}$. In practice this will hold provided $\boldsymbol{\theta}(\cdot)$ is Lipschitz continuous at $\boldsymbol{x}$ (*e.g.* continuously differentiable in a neighborhood of $\boldsymbol{x}$) and $m_1(\cdot)$ and $m_2(\cdot)$ are continuously differentiable with respect to $\boldsymbol{\theta}$. For example:

• The Fréchet model $F(y,\theta)=\exp(-y^{-1/\theta})$ (for $y,\theta>0$), is heavy-tailed with tail index $\gamma=\theta$. In this model, $m_1(\theta)=\Gamma(1-\theta)$ and $m_2(\theta)=\Gamma(1-2\theta)$ when $\theta<1/2$, where Γ is Euler's Gamma function, and therefore the functions m_1 and m_2 are infinitely differentiable functions of θ on $(0,1/2)$. Conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ then hold together if the function $\theta(\cdot)$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$ with $\theta(\boldsymbol{x})<1/(2+\delta)$.
• The absolute Student (or half-t) model with θ degrees of freedom, where $F(\cdot,\theta)$ has p.d.f.

$$f(y,\theta)=\frac{2\,\Gamma((\theta+1)/2)}{\sqrt{\theta\pi}\,\Gamma(\theta/2)}\left(1+\frac{y^2}{\theta}\right)^{-(\theta+1)/2}\quad\text{(for } y,\theta>0\text{)},$$

is heavy-tailed with tail index $\gamma=1/\theta$. In this model, $m_1(\theta)=\sqrt{\theta}\,\Gamma((\theta-1)/2)/(\sqrt{\pi}\,\Gamma(\theta/2))$ and $m_2(\theta)=\theta/(\theta-2)$ when $\theta>2$, so that again m_1 and m_2 are infinitely differentiable functions of θ on the interval $(2,\infty)$. Conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ then hold together if the function $\theta(\cdot)$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$ with $\theta(\boldsymbol{x})>2+\delta$.
• The Generalized Pareto model $F(y,\theta)=1-(1+\theta_1 y/\theta_2)^{-1/\theta_1}$ (for $y,\theta_1,\theta_2>0$), is heavy-tailed with tail index $\gamma=\theta_1$. In this model, $m_1(\theta_1,\theta_2)=\theta_2/(1-\theta_1)$ and $m_2(\theta_1,\theta_2)=2\theta_2^2/((1-\theta_1)(1-2\theta_1))$ when $\theta_1<1/2$, from which the functions m_1 and m_2 are infinitely differentiable functions of (θ_1,θ_2) on $(0,1/2)\times(0,\infty)$. Conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ then hold together if the function $\theta(\cdot)=(\theta_1(\cdot),\theta_2(\cdot))$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$ with $\theta_1(\boldsymbol{x})<1/(2+\delta)$.
• The Burr model $F(y,(\theta_1,\theta_2))=1-(1+y^{\theta_2/\theta_1})^{-1/\theta_2}$ (for $y,\theta_1,\theta_2>0$), is heavy-tailed with tail index $\gamma=\theta_1$. In this model, $m_1(\theta_1,\theta_2)=\theta_2^{-1}B((1-\theta_1)/\theta_2,1+\theta_1/\theta_2)$ and $m_2(\theta_1,\theta_2)=\theta_2^{-1}B((1-2\theta_1)/\theta_2,1+2\theta_1/\theta_2)$ when $\theta_1<1/2$, where B again denotes the Beta function. The Beta function is infinitely differentiable in both its arguments, so m_1 and m_2 are infinitely differentiable functions of (θ_1,θ_2) on $(0,1/2)\times(0,\infty)$. Conditions $\mathcal{H}_\delta$, $\mathcal{L}_m$ and $\mathcal{B}_m$ will hold together if the function $\theta(\cdot)=(\theta_1(\cdot),\theta_2(\cdot))$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$ with $\theta_1(\boldsymbol{x})<1/(2+\delta)$.

Condition $\mathcal{D}_m$ will similarly be satisfied if m_1 and m_2 are twice continuously differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}(\cdot)$ has a Lipschitz continuous gradient at $\boldsymbol{x}$ (a twice continuously differentiable mapping $\boldsymbol{\theta}(\cdot)$ in a neighborhood of $\boldsymbol{x}$ will satisfy this last assumption). If m_1 and/or m_2 have no simple closed form, a workaround is to use their (Choquet) integral expression,

$$m_1(\boldsymbol{\theta})=\int_0^\infty \overline{F}(y,\boldsymbol{\theta})\,dy\ \text{ and }\ m_2(\boldsymbol{\theta})=\int_0^\infty 2y\,\overline{F}(y,\boldsymbol{\theta})\,dy.$$

To check condition $\mathcal{L}_m$ (resp. $\mathcal{D}_m$), in addition to the regularity requirements on $\boldsymbol{\theta}(\cdot)$, it is thus sufficient to ascertain whether the parameter-dependent integrals $\int_0^\infty \overline{F}(y,\boldsymbol{\theta})\,dy$ and $\int_0^\infty y\,\overline{F}(y,\boldsymbol{\theta})\,dy$ exist, are finite, and continuously differentiable (resp. twice continuously differentiable) with respect to $\boldsymbol{\theta}$. This is most easily checked using general results on differentiation under the integral sign.

We finally examine the validity of conditions $\mathcal{L}_\omega$, $\mathcal{D}_\omega$ and $\mathcal{B}_\Omega$. Here

$$\frac{1}{\log(y)}\left|\log\frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})}\right|=\frac{1}{\log(y)}\left|\log\frac{\overline{F}(y,\boldsymbol{\theta}(\boldsymbol{x}'))}{\overline{F}(y,\boldsymbol{\theta}(\boldsymbol{x}))}\right|.$$

Models where $\boldsymbol{\theta}\mapsto\overline{F}(y,\boldsymbol{\theta})$ is continuously differentiable are very convenient, for then

$$\frac{1}{\log(y)}\left|\log\frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})}\right|\le\|\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{x}')\|\sup_{\boldsymbol{t}\in[\boldsymbol{\theta}(\boldsymbol{x}),\boldsymbol{\theta}(\boldsymbol{x}')]}\left\|\frac{\nabla_{\boldsymbol{\theta}}\log\overline{F}(y,\boldsymbol{t})}{\log(y)}\right\|$$

by the mean value theorem, where $[\boldsymbol{\theta}(\boldsymbol{x}),\boldsymbol{\theta}(\boldsymbol{x}')]$ denotes the segment in Θ linking $\boldsymbol{\theta}(\boldsymbol{x})$ to $\boldsymbol{\theta}(\boldsymbol{x}')$. Condition $\mathcal{L}_\omega$ then reduces to Lipschitz continuity of the function $\boldsymbol{\theta}$ at $\boldsymbol{x}$ and local boundedness of the above partial derivative, uniformly in y large enough, that is, there exist $y_0 > 0$ and a neighborhood W of $\boldsymbol{\theta}(\boldsymbol{x})$ with

$$\sup_{y\geq y_0}\sup_{\boldsymbol{t}\in W}\left\|\frac{\nabla_{\boldsymbol{\theta}}\log\overline{F}(y,\boldsymbol{t})}{\log(y)}\right\| = \sup_{y\geq y_0}\sup_{\boldsymbol{t}\in W}\frac{1}{\log(y)}\,\frac{\|\nabla_{\boldsymbol{\theta}}\overline{F}(y,\boldsymbol{t})\|}{\overline{F}(y,\boldsymbol{t})} < \infty.$$

This local boundedness property of course holds as soon as

$$\exists y_0 > 0,\ \forall\boldsymbol{\theta}\in\Theta,\ \sup_{y\geq y_0}\left\|\frac{\nabla_{\boldsymbol{\theta}}\log\overline{F}(y,\boldsymbol{\theta})}{\log(y)}\right\| = \sup_{y\geq y_0}\frac{1}{\log(y)}\,\frac{\|\nabla_{\boldsymbol{\theta}}\overline{F}(y,\boldsymbol{\theta})\|}{\overline{F}(y,\boldsymbol{\theta})} \leq \kappa(\boldsymbol{\theta})$$

where κ is continuous on the parameter space Θ. For instance:

- In the Fréchet model $F(y,\theta) = \exp(-y^{-1/\theta})$ (for $y,\theta > 0$),

$$\begin{aligned}\forall y > 1,\ \frac{1}{\log(y)}\frac{\partial\log\overline{F}}{\partial\theta}(y,\theta) &= \frac{1}{\theta^2}\times\frac{y^{-1/\theta}}{1-\exp(-y^{-1/\theta})}\times\exp(-y^{-1/\theta})\\ &\leq \frac{\max_{[0,1]}\psi}{\theta^2},\ \text{ with } \psi(z) = \frac{z}{1-e^{-z}}.\end{aligned}$$

[The upper bound is finite because ψ is everywhere continuous.] Then condition $\mathcal{L}_\omega$ is satisfied as soon as $\theta(\cdot)$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$.

- In the absolute Student model with θ degrees of freedom, that is,

$$\overline{F}(y,\theta) = \int_y^\infty D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}dz,\ \text{with } D(\theta) = \frac{2\,\Gamma((\theta+1)/2)}{\sqrt{\theta\pi}\,\Gamma(\theta/2)}\ \text{(for } y,\theta>0),$$

one has, for any $y > 1$,

$$\begin{aligned}\overline{F}(y,\theta) &= \theta^{(\theta+1)/2}D(\theta)\int_y^\infty z^{-\theta-1}\left(1+\frac{\theta}{z^2}\right)^{-(\theta+1)/2}dz\\ &\leq \theta^{(\theta-1)/2}D(\theta)y^{-\theta} = \kappa_1(\theta)y^{-\theta}\\ \text{and } \overline{F}(y,\theta) &\geq \theta^{(\theta-1)/2}D(\theta)\times\left(1+\frac{\theta}{y^2}\right)^{-(\theta+1)/2}y^{-\theta}\\ &\geq \theta^{(\theta-1)/2}D(\theta)(1+\theta)^{-(\theta+1)/2}y^{-\theta} = \kappa_2(\theta)y^{-\theta}\end{aligned}\tag{25}$$

where κ_1 and κ_2 are continuous and (strictly) positive on $(0,\infty)$. [Successive functions κ_j, here and in the verification of condition $\mathcal{D}_\omega$ below, will similarly be continuous without further mention.] Moreover, the function D is infinitely differentiable, and

$$\frac{\partial}{\partial\theta}\left[\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}\right] = \left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta}-\frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}.$$

Then on any interval of the form $[\theta_*,\theta^*]\subset(0,\infty)$, and for any $z > 0$,

$$\left|\frac{\partial}{\partial\theta}\left[\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}\right]\right| \leq \left(\frac{\theta_*+1}{2\theta_*}+\frac{1}{2}\log\left(1+\frac{z^2}{\theta_*}\right)\right)\left(1+\frac{z^2}{\theta^*}\right)^{-(\theta_*+1)/2}.$$

This obviously defines an integrable function of z on $(0,\infty)$, so $\overline{F}(y,\theta)$ is continuously differentiable with respect to θ under the integral, and

$$\begin{aligned}&\frac{\partial\overline{F}}{\partial\theta}(y,\theta)\\ &= \frac{D'(\theta)}{D(\theta)}\overline{F}(y,\theta)+\int_y^\infty\left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta}-\frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}dz.\end{aligned}$$

Hence the bound

$$\left|\frac{\partial\overline{F}}{\partial\theta}(y,\theta)+\log(y)\overline{F}(y,\theta)\right|\le\left(\frac{|D'(\theta)|}{D(\theta)}+\frac{\theta+1}{2\theta}\right)\overline{F}(y,\theta)$$

$$+\left|\frac{D(\theta)}{2}\int_y^\infty\left[\log\left(1+\frac{z^2}{\theta}\right)-2\log(y)\right]\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}dz\right|. \tag{26}$$

Writing, for $z>y>1$, $|\log(1+z^2/\theta)-2\log(y)|=|2\log(z/y)-\log\theta+\log(1+\theta/z^2)|\le 2\log(z/y)+|\log\theta|+\theta$, it follows that, for any $y>1$,

$$0\le\frac{D(\theta)}{2}\int_y^\infty\left|\log\left(1+\frac{z^2}{\theta}\right)-2\log(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}dz$$

$$\le\frac{|\log\theta|+\theta}{2}\overline{F}(y,\theta)+\theta^{(\theta+1)/2}D(\theta)\int_y^\infty\log(z/y)z^{-\theta}\,\frac{dz}{z}$$

$$=\frac{|\log\theta|+\theta}{2}\overline{F}(y,\theta)+\theta^{(\theta-3)/2}D(\theta)\times y^{-\theta}\le\kappa_3(\theta)\overline{F}(y,\theta) \tag{27}$$

by (25). [Here $\kappa_3(\theta)=(|\log\theta|+\theta)/2+\theta^{(\theta-3)/2}D(\theta)/\kappa_2(\theta)$.] Combine (26) and (27) to find, for any $y>1$,

$$\left|\frac{\partial\overline{F}}{\partial\theta}(y,\theta)+\log(y)\overline{F}(y,\theta)\right|\le\kappa_4(\theta)\overline{F}(y,\theta). \tag{28}$$

This readily yields

$$\sup_{y>2}\frac{1}{\log(y)}\times\frac{1}{\overline{F}(y,\theta)}\left|\frac{\partial\overline{F}}{\partial\theta}(y,\theta)\right|\le\kappa(\theta)$$

where κ is continuous on $(0,\infty)$. Then condition $\mathcal{L}_\omega$ is satisfied as soon as $\theta(\cdot)$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$.

- In the Generalized Pareto model $F(y,(\theta_1,\theta_2))=1-(1+\theta_1 y/\theta_2)^{-1/\theta_1}$ (for $y,\theta_1,\theta_2>0$),

$$\frac{\nabla_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))}{\log(y)}=\frac{1}{\log(y)}\left(\frac{1}{\theta_1^2}\log(1+\theta_1 y/\theta_2)-\frac{y}{\theta_1(\theta_2+\theta_1 y)},\ \frac{y}{\theta_2(\theta_2+\theta_1 y)}\right)^\top.$$

Writing, for any $y>1$, $|\log(1+\theta_1 y/\theta_2)|=|\log(\theta_1/\theta_2)+\log(y)+\log(1+\theta_2 y^{-1}/\theta_1)|\le|\log(\theta_2/\theta_1)|+\log(y)+\theta_2/\theta_1$, we find

$$\exists C>0,\ \sup_{y>2}\frac{\|\nabla_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))\|}{\log(y)}\le C\left(\frac{1+|\log(\theta_2/\theta_1)|+\theta_2/\theta_1}{\theta_1^2}+\frac{1}{\theta_2^2}\right).$$

Then condition $\mathcal{L}_\omega$ is satisfied as soon as $\boldsymbol{\theta}(\cdot)=(\theta_1(\cdot),\theta_2(\cdot))$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$.

- In the Burr model $F(y,(\theta_1,\theta_2))=1-(1+y^{\theta_2/\theta_1})^{-1/\theta_2}$ (for $y,\theta_1,\theta_2>0$),

$$\frac{\nabla_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))}{\log(y)}=\left(\frac{1}{\theta_1^2}\times\frac{1}{1+y^{-\theta_2/\theta_1}},\ \frac{1}{\theta_2^2}\frac{\log(1+y^{\theta_2/\theta_1})}{\log(y)}-\frac{1}{\theta_1\theta_2}\times\frac{1}{1+y^{-\theta_2/\theta_1}}\right)^\top.$$

Writing $|\log(1+y^{\theta_2/\theta_1})|=|(\theta_2/\theta_1)\log(y)+\log(1+y^{-\theta_2/\theta_1})|\le(\theta_2/\theta_1)\log(y)+1$ for any $y>1$, we find

$$\exists C>0,\ \sup_{y>2}\frac{\|\nabla_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))\|}{\log(y)}\le C\left(\frac{1}{\theta_1^2}+\frac{1}{\theta_2^2}\right).$$

Then condition $\mathcal{L}_\omega$ is satisfied as soon as $\boldsymbol{\theta}(\cdot)=(\theta_1(\cdot),\theta_2(\cdot))$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$.

Checking condition $\mathcal{D}_\omega$ again requires extra regularity assumptions. Notice that the chain rule for a function $\phi : \boldsymbol{x} \mapsto f(u(\boldsymbol{x}))$, where $f : \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{u} : \mathbb{R}^p \to \mathbb{R}^d$ are twice continuously differentiable, yields

$$H\phi(\boldsymbol{x}) = [J\boldsymbol{u}(\boldsymbol{x})]^\top Hf(\boldsymbol{u}(\boldsymbol{x}))J\boldsymbol{u}(\boldsymbol{x}) + H\boldsymbol{u}(\boldsymbol{x}) \times_1 \nabla f(\boldsymbol{u}(\boldsymbol{x})).$$

Here $H\boldsymbol{u}(\boldsymbol{x}) \times_1 \nabla f(\boldsymbol{u}(\boldsymbol{x}))$ denotes the 1-mode product of the (third-order tensor) Hessian of $\boldsymbol{u}$ by the gradient vector $\nabla f(\boldsymbol{u}(\boldsymbol{x}))$: in other words, if $\boldsymbol{u} = (u_1, \ldots, u_d)$ where each u_i is real-valued, this 1-mode product has (i,j)th element

$$[H\boldsymbol{u}(\boldsymbol{x}) \times_1 \nabla f(\boldsymbol{u}(\boldsymbol{x}))]_{i,j} = \sum_{k=1}^{d} \frac{\partial^2 u_k}{\partial x_i \partial x_j}(\boldsymbol{x}) \frac{\partial f}{\partial x_k}(\boldsymbol{u}(\boldsymbol{x}))$$

for any $1 \le i,j \le p$. As a consequence, if $\boldsymbol{\theta} \mapsto \gamma(\boldsymbol{\theta})$ is twice continuously differentiable on Θ (*i.e.* the tail index is a twice continuously differentiable function of the parameters), $\boldsymbol{\theta}(\cdot)$ is twice continuously differentiable in a suitable neighborhood of $\boldsymbol{x}$ and there exist $y_0 > 0$ and a neighborhood W of $\boldsymbol{\theta}(\boldsymbol{x})$ with

$$\sup_{y \ge y_0} \sup_{\boldsymbol{t} \in W} \left\{ \left\| \frac{\nabla_{\boldsymbol{\theta}} \log \overline{F}(y, \boldsymbol{t})}{\log(y)} \right\| + \left\| \frac{H_{\boldsymbol{\theta}} \log \overline{F}(y, \boldsymbol{t})}{\log(y)} \right\| \right\} < \infty,$$

then the partial Hessian matrix $H_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}')/\log(y)$ is uniformly bounded in y large enough and $\boldsymbol{x}'$ in a neighborhood of $\boldsymbol{x}$. Lemma A.4(ii) then shows that condition $\mathcal{D}_\omega$ will be satisfied, with

$$\lim_{y \to \infty} \frac{\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})}{\log(y)} = \frac{\nabla(\boldsymbol{x} \mapsto \gamma(\boldsymbol{\theta}(\boldsymbol{x})))}{\gamma^2(\boldsymbol{\theta}(\boldsymbol{x}))} = \left(\frac{\nabla \gamma(\boldsymbol{\theta}(\boldsymbol{x}))}{\gamma^2(\boldsymbol{\theta}(\boldsymbol{x}))} \right)^\top J\boldsymbol{\theta}(\boldsymbol{x}).$$

Again, the above boundedness assumption holds under the stronger property that

$$\exists y_0 > 0, \ \forall \boldsymbol{\theta} \in \Theta, \ \sup_{y \ge y_0} \left\{ \frac{1}{\log(y)} \left(\|\nabla_{\boldsymbol{\theta}} \log \overline{F}(y, \boldsymbol{\theta})\| + \|H_{\boldsymbol{\theta}} \log \overline{F}(y, \boldsymbol{\theta})\| \right) \right\} \le \kappa(\boldsymbol{\theta})$$

where κ is continuous on the parameter space Θ. In our above examples:

- In the Fréchet model $F(y, \theta) = \exp(-y^{-1/\theta})$ (for $y, \theta > 0$), recall the notation $\psi(z) = z/(1 - e^{-z})$. The function ψ is everywhere continuously differentiable, converges to 1 as $z \to 0$, and

$$\forall y > 1, \ \frac{1}{\log(y)} \frac{\partial \log \overline{F}}{\partial \theta}(y, \theta) = \frac{\psi(y^{-1/\theta})}{\theta^2} \exp(-y^{-1/\theta}).$$

$$\begin{aligned} \forall y > 1, \ \frac{1}{\log(y)} &\frac{\partial^2 \log \overline{F}}{\partial \theta^2}(y, \theta) \\ &= \frac{1}{\theta^4} \left(-2\theta\psi(y^{-1/\theta}) + y^{-1/\theta} \log(y) \times (\psi' - \psi)(y^{-1/\theta}) \right) \exp(-y^{-1/\theta}). \end{aligned}$$

Elementary calculus shows that, for any $\theta > 0$, $|y^{-1/\theta} \log(y)| \le \theta/e$ for any $y > 1$. Hence the inequality

$$\forall y > 1, \ \frac{1}{\log(y)} \left| \frac{\partial^2 \log \overline{F}}{\partial \theta^2}(y, \theta) \right| = \frac{1}{\theta^3} \left(2 \max_{[0,1]} \psi + \frac{\max_{[0,1]} |\psi' - \psi|}{e} \right).$$

Conclude that condition $\mathcal{D}_\omega$ is satisfied as soon as $\theta(\cdot)$ is twice continuously differentiable in a neighborhood of $\boldsymbol{x}$. [Here, as $y \to \infty$, $(\partial \log \overline{F}/\partial\theta)(y, \theta)/\log(y) \to 1/\theta^2 = \gamma'(\theta)/\gamma^2(\theta)$, since $\gamma(\theta) = \theta$.]

- In the absolute Student model with θ degrees of freedom,

$$\overline{F}(y,\theta) = \int_y^\infty D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz, \text{ with } D(\theta) = \frac{2\,\Gamma((\theta+1)/2)}{\sqrt{\theta\pi}\,\Gamma(\theta/2)} \text{ (for } y,\theta>0),$$

We may show, just as in our earlier analysis, that $\theta \mapsto \overline{F}(y,\theta)$ is twice continuously differentiable under the integral. Straightforward calculations yield

$$\begin{aligned}
&\frac{\partial^2 \overline{F}}{\partial\theta^2}(y,\theta) \\
&= \left(\frac{D''(\theta)}{D(\theta)} - \left(\frac{D'(\theta)}{D(\theta)}\right)^2\right)\overline{F}(y,\theta) + \frac{D'(\theta)}{D(\theta)}\frac{\partial \overline{F}}{\partial\theta}(y,\theta) \\
&+ \frac{\partial}{\partial\theta}\int_y^\infty \left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta} - \frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right) D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz \\
&= \frac{D''(\theta)}{D(\theta)}\overline{F}(y,\theta) + 2\frac{D'(\theta)}{D(\theta)}\left(\frac{\partial \overline{F}}{\partial\theta}(y,\theta) - \frac{D'(\theta)}{D(\theta)}\overline{F}(y,\theta)\right) \\
&+ \int_y^\infty \left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta} - \frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)^2 D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz \\
&+ \int_y^\infty \left(-\frac{1}{2\theta^2}\frac{z^2/\theta}{1+z^2/\theta} - \frac{\theta+1}{2\theta^2}\frac{z^2/\theta}{(1+z^2/\theta)^2} + \frac{1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta}\right) D(\theta)\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz.
\end{aligned}$$

Then, for any $y>1$,

$$\begin{aligned}
&\left|\frac{\partial^2 \overline{F}}{\partial\theta^2}(y,\theta) - \log^2(y)\overline{F}(y,\theta)\right| \\
&\le \left(\frac{|D''(\theta)|}{D(\theta)} + 2\left[\frac{D'(\theta)}{D(\theta)}\right]^2 + \frac{\theta+1}{\theta^2}\right)\overline{F}(y,\theta) + 2\frac{|D'(\theta)|}{D(\theta)}\left|\frac{\partial \overline{F}}{\partial\theta}(y,\theta)\right| \\
&+ \int_y^\infty D(\theta)\left|\left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta} - \frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)^2 - \log^2(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz.
\end{aligned}$$

Expanding the square in the last integral, we find

$$\begin{aligned}
&\int_y^\infty D(\theta)\left|\left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta} - \frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)^2 - \log^2(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz \\
&\le \int_y^\infty D(\theta)\left|\left(\frac{\theta+1}{2\theta}\frac{z^2/\theta}{1+z^2/\theta} - \frac{1}{2}\log\left(1+\frac{z^2}{\theta}\right)\right)^2 - \log^2(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz \\
&\le \frac{(\theta+1)^2}{4\theta^2}\overline{F}(y,\theta) + \frac{1}{4}\int_y^\infty D(\theta)\left|\log^2\left(1+\frac{z^2}{\theta}\right) - 4\log^2(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz \\
&+ \frac{\theta+1}{\theta}\times\frac{D(\theta)}{2}\int_y^\infty \left|\log\left(1+\frac{z^2}{\theta}\right) - 2\log(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2} dz + \frac{\theta+1}{\theta}\log(y)\overline{F}(y,\theta).
\end{aligned}$$

Using (27) and (28), we then obtain, for any $y>1$,

$$\left|\frac{\partial^2\overline{F}}{\partial\theta^2}(y,\theta)-\log^2(y)\overline{F}(y,\theta)\right|\le(\kappa_5(\theta)+\kappa_6(\theta)\log(y))\overline{F}(y,\theta)$$
$$+\frac{1}{4}\int_y^\infty D(\theta)\left|\log^2\left(1+\frac{z^2}{\theta}\right)-4\log^2(y)\right|\left(1+\frac{z^2}{\theta}\right)^{-(\theta+1)/2}dz.$$

The exact same arguments that led to (27) then entail, after somewhat cumbersome calculations,

$$\forall y>1,\ \left|\frac{\partial^2\overline{F}}{\partial\theta^2}(y,\theta)-\log^2(y)\overline{F}(y,\theta)\right|\le(\kappa_7(\theta)+\kappa_8(\theta)\log(y))\overline{F}(y,\theta).\tag{29}$$

Besides, recall (28) to get

$$\forall y>1,\ \left|\left(\frac{\partial\log\overline{F}}{\partial\theta}(y,\theta)\right)^2-\log^2(y)\right|$$
$$=\frac{1}{\overline{F}(y,\theta)}\left|\frac{\partial\overline{F}}{\partial\theta}(y,\theta)-\log(y)\overline{F}(y,\theta)\right|\times\frac{1}{\overline{F}(y,\theta)}\left|\frac{\partial\overline{F}}{\partial\theta}(y,\theta)+\log(y)\overline{F}(y,\theta)\right|$$
$$\le\kappa_9(\theta)+\kappa_{10}(\theta)\log(y).\tag{30}$$

Combine (29) and (30) with the identity

$$\frac{\partial^2\log\overline{F}}{\partial\theta^2}(y,\theta)=\left(\frac{1}{\overline{F}(y,\theta)}\frac{\partial^2\overline{F}}{\partial\theta^2}(y,\theta)-\log^2(y)\right)-\left(\left(\frac{\partial\log\overline{F}}{\partial\theta}(y,\theta)\right)^2-\log^2(y)\right)$$

to get

$$\sup_{y>2}\frac{1}{\log(y)}\left|\frac{\partial^2\log\overline{F}}{\partial\theta^2}(y,\theta)\right|\le\kappa(\theta)$$

where κ is continuous on $(0,\infty)$. Condition $\mathcal{D}_\omega$ is then satisfied as soon as $\theta(\cdot)$ is twice continuously differentiable in a neighborhood of $\boldsymbol{x}$. [Here, as $y\to\infty$, $(\partial\log\overline{F}/\partial\theta)(y,\theta)/\log(y)\to -1=\gamma'(\theta)/\gamma^2(\theta)$, since $\gamma(\theta)=1/\theta$.]

- In the Generalized Pareto model $F(y,(\theta_1,\theta_2))=1-(1+\theta_1 y/\theta_2)^{-1/\theta_1}$ (for $y,\theta_1,\theta_2>0$),

$$\frac{H_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))}{\log(y)}=\frac{1}{\log(y)}\begin{pmatrix}-\dfrac{2}{\theta_1^3}\log(1+\theta_1y/\theta_2)+\dfrac{y(2\theta_2+3\theta_1y)}{\theta_1^2(\theta_2+\theta_1y)^2} & -\dfrac{y^2}{\theta_2(\theta_2+\theta_1y)^2}\\ -\dfrac{y^2}{\theta_2(\theta_2+\theta_1y)^2} & -\dfrac{y(2\theta_2+\theta_1y)}{\theta_2^2(\theta_2+\theta_1y)^2}\end{pmatrix}.$$

Use again the inequality $|\log(1+\theta_1y/\theta_2)|\le|\log(\theta_2/\theta_1)|+\log(y)+\theta_2/\theta_1$ for any $y>1$ to obtain

$$\exists C>0,\ \sup_{y>2}\frac{\|H_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))\|}{\log(y)}\le\frac{C}{\theta_1}\left(\frac{1+|\log(\theta_2/\theta_1)|+\theta_2/\theta_1}{\theta_1^2}+\frac{1}{\theta_2^2}\right).$$

Conclude that condition $\mathcal{D}_\omega$ holds if $\boldsymbol{\theta}(\cdot)=(\theta_1(\cdot),\theta_2(\cdot))$ is twice continuously differentiable in a neighborhood of $\boldsymbol{x}$. [Here, as $y\to\infty$, $\nabla_{(\theta_1,\theta_2)}\log\overline{F}(y,(\theta_1,\theta_2))/\log(y)\to(1/\theta_1^2,0)^\top=\nabla\gamma((\theta_1,\theta_2))/\gamma^2((\theta_1,\theta_2))$, since $\gamma((\theta_1,\theta_2))=\theta_1$.]

- In the Burr model $F(y,(\theta_1,\theta_2))=1-(1+y^{\theta_2/\theta_1})^{-1/\theta_2}$ (for $y,\theta_1,\theta_2>0$),

$$\frac{1}{\log(y)}\times\frac{\partial^2\log\overline{F}}{\partial\theta_1^2}(y,(\theta_1,\theta_2))=\frac{2}{\theta_1^3}\times\frac{1}{1+y^{-\theta_2/\theta_1}}-\frac{\theta_2}{\theta_1^4}\times\frac{y^{-\theta_2/\theta_1}\log(y)}{(1+y^{-\theta_2/\theta_1})^2},$$

$$\frac{1}{\log(y)} \times \frac{\partial^2 \log \overline{F}}{\partial \theta_1 \theta_2}(y, (\theta_1, \theta_2)) = \frac{1}{\theta_1^3} \times \frac{y^{-\theta_2/\theta_1} \log(y)}{(1 + y^{-\theta_2/\theta_1})^2},$$

$$\frac{1}{\log(y)} \times \frac{\partial^2 \log \overline{F}}{\partial \theta_2^2}(y, (\theta_1, \theta_2)) = -\frac{2}{\theta_2^3} \frac{\log(1 + y^{\theta_2/\theta_1})}{\log(y)} + \frac{1}{\theta_1^2 \theta_2} \frac{1}{1 + y^{-\theta_2/\theta_1}} \left(2 - \frac{y^{-\theta_2/\theta_1} \log(y)}{1 + y^{-\theta_2/\theta_1}} \right).$$

Recall the inequalities $|\log(1+y^{\theta_2/\theta_1})| \leq (\theta_2/\theta_1)\log(y)+1$ and $|y^{-\theta_2/\theta_1}\log(y)| \leq \theta_1/(e\theta_2)$ for any $y > 1$ to get

$$\exists C > 0, \ \sup_{y>2} \frac{\|H_{(\theta_1,\theta_2)} \log \overline{F}(y, (\theta_1, \theta_2))\|}{\log(y)} \leq \frac{C}{\theta_1} \left(\frac{1}{\theta_1^2} + \frac{1}{\theta_2^2} \right).$$

Conclude that condition $\mathcal{D}_\omega$ holds if $\boldsymbol{\theta}(\cdot) = (\theta_1(\cdot), \theta_2(\cdot))$ is twice continuously differentiable in a neighborhood of $\boldsymbol{x}$. [Here, as $y \to \infty$, $\nabla_{(\theta_1,\theta_2)} \log \overline{F}(y, (\theta_1, \theta_2))/\log(y) \to (1/\theta_1^2, 0)^\top = \nabla\gamma((\theta_1, \theta_2))/\gamma^2((\theta_1, \theta_2))$, since $\gamma((\theta_1, \theta_2)) = \theta_1$.]

Meanwhile, we always have

$$\begin{aligned} &\frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y' | \boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}} \\ &= \frac{\mathbb{P}(U_1 > F(y, \boldsymbol{\theta}(\boldsymbol{x}')), U_{t+1} > F(y', \boldsymbol{\theta}(\boldsymbol{x}'')))}{\sqrt{[1 - F(y, \boldsymbol{\theta}(\boldsymbol{x}'))][1 - F(y', \boldsymbol{\theta}(\boldsymbol{x}''))]}} \leq 1 \end{aligned}$$

so that condition $\mathcal{B}_\Omega$ is satisfied in this nonlinear regression model as well.

Autoregressive model. Set $\boldsymbol{X}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\top$ and fix $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^\top \in \mathbb{R}^p$. Then

$$Y_t | \boldsymbol{X}_t = \boldsymbol{x} \stackrel{d}{=} \varepsilon + \sum_{j=1}^p \phi_j x_j, \text{ so } \overline{F}(y|\boldsymbol{x}) = \overline{F}_\varepsilon \left(y - \sum_{j=1}^p \phi_j x_j \right).$$

Under our stated assumptions on ε, (Y_t) is both geometrically $\beta-$mixing [20, Theorem 6 p.99] and, since the causal and invertible AR(p) process can be represented as a linear time series with geometrically decaying coefficients [7, proof of Theorem 3.1.1 p.85], it is also geometrically $\rho-$mixing [3, p.18]. Hence condition $\mathcal{M}$ on $((\boldsymbol{X}_t, Y_t))_{t\geq 1}$ holds (for the validity of the second-order regular variation property, see our above discussion in the location-scale model). Condition $\mathcal{H}_\delta$ holds when $\gamma < 1/(2+\delta)$ and $\mathbb{E}(\varepsilon_-^{2+\delta}) < \infty$.

To check the other assumptions, it will be convenient to rewrite the AR(p) model in vector form, more precisely as $\boldsymbol{Y}_t^{(p)} = A\boldsymbol{Y}_{t-1}^{(p)} + \boldsymbol{\varepsilon}_t^{(p)}$ with

$$\boldsymbol{Y}_t^{(p)} = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}, \ \boldsymbol{\varepsilon}_t^{(p)} = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } A = \begin{pmatrix} \phi_1 & \cdots & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

The previous equation is equivalently rewritten $\boldsymbol{X}_{t+1} = A\boldsymbol{X}_t + \boldsymbol{\varepsilon}_t^{(p)}$, and in particular $\boldsymbol{X}_{p+1} = A^p \boldsymbol{X}_1 + \sum_{j=1}^p A^{p-j} \boldsymbol{\varepsilon}_j^{(p)}$. Letting $\boldsymbol{e}_j$ be the jth vector in the canonical basis of $\mathbb{R}^p$, and noting that $\boldsymbol{\varepsilon}_j^{(p)} = \varepsilon_j \boldsymbol{e}_1$ and $A\boldsymbol{e}_j = \phi_j \boldsymbol{e}_1 + \boldsymbol{e}_{j+1}$ for any $j \in \{1, \ldots, p-1\}$, a straightforward calculation shows that there are real constants $b_{i,j}$ $(1 \le i \le j \le p-i)$ such that

$$\boldsymbol{Z}_p = \sum_{j=1}^{p} A^{p-j} \boldsymbol{\varepsilon}_j^{(p)} = \begin{pmatrix} \varepsilon_p + \sum_{j=1}^{p-1} b_{1,j}\varepsilon_j \\ \varepsilon_{p-1} + \sum_{j=1}^{p-2} b_{2,j}\varepsilon_j \\ \vdots \\ \varepsilon_2 + b_{p-1,1}\varepsilon_1 \\ \varepsilon_1 \end{pmatrix}.$$

This entails that $\boldsymbol{Z}_p$ has a p.d.f. $f_{\boldsymbol{Z}_p}$ on $\mathbb{R}^p$, given by

$$f_{\boldsymbol{Z}_p}(z_1, \ldots, z_p) = \prod_{i=1}^{p} f_\varepsilon \left(z_i - \sum_{j=1}^{p-i} b_{i,j} z_{p-j+1} \right).$$

Under our assumptions on f_ε, this defines a Lipschitz continuous, everywhere (strictly) positive and bounded function on $\mathbb{R}^p$. Recall now that if $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ are independent $\mathbb{R}^p-$valued random vectors, the joint distribution functions $F_{\boldsymbol{Z}+\boldsymbol{Z}'}$, $F_{\boldsymbol{Z}}$ and $F_{\boldsymbol{Z}'}$ of $\boldsymbol{Z}+\boldsymbol{Z}'$, $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ are linked by the convolution equation

$$F_{\boldsymbol{Z}+\boldsymbol{Z}'}(\boldsymbol{y}) = \int_{\mathbb{R}^p} F_{\boldsymbol{Z}}(\boldsymbol{y} - \boldsymbol{z}) \, dF_{\boldsymbol{Z}'}(\boldsymbol{z}).$$

If $\boldsymbol{Z}$ moreover has a p.d.f. $f_{\boldsymbol{Z}}$ on $\mathbb{R}^p$, writing

$$F_{\boldsymbol{Z}}(\boldsymbol{y} - \boldsymbol{z}) = \int_{\boldsymbol{t} \le \boldsymbol{y} - \boldsymbol{z}} f_{\boldsymbol{Z}}(\boldsymbol{t}) \, d\boldsymbol{t} = \int_{\boldsymbol{u} \le \boldsymbol{y}} f_{\boldsymbol{Z}}(\boldsymbol{u} - \boldsymbol{z}) \, d\boldsymbol{u}$$

(where the inequalities are to be understood componentwise) entails, by the Tonelli theorem, that $\boldsymbol{Z}+\boldsymbol{Z}'$ also has a p.d.f. $f_{\boldsymbol{Z}+\boldsymbol{Z}'}$ on $\mathbb{R}^p$, given by

$$f_{\boldsymbol{Z}+\boldsymbol{Z}'}(\boldsymbol{u}) = \int_{\mathbb{R}^p} f_{\boldsymbol{Z}}(\boldsymbol{u} - \boldsymbol{z}) \, dF_{\boldsymbol{Z}'}(\boldsymbol{z}). \tag{31}$$

Applying this to $\boldsymbol{Z} = \boldsymbol{Z}_p$ and $\boldsymbol{Z}' = A^p \boldsymbol{X}_1$, it follows that $\boldsymbol{Z}+\boldsymbol{Z}' = \boldsymbol{X}_{p+1} \stackrel{d}{=} \boldsymbol{X}_1$ has a p.d.f. g on $\mathbb{R}^p$, given by

$$g(\boldsymbol{u}) = f_{\boldsymbol{X}_1}(\boldsymbol{u}) = \int_{\mathbb{R}^p} f_{\boldsymbol{Z}_p}(\boldsymbol{u} - \boldsymbol{z}) \, dF_{A^p \boldsymbol{X}_1}(\boldsymbol{z}).$$

As a consequence, g is continuous and everywhere (strictly) positive because $f_{\boldsymbol{Z}_p}$ is so, is bounded by $\sup_{\mathbb{R}^p} f_{\boldsymbol{Z}_p} \le (\sup_{\mathbb{R}} f_\varepsilon)^p < \infty$, and, if c is the Lipschitz constant of $f_{\boldsymbol{Z}_p}$, then g is Lipschitz continuous with this same constant c, and therefore condition $\mathcal{L}_g$ holds. To verify condition $\mathcal{D}_g$, note that if f_ε has a uniformly bounded derivative, then $f_{\boldsymbol{Z}_p}$ has a uniformly bounded gradient, and so one may differentiate under the integral to obtain

$$\nabla g(\boldsymbol{u}) = \int_{\mathbb{R}^p} \nabla f_{\boldsymbol{Z}_p}(\boldsymbol{u} - \boldsymbol{z}) \, dF_{A^p \boldsymbol{X}_1}(\boldsymbol{z}).$$

Conclude similarly that ∇g is Lipschitz continuous as soon as $\nabla f_{\boldsymbol{Z}_p}$ has a uniformly bounded gradient that is itself Lipschitz continuous, which follows immediately from assuming that f_ε has a uniformly bounded and Lipschitz continuous derivative.

Condition $\mathcal{D}_m$ automatically follows from the square integrability of ε and the linearity of the regression function.

We next check condition $\mathcal{B}_p$ with $t_0 = p$. We start by checking the first half of the condition, for which there is nothing to be done if $p = 1$. Otherwise, note that for any t

$$\{\boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)\} \subset \bigcap_{j=1}^{p} \{Y_{t-j+1} \in [x_j - r, x_j + r]\}.$$

Then, for any integer t with $1 \le t < t_0 = p$,

$$\{\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)\} \subset \{\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), Y_1 \in [x_t - r, x_t + r]\}.$$

It follows that

$$\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) \le \int_{B(\boldsymbol{x},r)} \left(\int_{x_t - r}^{x_t + r} f_\varepsilon \left(y - \sum_{j=1}^{p} \phi_j z_j \right) dy \right) g(\boldsymbol{z})\, d\boldsymbol{z}$$

and so

$$\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) \le 2 \sup_{\mathbb{R}} f_\varepsilon \times r \int_{B(\boldsymbol{x},r)} g(\boldsymbol{z})\, d\boldsymbol{z} = O(r^{p+1})$$

as $r \to 0$. This shows that the first part of condition $\mathcal{B}_p$ indeed holds. To check the second part of this condition, the key point is to note that, for $t \ge t_0 = p$,

$$\boldsymbol{X}_{t+1} = A^t \boldsymbol{X}_1 + \sum_{j=1}^{t} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)} = A^t \boldsymbol{X}_1 + \sum_{j=1}^{t-p} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)} + \sum_{j=t-p+1}^{t} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)}.$$

As a consequence, $\boldsymbol{X}_{t+1} | \boldsymbol{X}_1 = \boldsymbol{x}_1 \stackrel{d}{=} A^t \boldsymbol{x}_1 + \sum_{j=1}^{t-p} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)} + \sum_{j=t-p+1}^{t} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)}$. The random vector $\sum_{j=t-p+1}^{t} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)} \stackrel{d}{=} \boldsymbol{Z}_p$ has p.d.f. g and is independent of $\boldsymbol{Z}'_{t,p} = \sum_{j=1}^{t-p} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)}$, so that $\boldsymbol{X}_{t+1} | \boldsymbol{X}_1 = \boldsymbol{x}_1$ has, by (31), the p.d.f.

$$f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}_{t+1} | \boldsymbol{x}_1) = \int_{\mathbb{R}^p} g(\boldsymbol{x}_{t+1} - A^t \boldsymbol{x}_1 - \boldsymbol{z})\, dF_{\boldsymbol{Z}'_{t,p}}(\boldsymbol{z}). \tag{32}$$

In particular, the conditional p.d.f.s $f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}$ are uniformly bounded (in both variables and in $t \ge t_0 = p$) by $\sup_{\mathbb{R}^p} g < \infty$. Conclude that

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) &= \int_{B(\boldsymbol{x},r) \times B(\boldsymbol{x},r)} f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}_{t+1} | \boldsymbol{x}_1) g(\boldsymbol{x}_1)\, d\boldsymbol{x}_1\, d\boldsymbol{x}_{t+1} \\
&\le \left(\sup_{\mathbb{R}^p} g \right)^2 \int_{B(\boldsymbol{x},r) \times B(\boldsymbol{x},r)} d\boldsymbol{x}_1\, d\boldsymbol{x}_{t+1}
\end{aligned}$$

so that

$$\limsup_{r \to 0} \sup_{t \ge t_0} \frac{1}{r^{2p}} \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) < \infty$$

which shows that assumption $\mathcal{B}_p$ is satisfied.

Condition $\mathcal{B}_m$ is unnecessary since (Y_t) is $\rho-$mixing.

To check condition $\mathcal{L}_\omega$, note that

$$\frac{1}{\log(y)} \left| \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} \right| = \frac{1}{\log(y)} \left| \log \frac{\overline{F}_\varepsilon(y - \sum_{j=1}^{p} \phi_j x'_j)}{\overline{F}_\varepsilon(y - \sum_{j=1}^{p} \phi_j x_j)} \right|.$$

This is handled exactly as in the example of the location-scale regression model, and it follows that condition $\mathcal{L}_\omega$ is satisfied (and even that $\omega_{h_n}(y_n|\boldsymbol{x}) = o(h_n)$), or that condition $\mathcal{D}_\omega$ is satisfied with $\lim_{y\to\infty} \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})/\log(y) = 0$ under the relevant assumptions stated in the result.

Checking condition $\mathcal{B}_\Omega$ is harder than in our other examples, because the covariate corresponds to a lagged value of the response, and therefore the sequence $(\boldsymbol{X}_t)$ is not independent from the sequence (ε_t), contrary to what is found in the previous examples. To control $\Omega_h(z|\boldsymbol{x})$, since (Y_t) is a Markov chain of order p, we check all assumptions of Lemma A.1(iii). It is sufficient to do so with $t_0 = p$, since for $t \le t_0 = p$, the conditional probability in the numerator of $\Omega_h(z|\boldsymbol{x})$ is equal to 0 for z large enough and h small enough. We then start by proving that there exists $r > 0$ with

$$\inf_{t>p} \inf_{\boldsymbol{x}_1, \boldsymbol{x}_{t+1} \in B(\boldsymbol{x}, r)} f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1) > 0. \tag{33}$$

Our argument is based on Equation (32). Note first that A is essentially the companion matrix of the polynomial $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$, and it is a standard exercise to show that

$$\det(\lambda \mathrm{I}_p - A) = \lambda^p \left(1 - \sum_{j=1}^p \phi_j \lambda^{-j}\right) = \lambda^p P(1/\lambda).$$

Our basic assumption on the AR(p) model is that the polynomial P has all its roots outside the closed unit disk in $\mathbb{C}$, meaning that the above characteristic polynomial has all its roots inside the open unit disk in $\mathbb{C}$, that is, the spectral radius $\rho(A)$ of A is smaller than 1. In particular, for any $\boldsymbol{u} \in \mathbb{R}^p$,

$$\|A^t \boldsymbol{u}\| \le \|A^t\| \|\boldsymbol{u}\| = (\rho(A) + o(1))^t \|\boldsymbol{u}\|$$

as $t \to \infty$, owing to the well-known convergence $\|A^t\|^{1/t} \to \rho(A)$ as $t \to \infty$. This means that for any bounded subset C of $\mathbb{R}^p$, the sequence of iterates $(A^t C) = (\{A^t \boldsymbol{u}, \boldsymbol{u} \in C\})$ is itself contained in a bounded subset C' of $\mathbb{R}^p$. Fix then $r > 0$. Thanks to the above argument, there is a bounded set I such that

$$\forall t \ge 1,\ (\boldsymbol{x}_1, \boldsymbol{x}_{t+1} \in B(\boldsymbol{x}, r) \Rightarrow \boldsymbol{x}_{t+1} - A^t \boldsymbol{x}_1 \in I).$$

Using Equation (32), we find that (33) shall be proven provided we show that

$$\inf_{t>p} \inf_{\boldsymbol{u} \in I} \int_{\mathbb{R}^p} g(\boldsymbol{u} - \boldsymbol{z})\, dF_{\boldsymbol{Z}'_{t,p}}(\boldsymbol{z}) > 0. \tag{34}$$

Observe now that

$$\boldsymbol{Z}'_{t,p} = \sum_{j=1}^{t-p} A^{t-j} \boldsymbol{\varepsilon}_j^{(p)} \stackrel{d}{=} A^p \sum_{j=0}^{t-p-1} A^j \boldsymbol{\varepsilon}_j^{(p)} \to A^p \sum_{j=0}^{\infty} A^j \boldsymbol{\varepsilon}_j^{(p)} = \boldsymbol{Z}'_p \text{ in } L^2, \text{ as } t \to \infty.$$

It follows that $\boldsymbol{Z}'_{t,p} \stackrel{d}{\longrightarrow} \boldsymbol{Z}'_p$ as $t \to \infty$, and as such the family of probability measures induced by the $\boldsymbol{Z}'_{t,p}$ (for $t > p$) is tight. Choose then $R > 0$ such that $\mathbb{P}(\boldsymbol{Z}'_{t,p} \in B_R) \ge 1/2$ for any $t > p$, where B_R is the closed ball centered at the origin with radius R. We have, by (32),

$$\inf_{t>p} \inf_{\boldsymbol{u} \in I} \int_{\mathbb{R}^p} g(\boldsymbol{u} - \boldsymbol{z})\, dF_{\boldsymbol{Z}'_{t,p}}(\boldsymbol{z}) \ge \frac{1}{2} \inf_{\boldsymbol{v} \in I - B_R} g(\boldsymbol{v})$$

where $I - B_R = \{\boldsymbol{v} - \boldsymbol{v}', (\boldsymbol{v}, \boldsymbol{v}') \in I \times B_R\}$. Clearly $I - B_R$ is bounded and g is continuous and (strictly) positive on $\mathbb{R}^p$, so that $\inf_{I - B_R} g > 0$, which means in particular that (34) is indeed satisfied. It only remains to check that

$$\sup_{t>p} \sup_{\boldsymbol{x}_1, \boldsymbol{x}_{t+1} \in B(\boldsymbol{x}, r)} \sup_{y_1 \ge y_0} f_{\boldsymbol{X}_{t+1}|\boldsymbol{X}_1, Y_1}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_1, y_1) < \infty$$

for some large y_0. Here though $\boldsymbol{X}_{t+1} = (Y_t, Y_{t-1}, \ldots, Y_{t-p+1})^\top$, $\boldsymbol{X}_1 = (Y_0, Y_{-1}, \ldots, Y_{-p+1})^\top$ and (Y_t) is a Markov chain of order p, so that the distribution of

$$\boldsymbol{X}_{t+1} | \{\boldsymbol{X}_1 = (y_0, y_{-1}, \ldots, y_{-p+1})^\top, Y_1 = y_1\}$$

is nothing but the distribution of

$$\begin{aligned}
&\boldsymbol{X}_{t+1} | \{(Y_1, Y_0, \ldots, Y_{-p+2})^\top = (y_1, y_0, \ldots, y_{-p+2})^\top\} \\
&= \boldsymbol{X}_{t+1} | \{\boldsymbol{X}_2 = (y_1, y_0, \ldots, y_{-p+2})^\top\} \\
&\stackrel{d}{=} \boldsymbol{X}_{(t-1)+1} | \{\boldsymbol{X}_1 = (y_1, y_0, \ldots, y_{-p+2})^\top\}.
\end{aligned}$$

We already know that the p.d.f. of this conditional distribution is uniformly bounded in t by $\sup_{\mathbb{R}^p} g$, see (32). Conclude that condition $\mathcal{B}_\Omega$ holds by Lemma A.1(iii).

APPENDIX C: FURTHER IMPLEMENTATION DETAILS AND FINITE-SAMPLE RESULTS

C.1. Explicit formulations of the estimators used as part of tuning parameter selection, bias and variance correction. We explain here how we estimate the second-order parameters ρ and b, using estimators inspired by those of [23] and [31]. We describe first the estimators in the unconditional setting, and we then explain how we adapt them to our nonparametric regression context. We start by the estimation of ρ. For a given sample $Z_1, \ldots, Z_N$, define

$$M_{\kappa_N}^{(j)} = \frac{1}{\kappa_N} \sum_{i=1}^{\kappa_N} (\log Z_{N-i+1,N} - \log Z_{N-\kappa_N,N})^j, \text{ for } j = 1, 2, 3.$$

Taking $j = 1$ yields the Hill estimator. The quantities $M_{\kappa_N}^{(j)}$ are the building blocks for the quantity $T_{\kappa_N}^{(\tau)}$ defined as

$$T_{\kappa_N}^{(\tau)} = \begin{cases} \dfrac{\left(M_{\kappa_N}^{(1)}\right)^\tau - \left(M_{\kappa_N}^{(2)}/2\right)^{\tau/2}}{\left(M_{\kappa_N}^{(2)}/2\right)^{\tau/2} - \left(M_{\kappa_N}^{(3)}/6\right)^{\tau/3}} & \text{if } \tau > 0, \\[2ex] \dfrac{\log\left(M_{\kappa_N}^{(1)}\right) - \frac{1}{2}\log\left(M_{\kappa_N}^{(2)}/2\right)}{\frac{1}{2}\log\left(M_{\kappa_N}^{(2)}/2\right) - \frac{1}{3}\log\left(M_{\kappa_N}^{(3)}/6\right)} & \text{if } \tau = 0. \end{cases}$$

The considered estimator of ρ is the following function of $T_{\kappa_N}^{(\tau)}$:

$$\widehat{\rho}_{\kappa_N}^{(\tau)} = -\left|\frac{3(T_{\kappa_N}^{(\tau)} - 1)}{T_{\kappa_N}^{(\tau)} - 3}\right|.$$

This estimator is implemented in the R function `mop` available as part of the `Expectrem` package (with credit due to B.G. Manjunath and F. Caeiro, who originally implemented this function in the now defunct R package `evt0`), with $\kappa_N = \lfloor N^{0.999} \rfloor$, and a choice of $\tau \in \{0, 1\}$ is made based on a stability criterion for $\kappa \mapsto \widehat{\rho}_\kappa^{(\tau)}$ for large κ (see Section 3.2 in [30] for more details). An estimator of b is then

$$\widehat{b}_{\kappa_N} = \left(\frac{\kappa_N}{N}\right)^{\overline{\rho}} \frac{\left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N}\left(\frac{i}{\kappa_N}\right)^{-\overline{\rho}}\right)\left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N} U_i\right) - \left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N}\left(\frac{i}{\kappa_N}\right)^{-\overline{\rho}} U_i\right)}{\left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N}\left(\frac{i}{\kappa_N}\right)^{-\overline{\rho}}\right)\left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N}\left(\frac{i}{\kappa_N}\right)^{-\overline{\rho}} U_i\right) - \left(\frac{1}{\kappa_N}\sum_{i=1}^{\kappa_N}\left(\frac{i}{\kappa_N}\right)^{-2\overline{\rho}} U_i\right)},$$

where $\overline{\rho} = \widehat{\rho}^{(\tau)}_{\kappa_N}$ and the $U_i = i \log(Z_{N-i+1,N}/Z_{N-i,N})$ are the weighted log-spacings. This estimator is also available from the R function `mop`.

To adapt these estimators to the regression setting, we choose a smoothing parameter h (in our case, $h = h_{n,\star}$), we compute $N = N_h(\boldsymbol{x}) = \sum_{t=1}^n \mathbb{1}_{\{\|\boldsymbol{X}_t - \boldsymbol{x}\| \le h\}}$ and we construct the sample $\{Z_1, \ldots, Z_N\}$ made of those Y_t such that $\boldsymbol{X}_t \in B(\boldsymbol{x}, h)$. Our estimators $\overline{\rho}(\boldsymbol{x})$ and $\overline{b}(\boldsymbol{x})$ are then given by $\overline{\rho}(\boldsymbol{x}) = \widehat{\rho}^{(\tau)}_{\kappa_N}$ and $\overline{b}(\boldsymbol{x}) = \widehat{b}_{\kappa_N}$. We also compute $\overline{\gamma}(\boldsymbol{x}) = M^{(1)}_{\lfloor N/4 \rfloor}$ as a preliminary estimator of $\gamma(\boldsymbol{x})$, used in the estimation of the bias and variance components during tuning parameter selection only.

C.2. Expression of the bias-reduced extreme expectile estimators. The full expression of the bias-reduced extrapolated estimator $\widehat{e}^{W,BR}_{n,\tau_n}(\tau'_n|\boldsymbol{x})$ is

$$\widehat{e}^{W,BR}_{n,\tau_n}(\tau'_n|\boldsymbol{x}) = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})} \widehat{e}^{W}_{n,\tau_n}(\tau'_n|\boldsymbol{x})(1+\overline{B}_{1,n}(\boldsymbol{x}))(1+\overline{B}_{2,n}(\boldsymbol{x}))(1+\overline{B}_{3,n}(\boldsymbol{x}))$$

where

$$1+\overline{B}_{1,n}(\boldsymbol{x}) = 1 + \frac{\left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\overline{\rho}(\boldsymbol{x})} - 1}{\overline{\rho}(\boldsymbol{x})} \overline{b}(\boldsymbol{x}) \widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})(1-\tau_n)^{-\overline{\rho}(\boldsymbol{x})},$$

$$1+\overline{B}_{2,n}(\boldsymbol{x})$$

$$= (1+\widehat{r}(\tau_n|\boldsymbol{x}))^{\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})} \left(1 + \frac{\frac{\left(1/\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})-1\right)^{-\overline{\rho}(\boldsymbol{x})}}{(1+\widehat{r}(\tau_n|\boldsymbol{x}))^{\overline{\rho}(\boldsymbol{x})}} - 1}{\overline{\rho}(\boldsymbol{x})} \overline{b}(\boldsymbol{x}) \widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})(1-\tau_n)^{-\overline{\rho}(\boldsymbol{x})}\right)^{-1},$$

$$1+\overline{B}_{3,n}(\boldsymbol{x})$$

$$= (1+\widehat{r}^{\star}(\tau'_n|\boldsymbol{x}))^{-\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})} \left(1 + \frac{\frac{\left(1/\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})-1\right)^{-\overline{\rho}(\boldsymbol{x})}}{(1+\widehat{r}^{\star}(\tau'_n|\boldsymbol{x}))^{\overline{\rho}(\boldsymbol{x})}} - 1}{\overline{\rho}(\boldsymbol{x})} \overline{b}(\boldsymbol{x}) \widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})(1-\tau'_n)^{-\overline{\rho}(\boldsymbol{x})}\right)$$

and, in $\overline{B}_{3,n}(\boldsymbol{x})$,

$$1+\widehat{r}^{\star}(\tau'_n|\boldsymbol{x})$$

$$= \left(1 - \frac{\widehat{m}_n(\boldsymbol{x})}{\widehat{e}^{W}_{n,\tau_n}(\tau'_n|\boldsymbol{x})}\right) \frac{1}{2\tau'_n - 1} \left(1 + \frac{\overline{b}(\boldsymbol{x})\left(1/\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x}) - 1\right)^{-\overline{\rho}(\boldsymbol{x})}}{1 - \widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x}) - \overline{\rho}(\boldsymbol{x})}(1-\tau'_n)^{-\overline{\rho}(\boldsymbol{x})}\right)^{-1}.$$

Meanwhile, the bias-reduced version of the extrapolated quantile-based estimator

$$\check{e}^{W}_{n,\tau_n}(\tau'_n|\boldsymbol{x}) = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})} \check{e}_n(\tau_n|\boldsymbol{x})$$

is

$$\check{e}^{W,BR}_{n,\tau_n}(\tau'_n|\boldsymbol{x}) = \check{e}^{W}_{n,\tau_n}(\tau'_n|\boldsymbol{x})(1+\check{B}_{1,n}(\boldsymbol{x}))(1+\check{B}_{3,n}(\boldsymbol{x})).$$

Here $\check{B}_{1,n}(\boldsymbol{x})$ and $\check{B}_{3,n}(\boldsymbol{x})$ are defined as $\overline{B}_{1,n}(\boldsymbol{x})$ and $\overline{B}_{3,n}(\boldsymbol{x})$ above, only with $\widehat{\gamma}^{(J,BR)}_{\tau_n}(\boldsymbol{x})$ in place of $\widehat{\gamma}^{E,BR}_{\tau_n}(\boldsymbol{x})$. These bias-corrected versions are obtained by adapting the bias correction methodology of [28], developed for the unconditional extrapolated expectile estimators, to the nonparametric regression setting.

C.3. Construction of the corrected asymptotic variance estimator for extreme expectiles. We provide further details regarding the design of the variance correction for extreme conditional expectile estimators. The proof of Theorem 3.4 seeks, as an intermediate step, the asymptotic distribution of

$$\sqrt{nh_n^p(1-\tau_n)}\left(\frac{\widehat{\overline{F}}_n(\widehat{e}_n(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{1-\tau_n}-(1/\gamma(\boldsymbol{x})-1),\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right).$$

It partly relies on the convergence $q(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})/e(\tau_n|\boldsymbol{x})\to 1$. A more accurate approximation, due to Proposition 1 in [17], is

$$\frac{q(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}\approx 1-\frac{\gamma(\boldsymbol{x})m(\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}.$$

Retracing the steps of the proof of Theorem 3.4, we find that the asymptotic distribution of

$$\sqrt{nh_n^p(1-\tau_n)}\left(\frac{\widehat{\overline{F}}_n(\widehat{e}_n(\tau_n|\boldsymbol{x})|\boldsymbol{x})}{1-\tau_n}-(1/\gamma(\boldsymbol{x})-1),\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\right)$$

can be more accurately approximated by that of

$$\begin{pmatrix}-\dfrac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\dfrac{1}{1-\frac{\gamma(\boldsymbol{x})m(\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}} & \dfrac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\\ 1 & 0\end{pmatrix}\sqrt{nh_n^p(1-\tau_n)}\begin{pmatrix}\dfrac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1\\ \dfrac{\widehat{q}_n(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})}{q(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})}-1\end{pmatrix}.$$

Moreover, the key approximation in Equation (20) (part of the proof of Theorem 3.1) is an approximation of $\overline{F}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x})$ by $(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)$. This can be refined as

$$\overline{F}(e(\tau_n|\boldsymbol{x})|\boldsymbol{x})\approx(1-\tau_n)\left(\frac{1}{\gamma(\boldsymbol{x})}-1\right)(1+r(\tau_n|\boldsymbol{x}))$$

instead, where

$$r(\tau|\boldsymbol{x})=\left(1-\frac{m(\boldsymbol{x})}{e(\tau|\boldsymbol{x})}\right)\frac{1}{2\tau-1}-1.$$

This refined approximation is motivated by an inspection of the proof of Proposition 1 in [16]. Then, the distribution of

$$\sqrt{nh_n^p(1-\tau_n)}\left(\frac{\widehat{e}_n(\tau_n|\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}-1,\frac{\widehat{q}_n(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})}{q(1-(1-\tau_n)(1/\gamma(\boldsymbol{x})-1)|\boldsymbol{x})}-1\right)$$

may also be more accurately approximated by a zero-mean Gaussian distribution with covariance matrix

$$\frac{\int_{\mathbb{R}^p}K^2}{g(\boldsymbol{x})}\begin{pmatrix}\dfrac{2\gamma^3(\boldsymbol{x})}{1-2\gamma(\boldsymbol{x})}\dfrac{2\tau_n-1}{1-\frac{m(\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}} & \dfrac{\gamma^3(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}\sqrt{\dfrac{2\tau_n-1}{1-\frac{m(\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}}}\\ \dfrac{\gamma^3(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}\sqrt{\dfrac{2\tau_n-1}{1-\frac{m(\boldsymbol{x})}{e(\tau_n|\boldsymbol{x})}}} & \dfrac{\gamma^3(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}\end{pmatrix}.$$

Using the delta method, it is then straightforward that the desired distribution may then be better approximated by a Gaussian distribution with covariance matrix $(\int_{\mathbb{R}^p}K^2/g(\boldsymbol{x}))\boldsymbol{T}_n(\boldsymbol{x})$, with $\boldsymbol{T}_n(\boldsymbol{x})$ given in Section 5.3.2.

C.4. Full set of finite-sample results. We provide here a full set of finite-sample results linked to our simulation study. We consider the following models spanning our list of worked-out examples:

(LS) The location-scale model $Y_t = m(X_t) + \sigma(X_t)\varepsilon_t$, with $m(x) = 1 + \sin(2\pi x)/5$, $\sigma(x) = (5+x)/(4(1+x))$ (for $x \in [0,1]$). Here:
- $X_t = \Phi(Z_t)$, where Φ is the standard Gaussian distribution function and (Z_t) is a GARCH(1,1) process with $\omega = 0.25$, $\alpha = 0.75$, $\beta = 0.2$, *i.e.* $Z_{t+1} = \Sigma_{t+1}\delta_{t+1}$, where the δ_t are i.i.d. standard Gaussian and Σ_{t+1} is defined recursively as $\Sigma_{t+1} = (\omega + \alpha Z_t^2 + \beta\Sigma_t^2)^{1/2}$. The process (X_t) is geometrically $\beta-$mixing because (Z_t) (simulated using the `garch.sim` routine from the R package `TSA`) is so, see Theorem 3.4 on p.66 of [24].
- $\varepsilon_t = \Phi_\nu^{-1}(U_t)$, where Φ_ν is the Student distribution function with $\nu = 10/3$ degrees of freedom, and (U_t) is defined recursively as $U_0 \sim$ Uniform$[0,1]$ and, for $t \geq 1$, $U_t = \frac{1}{r}U_{t-1} + \eta_t$, where the η_t are i.i.d. uniform over $\left\{0, \frac{1}{r}, \ldots, \frac{r-1}{r}\right\}$ (for a fixed integer $r \geq 2$). The uniform AR(1) process (U_t), in the terminology of [12], is stationary with standard uniform margins and satisfies

$$\mathbb{E}(|x/r + \eta_0|) \leq \frac{|x|}{r} + \frac{r-1}{2r} \leq \begin{cases} 3|x|/4 & \text{if } |x| > (2r-2)/(3r-4), \\ 3/2 < \infty & \text{otherwise.} \end{cases}$$

As such, it is geometrically $\alpha-$mixing, see p.xvi of [41]. We take $r = 5$.

Here $\gamma(x) = 0.3$ and $\rho(x) = -2\gamma(x) = -0.6$ for any x, and $q(\tau|x) = m(x) + \sigma(x)\Phi_\nu^{-1}(\tau)$. The theoretical conditional expectile $e(\tau|x)$ is similarly computed by replacing $\Phi_\nu^{-1}(\tau)$ by the τ-expectile of the Student distribution with ν degrees of freedom, computed using the R function `et` from the R package `Expectrem`.

(NL) A nonlinear Burr process $Y_t = ((1-U_t)^{\rho(X_t)} - 1)^{-\gamma(X_t)/\rho(X_t)}$, with (X_t) and (U_t) generated as in model (LS). We fix $\rho(x) = -1$ for all $x \in [0,1]$ and consider three different models for $\gamma(x)$, $x \in [0,1]$:

(NL-P) The polynomial model $\gamma(x) = 0.15 + 0.5x(1-x)$;

(NL-S) The sinusoidal model $\gamma(x) = 0.2 + 0.05\sin(2\pi x)$;

(NL-C) The constant model $\gamma(x) = 0.2$.

In these three cases, $\gamma(x) \in (0, 1/2)$ for any $x \in [0,1]$. The true value of the conditional quantile is $q(\tau|x) = ((1-\tau)^{-1} - 1)^{\gamma(x)}$. To specifically assess the effect of a stronger bias component on the estimation and inference, we also consider the model

(NL-CB) The constant model $\gamma(x) = 0.2$ with $\rho(x) = -1/2$.

In this model, $q(\tau|x) = ((1-\tau)^{-1/2} - 1)^{2\gamma(x)}$, and again the theoretical conditional expectile $e(\tau|x)$ can be computed numerically using the R function `eburr`.

(AR) The autoregressive process $Y_t = \phi Y_{t-1} + \varepsilon_t$, where $\phi = 0.1$ and (ε_t) is an i.i.d. sequence of realizations of a random variable whose distribution is the following mixture:
- With probability $3/4$, a uniform distribution over $[-1,1]$;
- With probability $1/4$, a symmetrized Pareto distribution, *i.e.* having density $f(x) = |x|^{-1/\gamma-1}/(2\gamma)$ for $|x| > 1$, with tail index $\gamma = 0.4$.

In this time series example, the natural covariate for the response Y_t is $X_t = Y_{t-1}$, and we have $\gamma(x) = 0.4$ for any x. The theoretical conditional extreme quantile for tomorrow's unobserved value Y_{t+1} given our knowledge $Y_t = x$ of today is, for $\tau > 7/8$, $q(\tau|x) = \phi x + (1 - 8(\tau - 7/8))^{-\gamma}$, *i.e.* $q(1-1/t|x) = 8t^\gamma$ for t large enough. Therefore, strictly speaking, the second-order regular variation condition $\mathcal{C}_2(\gamma(x), \rho(x), A(\cdot|x))$ is not satisfied, but we are in the case when the conditional right tail is pure Pareto, which is usually understood as $\rho(x) = -\infty$ and $A(\cdot|x) = 0$. Besides, the theoretical conditional extreme expectile

$e(\tau|x)$ is numerically computed as $\phi x + e_\tau$, where e_τ satisfies, for τ large enough,

$$\frac{\int_{e_\tau}^{\infty} \frac{1}{8} z^{-1/\gamma}\, dz}{2\int_{e_\tau}^{\infty} \frac{1}{8} z^{-1/\gamma}\, dz + e_\tau} = 1 - \tau,$$

and is determined by making use of the R functions `integrate` and `uniroot`.

The covariate in all these models is one-dimensional. We also consider the following two-dimensional example:

(NL-S-2) A nonlinear Burr process $Y_t = ((1-U_t)^{\rho(X_{1,t},X_{2,t})} - 1)^{-\gamma(X_{1,t},X_{2,t})/\rho(X_{1,t},X_{2,t})}$, where we take $\rho(x_1,x_2) \equiv -1$ and $\gamma(x_1,x_2) = 0.2 + 0.05\sin(2\pi x_1)\cos(2\pi x_2)$. The covariate $(X_{1,t}, X_{2,t})$ is generated as $(X_{1,t}, X_{2,t}) = (\Phi(Z_{1,t}), \Phi(Z_{2,t}))$, where Φ is the standard Gaussian distribution function, and $(Z_{1,t}, Z_{2,t})$ are observations from the bivariate GARCH(1,1) process $Z_{j,t+1} = \Sigma_{j,t+1}\delta_{j,t+1},\ \Sigma^2_{j,t+1} = \omega + \alpha Z^2_{j,t} + \beta\Sigma^2_{j,t}\ (j = 1,2)$ with $\omega = 0.25$, $\alpha = 0.75$, $\beta = 0.2$ and $(\delta_{1,t}, \delta_{2,t})$ are i.i.d. bivariate standard Gaussian with correlation 0.5. The true value of the conditional quantile is $q(\tau|x_1,x_2) = ((1-\tau)^{-1} - 1)^{\gamma(x_1,x_2)}$.

In this final model, in line with the univariate examples, K is the uniform kernel over $[-1,1]^2$, that is, $K(u_1,u_2) = 0.25\,\mathbb{1}_{\{|u_1|\le 1\}}\mathbb{1}_{\{|u_2|\le 1\}}$, and the bandwidth is chosen as $h_{n,\star} = \sigma n^{-1/6}$ where $\sigma = 1/\sqrt{12}$ (the standard deviation of a uniform distribution), in accordance with [4]. Due to the difficulty of estimating the second-order parameters ρ and b in the situation $p = 2$ where local effective sample sizes tend to be low, we fix the values of their estimates at -1 and 1, respectively. These (misspecified) choices constitute *ad hoc* compromises between bias reduction and variability of the estimators; in particular, misspecifying the estimate of ρ as -1 has been recommended on p.117 of Section 4.5.1 in [1], on p.212-215 in [29] and on p.195 of Section 6.6 in [40].

We illustrate the performance of the extreme quantile and asymmetric least squares expectile estimators and the pertaining confidence intervals in models (LS), (NL-P), (NL-S), (NL-C), (NL-CB) and (AR), with $n \in \{1{,}000, 5{,}000, 10{,}000\}$. In model (NL-S-2), we consider $n = 10{,}000$. We represent boxplots of the extreme conditional log-quantile and log-expectile estimates as well as the coverage probabilities and pointwise median lengths of the intervals $\widehat{I}_{q,j}(\tau_n'|\boldsymbol{x})$ and $\widehat{I}_{E,j}(\tau_n'|\boldsymbol{x})$, $1 \le j \le 4$, for $\tau_n' = 1 - 10/n \in \{0.99, 0.998, 0.999\}$.

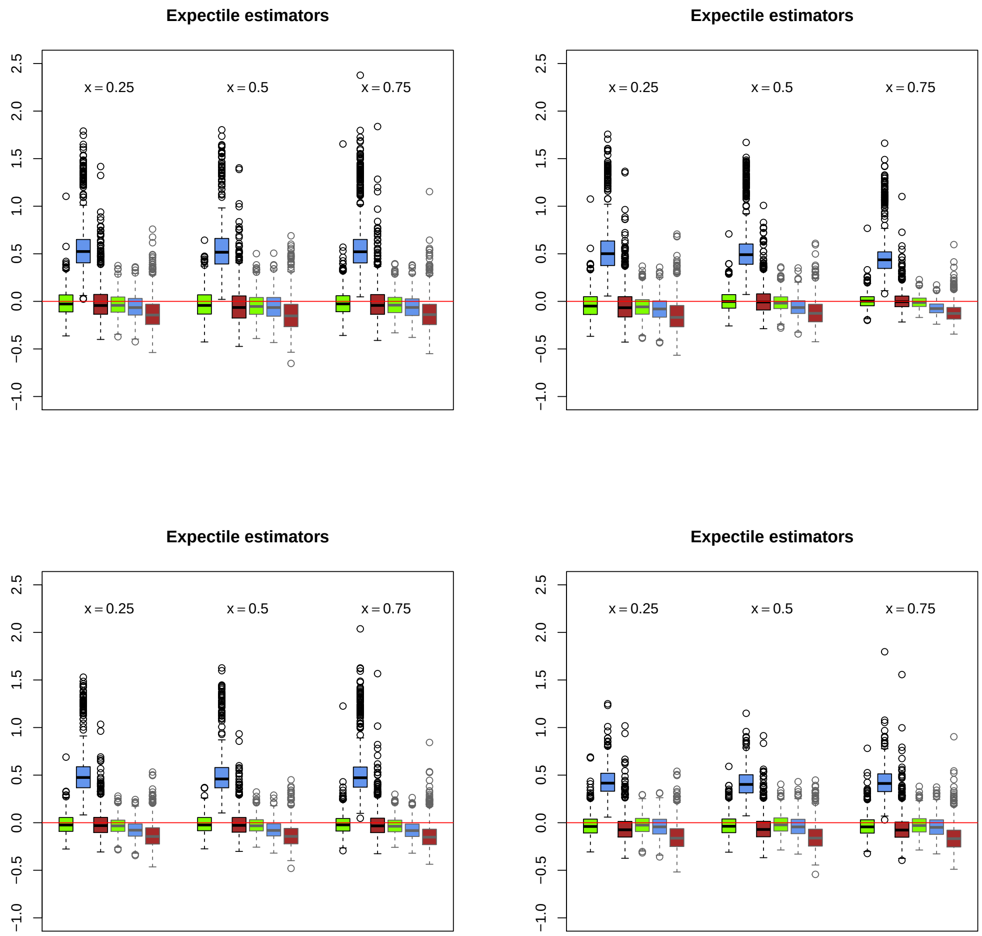


FIG C.1. *Simulation results in the nonlinear Burr model for* $\tau_n' = 1 - 10/n$, *with* $n = 5{,}000$. *Boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue),* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown),* $\log(\breve{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green, grayed out),* $\log(\breve{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue, grayed out) and* $\log(\breve{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown, grayed out). Top left: model (NL-P), top right: model (NL-S), bottom left: model (NL-C), bottom right: model (NL-CB). In each panel, the boxplots on the left correspond to* $x = 1/4$, *those in the middle to* $x = 1/2$, *and those on the right to* $x = 3/4$.

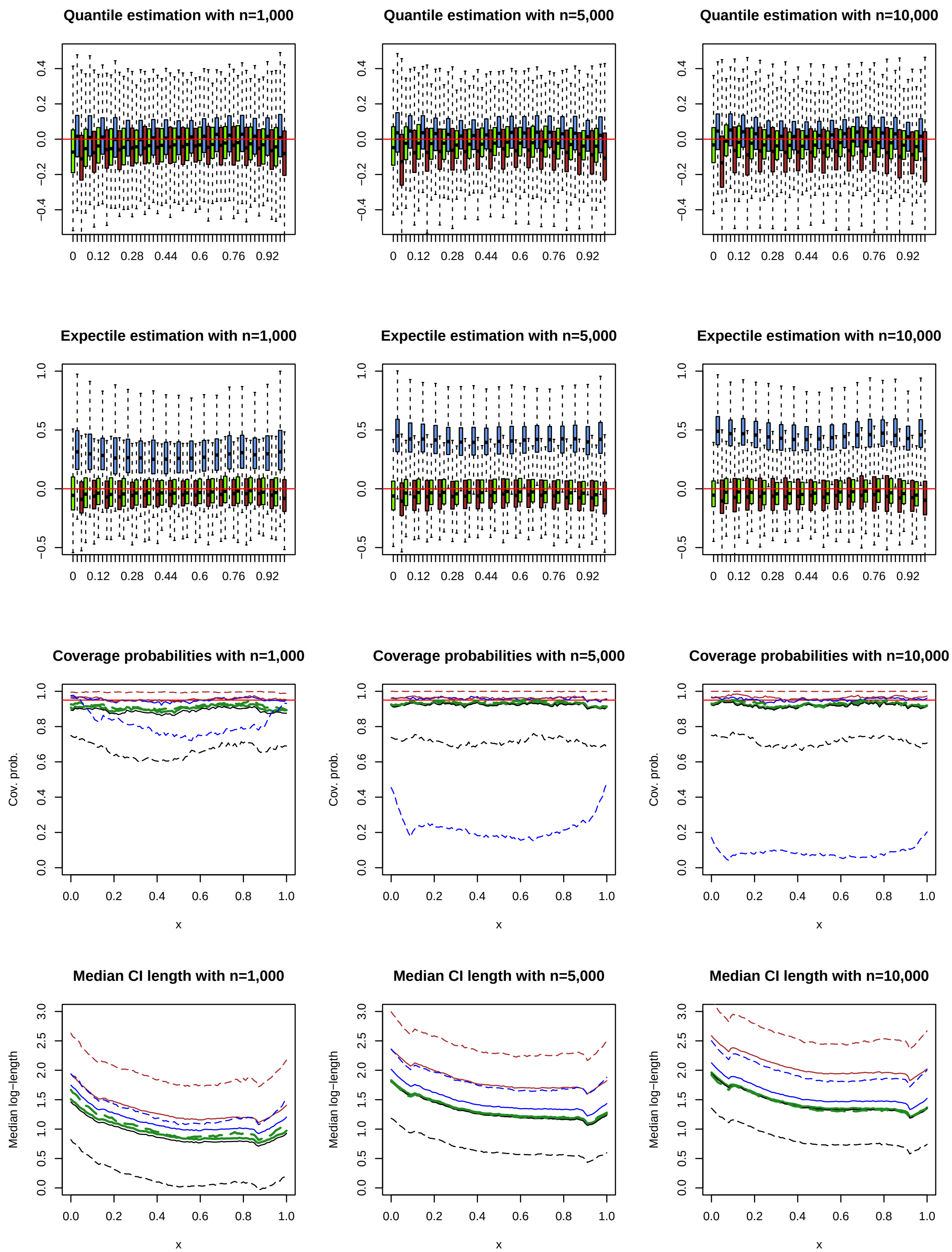


FIG C.2. *Simulation results in the location-scale model (LS) for* $\tau_n' = 1 - 10/n$*, with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

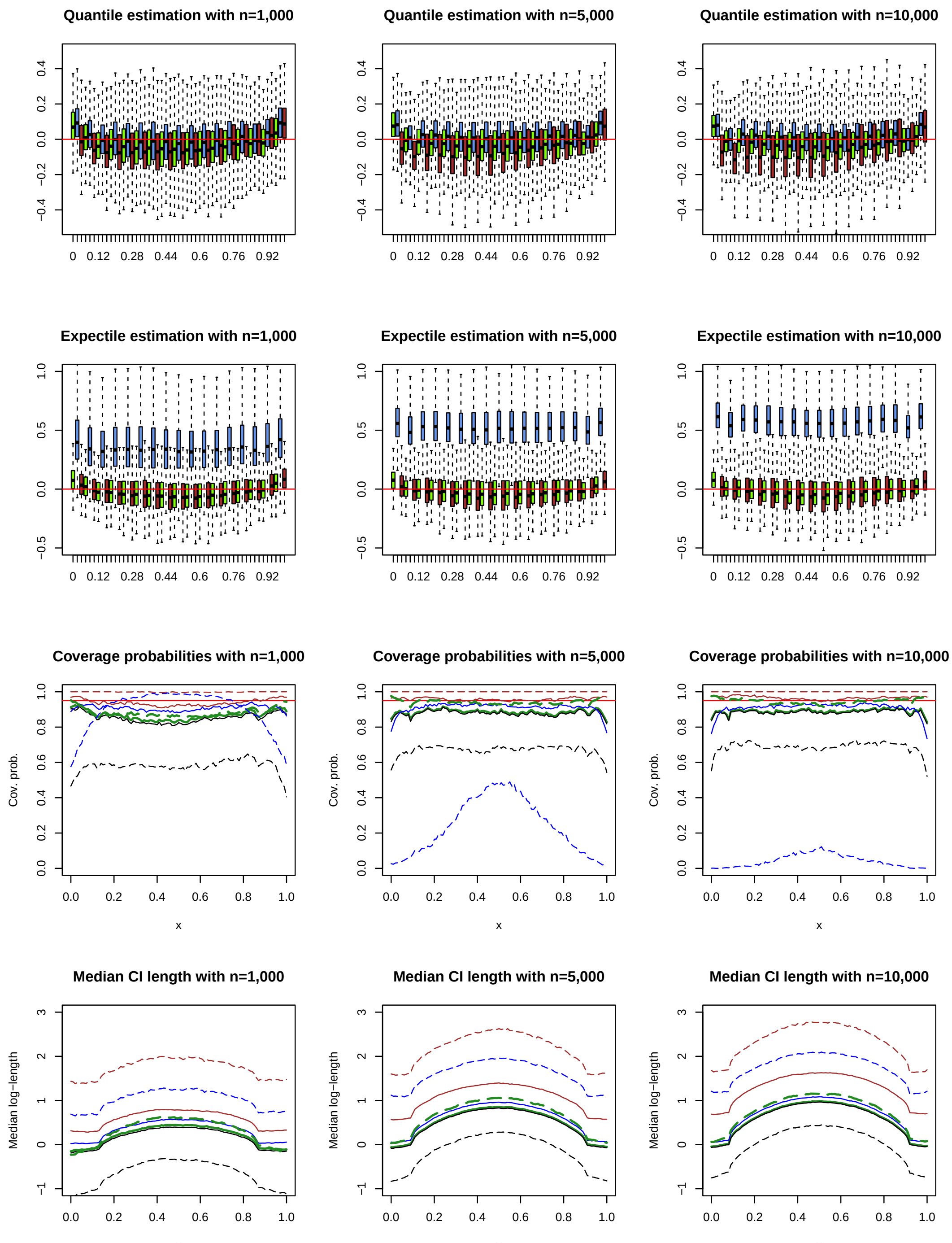


FIG C.3. *Simulation results in the nonlinear Burr model (NL-P) for* $\tau_n' = 1 - 10/n$*, with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

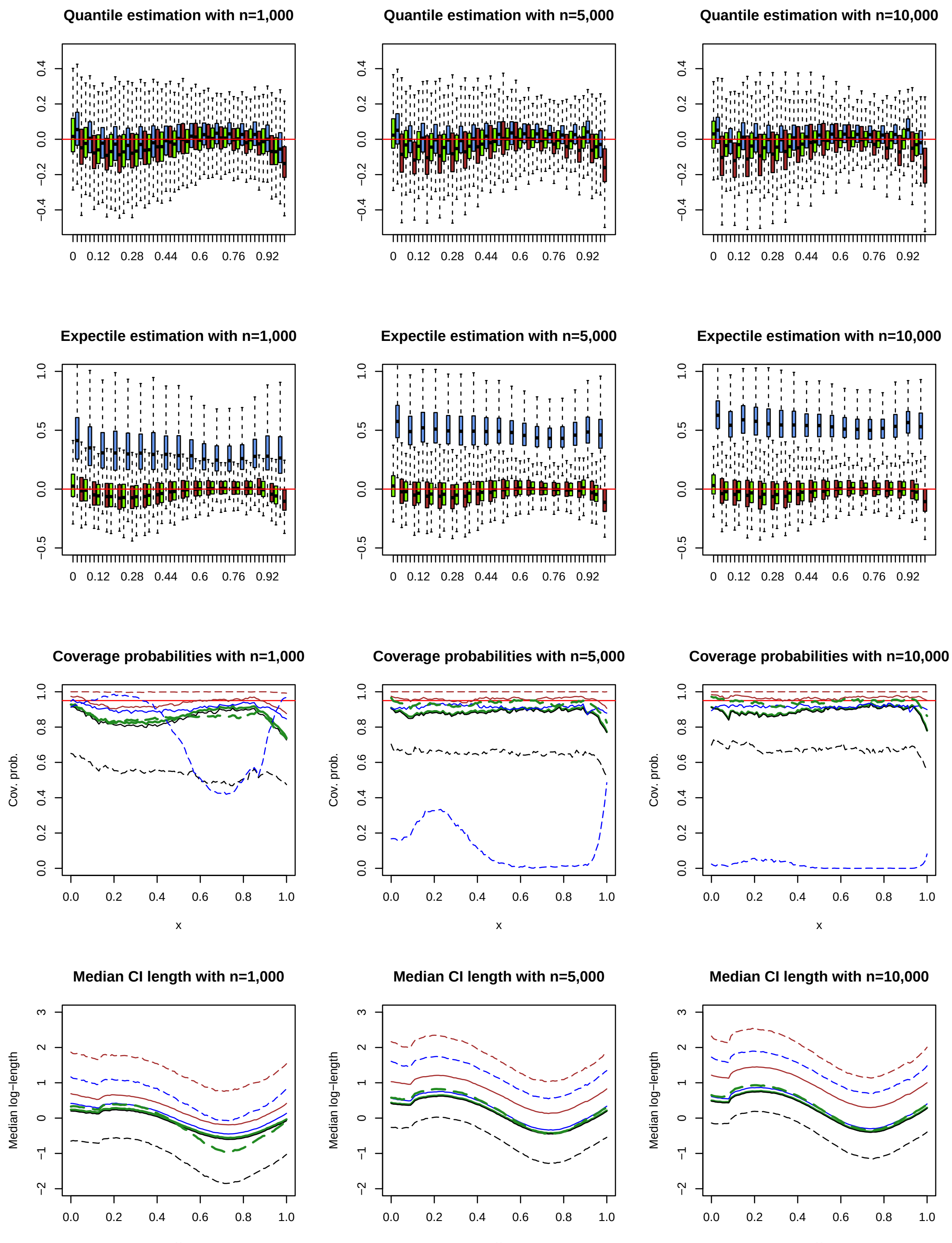


FIG C.4. *Simulation results in the nonlinear Burr model (NL-S) for* $\tau_n' = 1 - 10/n$, *with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

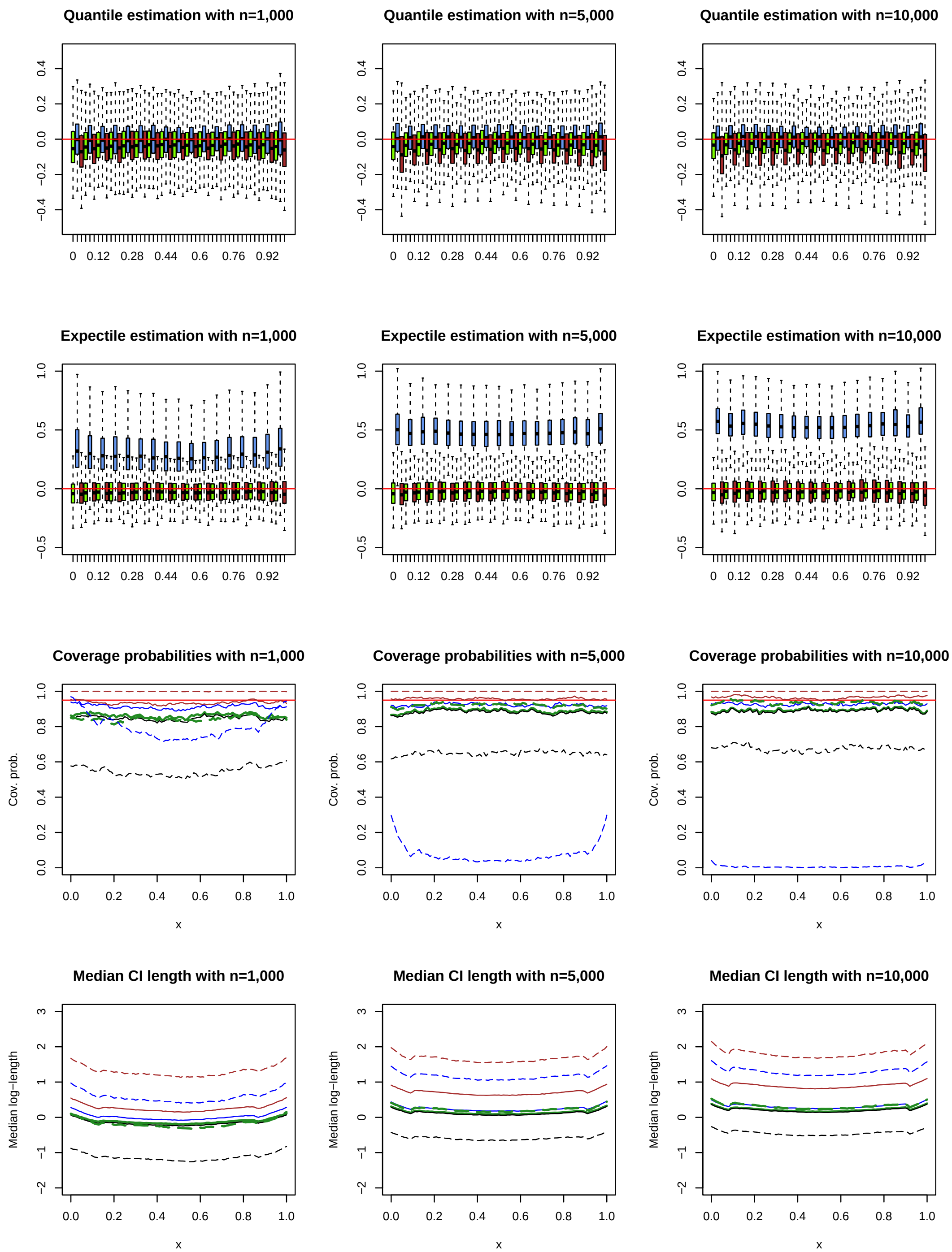


FIG C.5. *Simulation results in the nonlinear Burr model (NL-C) for* $\tau_n' = 1 - 10/n$*, with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

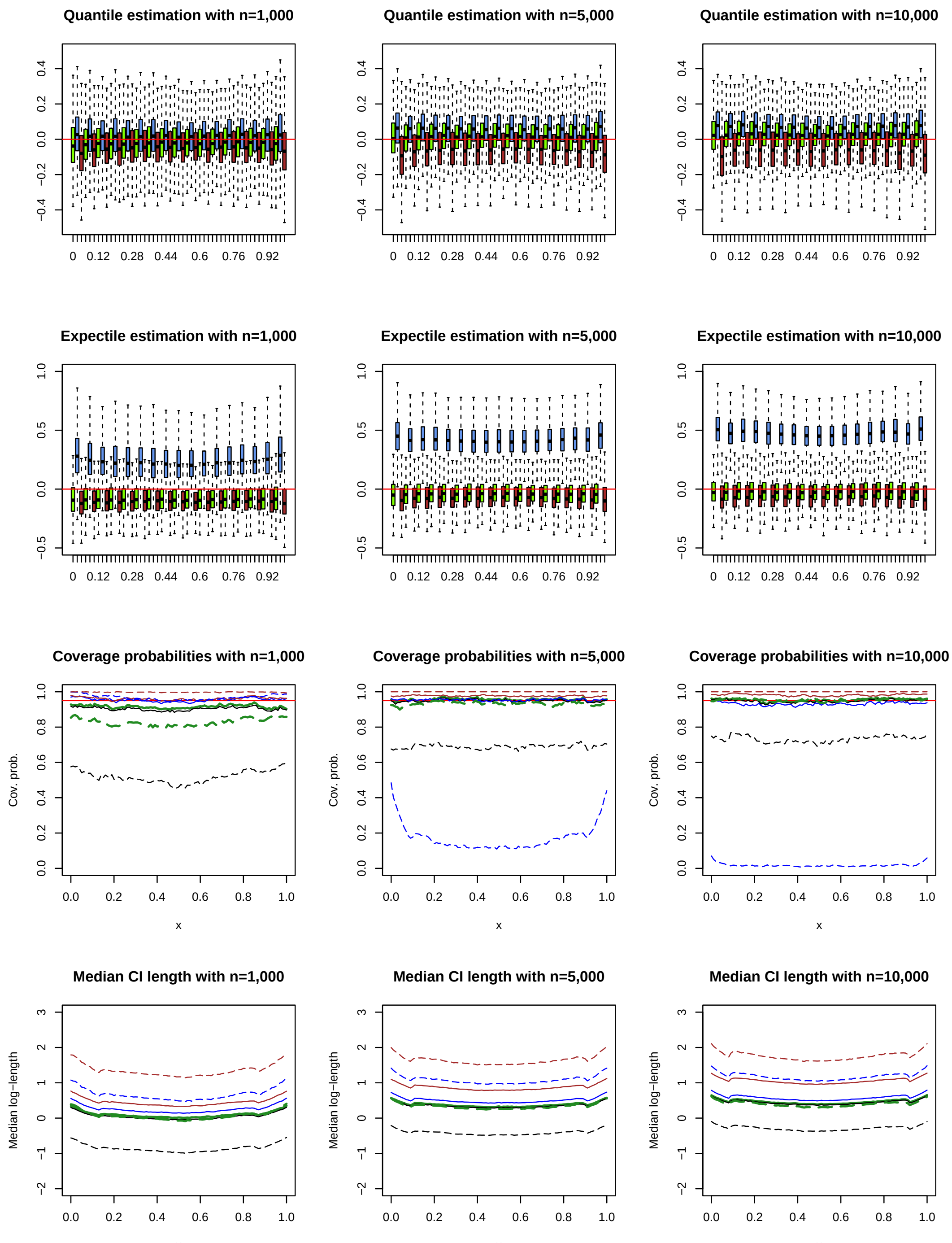


FIG C.6. *Simulation results in the nonlinear Burr model (NL-CB) for* $\tau_n' = 1 - 10/n$*, with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

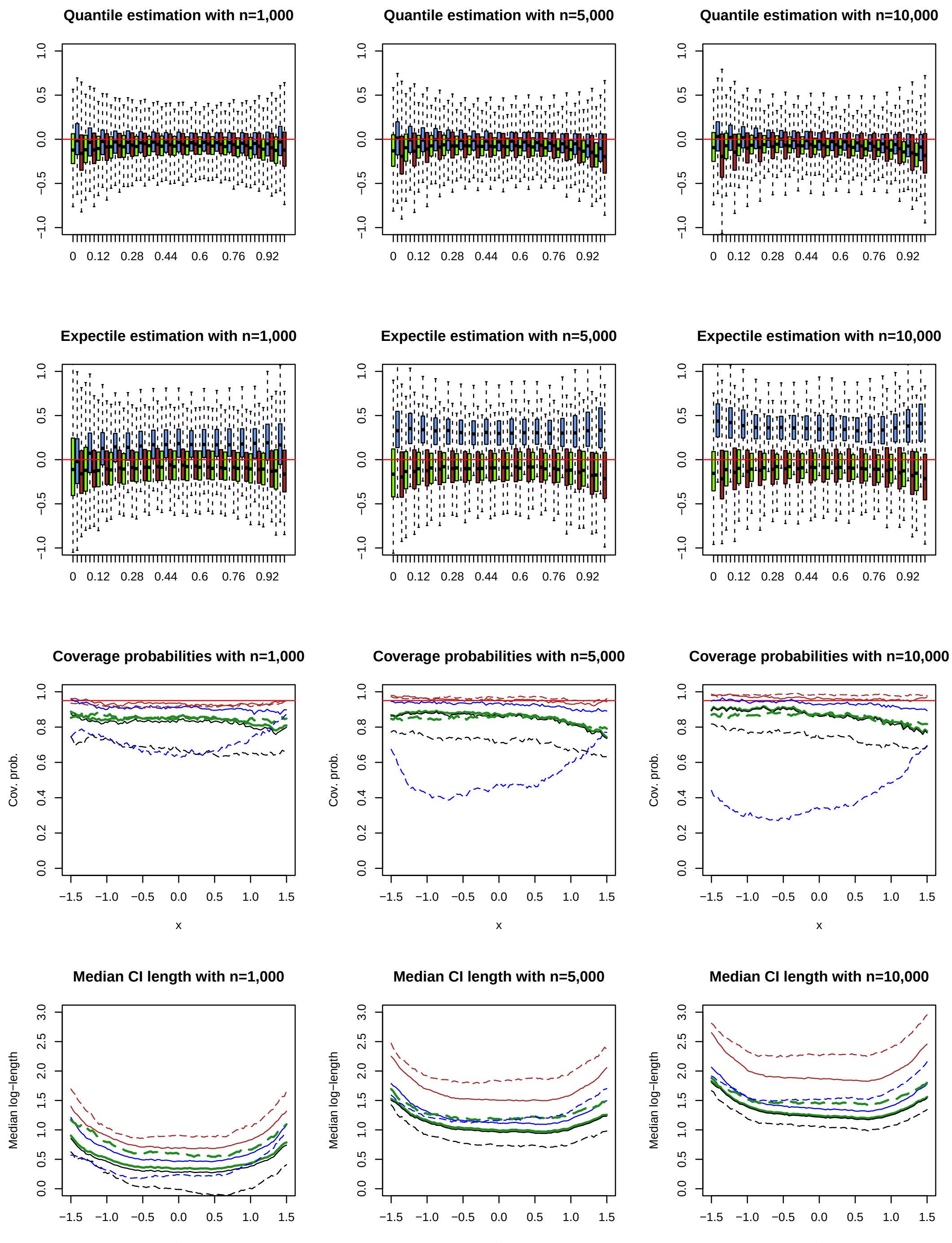


FIG C.7. *Simulation results in the autoregressive model (AR) for* $\tau_n' = 1 - 10/n$, *with* $n = 1{,}000$ *(left panels),* $n = 5{,}000$ *(middle panels), and* $n = 10{,}000$ *(right panels). First row: boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/q(\tau_n'|x))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|x)/q(\tau_n'|x))$ *(brown). Second row: boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|x)/e(\tau_n'|x))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|x)/e(\tau_n'|x))$ *(brown). Third row: empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|x)$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|x)$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|x)$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|x)$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|x)$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|x)$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|x)$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|x)$ *(dashed brown line), with the* 95% *nominal level in full red line. Fourth row: pointwise median length of these confidence intervals, with the same color code, on the log-scale.*

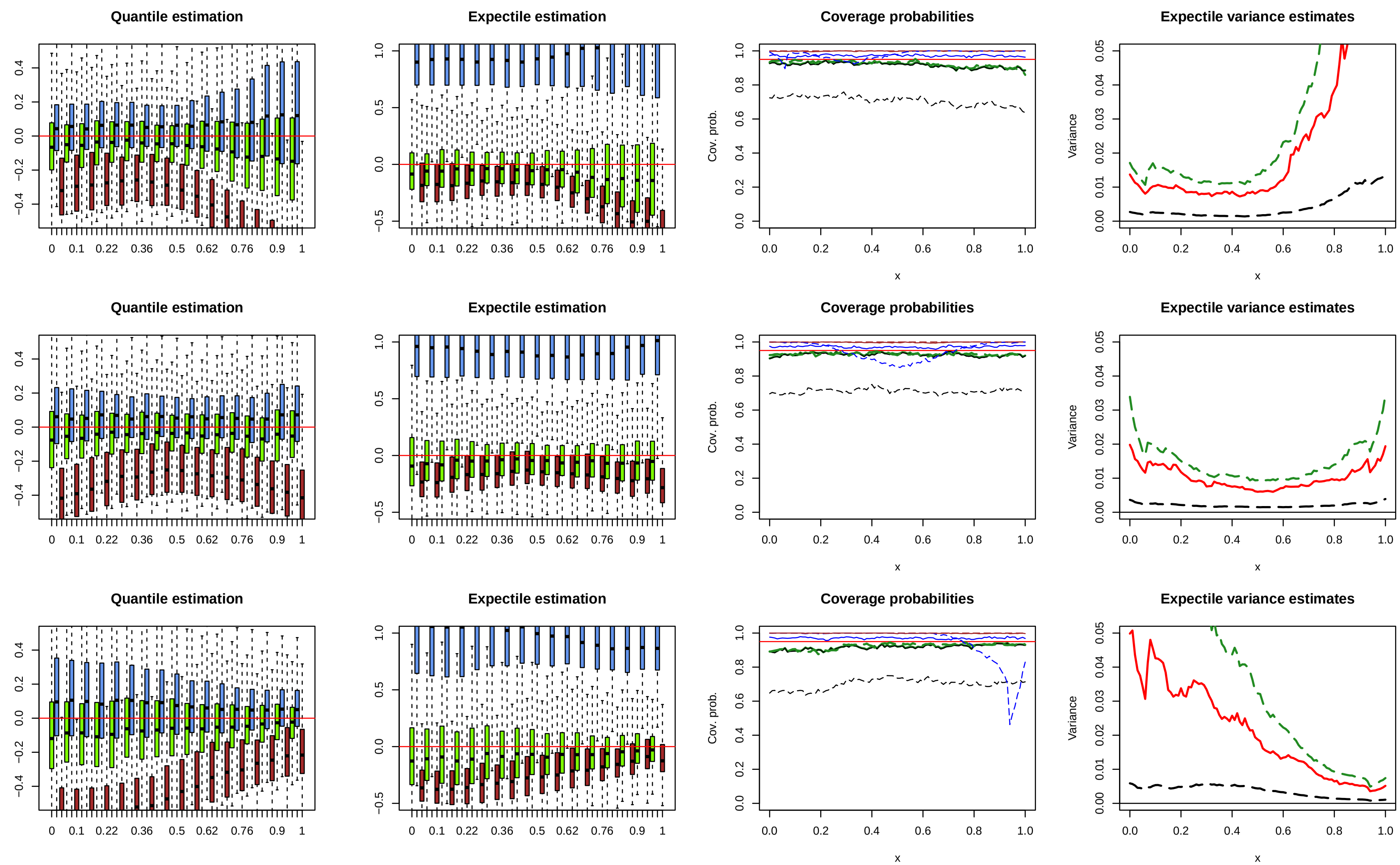


FIG C.8. *Simulation results in model (NL-S-2). From left to right, boxplots of* $\log(\widehat{q}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/q(\tau_n'|\boldsymbol{x}))$ *(green),* $\log(\widehat{q}^{W}_{1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/q(\tau_n'|\boldsymbol{x}))$ *(blue) and* $\log(\widehat{q}_n(\tau_n'|\boldsymbol{x})/q(\tau_n'|\boldsymbol{x}))$ *(brown); boxplots of* $\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/e(\tau_n'|\boldsymbol{x}))$ *(green),* $\log(\widehat{e}^{W}_{1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/e(\tau_n'|\boldsymbol{x}))$ *(blue) and* $\log(\widehat{e}_n(\tau_n'|\boldsymbol{x})/e(\tau_n'|\boldsymbol{x}))$ *(brown); empirical pointwise coverage probabilities of the asymptotic* 95% *confidence intervals* $\widehat{I}_{q,1}(\tau_n'|\boldsymbol{x})$ *(full black line),* $\widehat{I}_{q,2}(\tau_n'|\boldsymbol{x})$ *(full green line),* $\widehat{I}_{q,3}(\tau_n'|\boldsymbol{x})$ *(full blue line),* $\widehat{I}_{q,4}(\tau_n'|\boldsymbol{x})$ *(full brown line),* $\widehat{I}_{E,1}(\tau_n'|\boldsymbol{x})$ *(dashed black line),* $\widehat{I}_{E,2}(\tau_n'|\boldsymbol{x})$ *(dashed green line),* $\widehat{I}_{E,3}(\tau_n'|\boldsymbol{x})$ *(dashed blue line) and* $\widehat{I}_{E,4}(\tau_n'|\boldsymbol{x})$ *(dashed brown line), with the target* 95% *nominal level in full red line; comparison of the pointwise empirical variances of the* $(\sqrt{k_{n,\star}h_{n,\star}^2}/\log(k_{n,\star}/(n(1-\tau_n'))))\log(\widehat{e}^{W,BR}_{1-k_{n,\star}/n}(\tau_n'|\boldsymbol{x})/e(\tau_n'|\boldsymbol{x}))$ *(full red line), the pointwise median of the corrected variance estimates* $(\int_{\mathbb{R}^2}K^2/\widehat{g}_n(\boldsymbol{x}))\widetilde{v}_E(\boldsymbol{x})$ *(dashed green line) and the pointwise median of the naive variance estimates* $(\int_{\mathbb{R}^2}K^2/\widehat{g}_n(\boldsymbol{x}))\widehat{v}_E(\boldsymbol{x})$ *(dashed black line). Top row:* $x_1=0.1$ *and* $x_2\in[0,1]$*, middle row:* $x_1=0.5$ *and* $x_2\in[0,1]$*, bottom row:* $x_1=0.9$ *and* $x_2\in[0,1]$*, with* $n=10{,}000$ *and estimates at level* $\tau_n'=1-10/n=0.999$ *in each case.*

$x = 0.1$	Estimate	Bias	Variance	MSE	C.I.	Med. est. var.	Cov.
NL-P	$\widehat{q}$	$-2.87 \cdot 10^{-2}$	$2.07 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$I_{q,4}$	$3.09 \cdot 10^{-2}$	0.983
	$\widehat{q}^{W}$	$3.35 \cdot 10^{-2}$	$9.57 \cdot 10^{-3}$	$1.07 \cdot 10^{-2}$	$I_{q,3}$	$8.11 \cdot 10^{-3}$	0.904
	$\widehat{q}^{W,BR}$	$-4.96 \cdot 10^{-3}$	$9.23 \cdot 10^{-3}$	$9.25 \cdot 10^{-3}$	$I_{q,1}$	$6.91 \cdot 10^{-3}$	0.886
					$I_{q,2}$	$7.18 \cdot 10^{-3}$	0.894
	$\widehat{e}$	$1.45 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$I_{E,4}$	$3.11 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$6.36 \cdot 10^{-1}$	$6.24 \cdot 10^{-2}$	$4.66 \cdot 10^{-1}$	$I_{E,3}$	$3.35 \cdot 10^{-2}$	0.006
	$\widehat{e}^{W,BR}$	$9.71 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$I_{E,1}$	$2.94 \cdot 10^{-3}$	0.711
					$I_{E,2}$	$1.38 \cdot 10^{-2}$	0.959
NL-S	$\widehat{q}$	$-6.01 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$3.05 \cdot 10^{-2}$	$I_{q,4}$	$4.07 \cdot 10^{-2}$	0.974
	$\widehat{q}^{W}$	$1.79 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$I_{q,3}$	$1.09 \cdot 10^{-2}$	0.926
	$\widehat{q}^{W,BR}$	$-2.74 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$	$I_{q,1}$	$9.32 \cdot 10^{-3}$	0.890
					$I_{q,2}$	$9.69 \cdot 10^{-3}$	0.892
	$\widehat{e}$	$-1.58 \cdot 10^{-2}$	$2.96 \cdot 10^{-2}$	$2.98 \cdot 10^{-2}$	$I_{E,4}$	$2.24 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$6.19 \cdot 10^{-1}$	$5.11 \cdot 10^{-2}$	$4.35 \cdot 10^{-1}$	$I_{E,3}$	$4.23 \cdot 10^{-2}$	0.026
	$\widehat{e}^{W,BR}$	$-1.12 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$	$I_{E,1}$	$4.44 \cdot 10^{-3}$	0.723
					$I_{E,2}$	$1.98 \cdot 10^{-2}$	0.946
NL-C	$\widehat{q}$	$-5.43 \cdot 10^{-2}$	$2.07 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$	$I_{q,4}$	$3.09 \cdot 10^{-2}$	0.979
	$\widehat{q}^{W}$	$1.64 \cdot 10^{-2}$	$8.91 \cdot 10^{-3}$	$9.17 \cdot 10^{-3}$	$I_{q,3}$	$8.30 \cdot 10^{-3}$	0.934
	$\widehat{q}^{W,BR}$	$-2.33 \cdot 10^{-2}$	$8.39 \cdot 10^{-3}$	$8.93 \cdot 10^{-3}$	$I_{q,1}$	$7.10 \cdot 10^{-3}$	0.903
					$I_{q,2}$	$7.38 \cdot 10^{-3}$	0.906
	$\widehat{e}$	$-1.40 \cdot 10^{-2}$	$1.98 \cdot 10^{-2}$	$2.00 \cdot 10^{-2}$	$I_{E,4}$	$1.71 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.84 \cdot 10^{-1}$	$3.70 \cdot 10^{-2}$	$3.78 \cdot 10^{-1}$	$I_{E,3}$	$3.22 \cdot 10^{-2}$	0.006
	$\widehat{e}^{W,BR}$	$-9.17 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$I_{E,1}$	$2.99 \cdot 10^{-3}$	0.710
					$I_{E,2}$	$1.37 \cdot 10^{-2}$	0.947
$x = 0.3$	Estimate	Bias	Variance	MSE	C.I.	Med. est. var.	Cov.
NL-P	$\widehat{q}$	$-6.51 \cdot 10^{-2}$	$3.17 \cdot 10^{-2}$	$3.59 \cdot 10^{-2}$	$I_{q,4}$	$4.11 \cdot 10^{-2}$	0.964
	$\widehat{q}^{W}$	$1.58 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$I_{q,3}$	$1.15 \cdot 10^{-2}$	0.922
	$\widehat{q}^{W,BR}$	$-3.00 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$I_{q,1}$	$9.93 \cdot 10^{-3}$	0.887
					$I_{q,2}$	$1.03 \cdot 10^{-2}$	0.896
	$\widehat{e}$	$-2.62 \cdot 10^{-2}$	$3.94 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$	$I_{E,4}$	$2.56 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.94 \cdot 10^{-1}$	$5.05 \cdot 10^{-2}$	$4.04 \cdot 10^{-1}$	$I_{E,3}$	$4.03 \cdot 10^{-2}$	0.048
	$\widehat{e}^{W,BR}$	$-2.14 \cdot 10^{-2}$	$1.82 \cdot 10^{-2}$	$1.87 \cdot 10^{-2}$	$I_{E,1}$	$4.99 \cdot 10^{-3}$	0.691
					$I_{E,2}$	$2.12 \cdot 10^{-2}$	0.935
NL-S	$\widehat{q}$	$-7.53 \cdot 10^{-2}$	$3.02 \cdot 10^{-2}$	$3.59 \cdot 10^{-2}$	$I_{q,4}$	$3.81 \cdot 10^{-2}$	0.954
	$\widehat{q}^{W}$	$1.10 \cdot 10^{-3}$	$1.24 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$I_{q,3}$	$1.07 \cdot 10^{-2}$	0.924
	$\widehat{q}^{W,BR}$	$-4.30 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$I_{q,1}$	$9.23 \cdot 10^{-3}$	0.861
					$I_{q,2}$	$9.61 \cdot 10^{-3}$	0.867
	$\widehat{e}$	$-3.85 \cdot 10^{-2}$	$3.55 \cdot 10^{-2}$	$3.70 \cdot 10^{-2}$	$I_{E,4}$	$1.84 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.71 \cdot 10^{-1}$	$4.70 \cdot 10^{-2}$	$3.73 \cdot 10^{-1}$	$I_{E,3}$	$3.69 \cdot 10^{-2}$	0.044
	$\widehat{e}^{W,BR}$	$-3.54 \cdot 10^{-2}$	$1.66 \cdot 10^{-2}$	$1.78 \cdot 10^{-2}$	$I_{E,1}$	$4.36 \cdot 10^{-3}$	0.667
					$I_{E,2}$	$1.86 \cdot 10^{-2}$	0.913
NL-C	$\widehat{q}$	$-5.09 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	$2.21 \cdot 10^{-2}$	$I_{q,4}$	$2.52 \cdot 10^{-2}$	0.961
	$\widehat{q}^{W}$	$1.37 \cdot 10^{-2}$	$8.12 \cdot 10^{-3}$	$8.31 \cdot 10^{-3}$	$I_{q,3}$	$7.12 \cdot 10^{-3}$	0.930
	$\widehat{q}^{W,BR}$	$-2.21 \cdot 10^{-2}$	$7.71 \cdot 10^{-3}$	$8.20 \cdot 10^{-3}$	$I_{q,1}$	$6.14 \cdot 10^{-3}$	0.882
					$I_{q,2}$	$6.38 \cdot 10^{-3}$	0.890
	$\widehat{e}$	$-1.47 \cdot 10^{-2}$	$1.99 \cdot 10^{-2}$	$2.01 \cdot 10^{-2}$	$I_{E,4}$	$1.39 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.49 \cdot 10^{-1}$	$3.54 \cdot 10^{-2}$	$3.36 \cdot 10^{-1}$	$I_{E,3}$	$2.47 \cdot 10^{-2}$	0.004
	$\widehat{e}^{W,BR}$	$-1.63 \cdot 10^{-2}$	$9.79 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$I_{E,1}$	$2.39 \cdot 10^{-3}$	0.665
					$I_{E,2}$	$1.08 \cdot 10^{-2}$	0.924

TABLE C.1

Estimation and inference results related to the models considered in Figure 3, at $x = 0.1$ and 0.3. Biases, variances and MSEs are reported for the log-estimates. The median estimated variance reported in the penultimate column is the median of the estimated asymptotic variance for each estimator divided by the square of its convergence rate. The last column gives the empirical coverage of each confidence interval.

$x = 0.5$	Estimate	Bias	Variance	MSE	C.I.	Med. est. var.	Cov.
NL-P	$\widehat{q}$	$-7.57 \cdot 10^{-2}$	$3.69 \cdot 10^{-2}$	$4.26 \cdot 10^{-2}$	$I_{q,4}$	$4.32 \cdot 10^{-2}$	0.946
	$\widehat{q}^{W}$	$9.08 \cdot 10^{-3}$	$1.37 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$I_{q,3}$	$1.23 \cdot 10^{-2}$	0.918
	$\widehat{q}^{W,BR}$	$-3.81 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	$I_{q,1}$	$1.07 \cdot 10^{-2}$	0.883
					$I_{q,2}$	$1.11 \cdot 10^{-2}$	0.888
	$\widehat{e}$	$-3.97 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$I_{E,4}$	$2.37 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.80 \cdot 10^{-1}$	$4.91 \cdot 10^{-2}$	$3.86 \cdot 10^{-1}$	$I_{E,3}$	$4.31 \cdot 10^{-2}$	0.113
	$\widehat{e}^{W,BR}$	$-3.15 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$I_{E,1}$	$5.60 \cdot 10^{-3}$	0.667
					$I_{E,2}$	$2.31 \cdot 10^{-2}$	0.920
NL-S	$\widehat{q}$	$-3.19 \cdot 10^{-2}$	$2.03 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	$I_{q,4}$	$2.36 \cdot 10^{-2}$	0.961
	$\widehat{q}^{W}$	$2.54 \cdot 10^{-2}$	$7.93 \cdot 10^{-3}$	$8.58 \cdot 10^{-3}$	$I_{q,3}$	$6.76 \cdot 10^{-3}$	0.915
	$\widehat{q}^{W,BR}$	$-9.36 \cdot 10^{-3}$	$7.53 \cdot 10^{-3}$	$7.62 \cdot 10^{-3}$	$I_{q,1}$	$5.85 \cdot 10^{-3}$	0.888
					$I_{q,2}$	$6.09 \cdot 10^{-3}$	0.902
	$\widehat{e}$	$3.75 \cdot 10^{-3}$	$2.19 \cdot 10^{-2}$	$2.19 \cdot 10^{-2}$	$I_{E,4}$	$1.71 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.60 \cdot 10^{-1}$	$3.66 \cdot 10^{-2}$	$3.50 \cdot 10^{-1}$	$I_{E,3}$	$2.37 \cdot 10^{-2}$	0.002
	$\widehat{e}^{W,BR}$	$-1.03 \cdot 10^{-3}$	$9.84 \cdot 10^{-3}$	$9.84 \cdot 10^{-3}$	$I_{E,1}$	$2.37 \cdot 10^{-3}$	0.681
					$I_{E,2}$	$1.07 \cdot 10^{-2}$	0.932
NL-C	$\widehat{q}$	$-4.98 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	$I_{q,4}$	$2.30 \cdot 10^{-2}$	0.948
	$\widehat{q}^{W}$	$1.27 \cdot 10^{-2}$	$7.32 \cdot 10^{-3}$	$7.48 \cdot 10^{-3}$	$I_{q,3}$	$6.60 \cdot 10^{-3}$	0.920
	$\widehat{q}^{W,BR}$	$-2.16 \cdot 10^{-2}$	$6.96 \cdot 10^{-3}$	$7.42 \cdot 10^{-3}$	$I_{q,1}$	$5.70 \cdot 10^{-3}$	0.890
					$I_{q,2}$	$5.94 \cdot 10^{-3}$	0.894
	$\widehat{e}$	$-1.44 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$2.07 \cdot 10^{-2}$	$I_{E,4}$	$1.28 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.39 \cdot 10^{-1}$	$3.43 \cdot 10^{-2}$	$3.25 \cdot 10^{-1}$	$I_{E,3}$	$2.24 \cdot 10^{-2}$	0.003
	$\widehat{e}^{W,BR}$	$-1.57 \cdot 10^{-2}$	$9.31 \cdot 10^{-3}$	$9.56 \cdot 10^{-3}$	$I_{E,1}$	$2.24 \cdot 10^{-3}$	0.648
					$I_{E,2}$	$1.00 \cdot 10^{-2}$	0.920

$x = 0.7$	Estimate	Bias	Variance	MSE	C.I.	Med. est. var.	Cov.
NL-P	$\widehat{q}$	$-4.67 \cdot 10^{-2}$	$3.38 \cdot 10^{-2}$	$3.60 \cdot 10^{-2}$	$I_{q,4}$	$4.09 \cdot 10^{-2}$	0.964
	$\widehat{q}^{W}$	$1.75 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	$I_{q,3}$	$1.15 \cdot 10^{-2}$	0.931
	$\widehat{q}^{W,BR}$	$-2.85 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	$1.28 \cdot 10^{-2}$	$I_{q,1}$	$9.97 \cdot 10^{-3}$	0.895
					$I_{q,2}$	$1.04 \cdot 10^{-2}$	0.898
	$\widehat{e}$	$-1.18 \cdot 10^{-2}$	$4.03 \cdot 10^{-2}$	$4.05 \cdot 10^{-2}$	$I_{E,4}$	$2.57 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$6.02 \cdot 10^{-1}$	$4.71 \cdot 10^{-2}$	$4.09 \cdot 10^{-1}$	$I_{E,3}$	$4.17 \cdot 10^{-2}$	0.050
	$\widehat{e}^{W,BR}$	$-1.27 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$1.82 \cdot 10^{-2}$	$I_{E,1}$	$5.10 \cdot 10^{-3}$	0.711
					$I_{E,2}$	$2.17 \cdot 10^{-2}$	0.948
NL-S	$\widehat{q}$	$-1.00 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$I_{q,4}$	$1.53 \cdot 10^{-2}$	0.975
	$\widehat{q}^{W}$	$3.06 \cdot 10^{-2}$	$4.72 \cdot 10^{-3}$	$5.66 \cdot 10^{-3}$	$I_{q,3}$	$4.29 \cdot 10^{-3}$	0.925
	$\widehat{q}^{W,BR}$	$2.75 \cdot 10^{-3}$	$4.47 \cdot 10^{-3}$	$4.48 \cdot 10^{-3}$	$I_{q,1}$	$3.70 \cdot 10^{-3}$	0.915
					$I_{q,2}$	$3.85 \cdot 10^{-3}$	0.920
	$\widehat{e}$	$1.60 \cdot 10^{-2}$	$1.03 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$I_{E,4}$	$1.86 \cdot 10^{-2}$	1.000
	$\widehat{e}^{W}$	$5.23 \cdot 10^{-1}$	$2.69 \cdot 10^{-2}$	$3.01 \cdot 10^{-1}$	$I_{E,3}$	$1.70 \cdot 10^{-2}$	0.000
	$\widehat{e}^{W,BR}$	$1.02 \cdot 10^{-2}$	$5.33 \cdot 10^{-3}$	$5.43 \cdot 10^{-3}$	$I_{E,1}$	$1.19 \cdot 10^{-3}$	0.663
					$I_{E,2}$	$5.81 \cdot 10^{-3}$	0.952
NL-C	$\widehat{q}$	$-3.67 \cdot 10^{-2}$	$2.03 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$I_{q,4}$	$2.52 \cdot 10^{-2}$	0.967
	$\widehat{q}^{W}$	$1.55 \cdot 10^{-2}$	$7.77 \cdot 10^{-3}$	$8.01 \cdot 10^{-3}$	$I_{q,3}$	$7.10 \cdot 10^{-3}$	0.930
	$\widehat{q}^{W,BR}$	$-2.04 \cdot 10^{-2}$	$7.38 \cdot 10^{-3}$	$7.80 \cdot 10^{-3}$	$I_{q,1}$	$6.13 \cdot 10^{-3}$	0.896
					$I_{q,2}$	$6.38 \cdot 10^{-3}$	0.902
	$\widehat{e}$	$-4.02 \cdot 10^{-3}$	$2.08 \cdot 10^{-2}$	$2.08 \cdot 10^{-2}$	$I_{E,4}$	$1.40 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.59 \cdot 10^{-1}$	$3.64 \cdot 10^{-2}$	$3.49 \cdot 10^{-1}$	$I_{E,3}$	$2.55 \cdot 10^{-2}$	0.004
	$\widehat{e}^{W,BR}$	$-9.14 \cdot 10^{-3}$	$9.64 \cdot 10^{-3}$	$9.72 \cdot 10^{-3}$	$I_{E,1}$	$2.47 \cdot 10^{-3}$	0.680
					$I_{E,2}$	$1.12 \cdot 10^{-2}$	0.949

TABLE C.2
As in Table C.1, at $x = 0.5$ and 0.7.

$x = 0.9$	Estimate	Bias	Variance	MSE	C.I.	Med. est. var.	Cov.
NL-P	$\widehat{q}$	$-2.88 \cdot 10^{-2}$	$2.29 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$	$I_{q,4}$	$3.08 \cdot 10^{-2}$	0.970
	$\widehat{q}^{W}$	$3.47 \cdot 10^{-2}$	$9.58 \cdot 10^{-3}$	$1.08 \cdot 10^{-2}$	$I_{q,3}$	$8.13 \cdot 10^{-3}$	0.895
	$\widehat{q}^{W,BR}$	$-3.98 \cdot 10^{-3}$	$9.24 \cdot 10^{-3}$	$9.26 \cdot 10^{-3}$	$I_{q,1}$	$6.94 \cdot 10^{-3}$	0.888
					$I_{q,2}$	$7.21 \cdot 10^{-3}$	0.893
	$\widehat{e}$	$1.57 \cdot 10^{-2}$	$2.52 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	$I_{E,4}$	$3.19 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$6.23 \cdot 10^{-1}$	$6.03 \cdot 10^{-2}$	$4.49 \cdot 10^{-1}$	$I_{E,3}$	$3.22 \cdot 10^{-2}$	0.005
	$\widehat{e}^{W,BR}$	$5.31 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$I_{E,1}$	$2.86 \cdot 10^{-3}$	0.706
					$I_{E,2}$	$1.34 \cdot 10^{-2}$	0.956
NL-S	$\widehat{q}$	$-2.31 \cdot 10^{-2}$	$1.84 \cdot 10^{-2}$	$1.89 \cdot 10^{-2}$	$I_{q,4}$	$2.37 \cdot 10^{-2}$	0.974
	$\widehat{q}^{W}$	$3.64 \cdot 10^{-2}$	$6.77 \cdot 10^{-3}$	$8.10 \cdot 10^{-3}$	$I_{q,3}$	$6.30 \cdot 10^{-3}$	0.914
	$\widehat{q}^{W,BR}$	$2.22 \cdot 10^{-3}$	$6.48 \cdot 10^{-3}$	$6.49 \cdot 10^{-3}$	$I_{q,1}$	$5.39 \cdot 10^{-3}$	0.902
					$I_{q,2}$	$5.59 \cdot 10^{-3}$	0.909
	$\widehat{e}$	$1.46 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	$I_{E,4}$	$2.31 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.84 \cdot 10^{-1}$	$4.45 \cdot 10^{-2}$	$3.86 \cdot 10^{-1}$	$I_{E,3}$	$2.52 \cdot 10^{-2}$	0.000
	$\widehat{e}^{W,BR}$	$8.75 \cdot 10^{-3}$	$7.57 \cdot 10^{-3}$	$7.65 \cdot 10^{-3}$	$I_{E,1}$	$1.95 \cdot 10^{-3}$	0.689
					$I_{E,2}$	$9.44 \cdot 10^{-3}$	0.960
NL-C	$\widehat{q}$	$-5.37 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	$I_{q,4}$	$3.11 \cdot 10^{-2}$	0.967
	$\widehat{q}^{W}$	$1.89 \cdot 10^{-2}$	$8.97 \cdot 10^{-3}$	$9.33 \cdot 10^{-3}$	$I_{q,3}$	$8.40 \cdot 10^{-3}$	0.921
	$\widehat{q}^{W,BR}$	$-2.06 \cdot 10^{-2}$	$8.54 \cdot 10^{-3}$	$8.97 \cdot 10^{-3}$	$I_{q,1}$	$7.16 \cdot 10^{-3}$	0.891
					$I_{q,2}$	$7.44 \cdot 10^{-3}$	0.897
	$\widehat{e}$	$-1.21 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$I_{E,4}$	$1.72 \cdot 10^{-1}$	1.000
	$\widehat{e}^{W}$	$5.81 \cdot 10^{-1}$	$4.38 \cdot 10^{-2}$	$3.82 \cdot 10^{-1}$	$I_{E,3}$	$3.08 \cdot 10^{-2}$	0.008
	$\widehat{e}^{W,BR}$	$-1.37 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$1.07 \cdot 10^{-2}$	$I_{E,1}$	$2.86 \cdot 10^{-3}$	0.683
					$I_{E,2}$	$1.32 \cdot 10^{-2}$	0.942

TABLE C.3

As in Table C.1, at $x = 0.9$.

REFERENCES

[1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

[2] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge University Press.

[3] BOSQ, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer Science & Business Media.

[4] BOWMAN, A. W. and AZZALINI, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations (Oxford Statistical Science Series, volume 18)*. Oxford University Press.

[5] BRADLEY, R. C. (1981). Central limit theorems under weak dependence. *Journal of Multivariate Analysis* **11** 1–16.

[6] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2** 107–144.

[7] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods (Second Edition)*. Springer.

[8] BÜCHER, A. and RUPPERT, M. (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis* **116** 208–229.

[9] CHAUDHURI, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* **39** 246–269.

[10] CHEN, J., LI, D., LINTON, O. and LU, Z. (2018). Semiparametric Ultra-High Dimensional Model Averaging of Nonlinear Dynamic Time Series. *Journal of the American Statistical Association* **113** 919–932.

[11] CHEN, X., LINTON, O. and ROBINSON, P. M. (2000). The estimation of conditional densities. In *Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas* (M. L. Puri, ed.) 71–84. VSP International Science Publishers, Netherlands.

[12] CHERNICK, M. R. (1981). A Limit Theorem for the Maximum of Autoregressive Processes with Uniform Marginal Distributions. *Annals of Probability* **9** 145–149.

[13] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and KAJI, T. (2017). Extremal Quantile Regression. In *Handbook of Quantile Regression* (R. Koenker, V. Chernozhukov, X. He and L. Peng, eds.) Chapman and Hall/CRC.

[14] DAOUIA, A., GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19** 2557–2589.

[15] DAOUIA, A., GARDES, L., GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves. *TEST* **20** 311–333.

[16] DAOUIA, A., GIRARD, S. and STUPFLER, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B* **80** 263–292.

[17] DAOUIA, A., GIRARD, S. and STUPFLER, G. (2020). Tail expectile process and risk assessment. *Bernoulli* **26** 531–556.

[18] DAVISON, A. C., PADOAN, S. A. and STUPFLER, G. (2023). Tail risk inference via expectiles in heavy-tailed time series. *Journal of Business and Economic Statistics* **41** 876–889.

[19] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

[20] DOUKHAN, P. (1994). *Mixing: Properties and Examples*. Springer, New York.

[21] DREES, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9** 617–657.

[22] EL METHNI, J., GARDES, L. and GIRARD, S. (2014). Non-parametric Estimation of Extreme Risk Measures from Conditional Heavy-tailed Distributions. *Scandinavian Journal of Statistics* **41** 988–1012.

[23] FRAGA ALVES, M., GOMES, M. I. and DE HAAN, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60** 193–214.

[24] FRANCQ, C. and ZAKOÏAN, J. M. (2019). *GARCH Models (Second Edition)*. John Wiley & Sons, Hoboken, NJ.

[25] GARDES, L. and STUPFLER, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes* **17** 45–75.

[26] GARDES, L. and STUPFLER, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT: Statistical Journal* **17** 109–144.

[27] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2022). Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics* **49** 78–115.

[28] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statistics & Computing* **32** 64.

[29] GOMES, M. I., MARTINS, M. J. and NEVES, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events - The jackknife methodology. *Extremes* **3** 207-229.

[30] GOMES, M. I., BRILHANTE, M. F. and PESTANA, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics-Simulation and Computation* **45** 833–862.

[31] GOMES, M. I. and MARTINS, M. J. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5** 5–31.

[32] GOMES, M. I. and PESTANA, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association* **102** 280–292.

[33] HALL, P. (1982). On Some Simple Estimates of an Exponent of Regular Variation. *Journal of the Royal Statistical Society: Series B* **44** 37–42.

[34] IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications* **7** 349–382.

[35] MARTINS-FILHO, C., YAO, F. and TORERO, M. (2018). Nonparametric estimation of Conditional Value-at-Risk and Expected Shortfall based on extreme value theory. *Econometric Theory* **34** 23–67.

[36] MASRY, E. and FAN, J. (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian Journal of Statistics* **24** 165–179.

[37] MASRY, E. and TJØSTHEIM, D. (1995). Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality. *Econometric Theory* **11** 258–289.

[38] MCMURRY, T. L. and POLITIS, D. N. (2008). Minimally biased nonparametric regression and autoregression. *REVSTAT - Statistical Journal* **6** 123–150.

[39] PHILIPP, W. and WEBB, G. R. (1973). An invariance principle for mixing sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **25** 223–237.

[40] REISS, R. D. and THOMAS, M. (2007). *Statistical Analysis of Extreme Values*. Birkhäuser Basel.

[41] RIO, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. *In Probability Theory and Stochastic Modelling*, volume 80, Springer-Verlag, Berlin-Heidelberg.

[42] ROOTZÉN, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications* **119** 468–490.

[43] ROUSSAS, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications* **36** 107–116.

[44] STUPFLER, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics* **7** 2298–2343.

[45] STUPFLER, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis* **144** 1–24.

[46] STUPFLER, G. and USSEGLIO-CARLEVE, A. (2023). Composite bias-reduced L^p-quantile-based estimators of extreme quantiles and expectiles. *Canadian Journal of Statistics* **51** 704–742.

[47] UTEV, S. A. (1990). On the central limit theorem for $\varphi-$mixing arrays of random variables. *Theory of Probability and its Applications* **35** 131–139.

[48] VILAR-FERNÁNDEZ, J. A. and VILAR-FERNÁNDEZ, J. M. (1998). Recursive estimation of regression functions by local polynomial fitting. *Annals of the Institute of Statistical Mathematics* **50** 729–754.